



# IJCSI

## **International Journal of Computer Science Issues**

**Volume 8, Issue 3, No 1, May 2011  
ISSN (Online): 1694-0814**

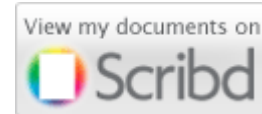
**© IJCSI PUBLICATION  
[www.IJCSI.org](http://www.IJCSI.org)**

**IJCSI proceedings are currently indexed by:**



**Cogprints**

**Google scholar**



**SciRate.com**

**CiteSeer<sup>x</sup> beta**



**DOAJ** DIRECTORY OF OPEN ACCESS JOURNALS



**ProQuest**

## **IJCSI Publicity Board 2011**

**Dr. Borislav D Dimitrov**

Department of General Practice, Royal College of Surgeons in Ireland  
Dublin, Ireland

**Dr. Vishal Goyal**

Department of Computer Science, Punjabi University  
Patiala, India

**Mr. Nehinbe Joshua**

University of Essex  
Colchester, Essex, UK

**Mr. Vassilis Papataxiarhis**

Department of Informatics and Telecommunications  
National and Kapodistrian University of Athens, Athens, Greece

## **EDITORIAL**

In this third edition of 2011, we bring forward issues from various dynamic computer science fields ranging from system performance, computer vision, artificial intelligence, software engineering, multimedia, pattern recognition, information retrieval, databases, security and networking among others.

Considering the growing interest of academics worldwide to publish in IJCSI, we invite universities and institutions to partner with us to further encourage open-access publications.

As always we thank all our reviewers for providing constructive comments on papers sent to them for review. This helps enormously in improving the quality of papers published in this issue.

Google Scholar reported a large amount of cited papers published in IJCSI. We will continue to encourage the readers, authors and reviewers and the computer science scientific community and interested authors to continue citing papers published by the journal.

It was with pleasure and a sense of satisfaction that we announced in mid March 2011 our 2-year Impact Factor which is evaluated at 0.242. For more information about this please see the FAQ section of the journal.

Apart from availability of the full-texts from the journal website, all published papers are deposited in open-access repositories to make access easier and ensure continuous availability of the proceedings free of charge for all researchers.

We are pleased to present IJCSI Volume 8, Issue 3, No 1, May 2011 (IJCSI Vol. 8, Issue 3, No. 1). The acceptance rate for this issue is 33.6%.

IJCSI Editorial Board  
May 2011 Issue  
ISSN (Online): 1694-0814  
© IJCSI Publications  
[www.IJCSI.org](http://www.IJCSI.org)



## **IJCSI Editorial Board 2011**

### **Dr Tristan Vanrullen**

Chief Editor

LPL, Laboratoire Parole et Langage - CNRS - Aix en Provence, France

LABRI, Laboratoire Bordelais de Recherche en Informatique - INRIA - Bordeaux, France

LEEE, Laboratoire d'Esthétique et Expérimentations de l'Espace - Université d'Auvergne, France

### **Dr Constantino Malagón**

Associate Professor

Nebrija University

Spain

### **Dr Lamia Fourati Chaari**

Associate Professor

Multimedia and Informatics Higher Institute in SFAX

Tunisia

### **Dr Mokhtar Beldjehem**

Professor

Sainte-Anne University

Halifax, NS, Canada

### **Dr Pascal Chatonnay**

Assistant Professor

Maître de Conférences

Laboratoire d'Informatique de l'Université de Franche-Comté

Université de Franche-Comté

France

### **Dr Karim Mohammed Rezaul**

Centre for Applied Internet Research (CAIR)

Glyndwr University

Wrexham, United Kingdom

### **Dr Yee-Ming Chen**

Professor

Department of Industrial Engineering and Management

Yuan Ze University

Taiwan

### **Dr Vishal Goyal**

Assistant Professor

Department of Computer Science

Punjabi University

Patiala, India

**Dr Dalbir Singh**

Faculty of Information Science And Technology  
National University of Malaysia  
Malaysia

**Dr Natarajan Meghanathan**

Assistant Professor  
REU Program Director  
Department of Computer Science  
Jackson State University  
Jackson, USA

**Dr Deepak Laxmi Narasimha**

Department of Software Engineering,  
Faculty of Computer Science and Information Technology,  
University of Malaya,  
Kuala Lumpur, Malaysia

**Dr. Prabhat K. Mahanti**

Professor  
Computer Science Department,  
University of New Brunswick  
Saint John, N.B., E2L 4L5, Canada

**Dr Navneet Agrawal**

Assistant Professor  
Department of ECE,  
College of Technology & Engineering,  
MPUAT, Udaipur 313001 Rajasthan, India

**Dr Panagiotis Michailidis**

Division of Computer Science and Mathematics,  
University of Western Macedonia,  
53100 Florina, Greece

**Dr T. V. Prasad**

Professor  
Department of Computer Science and Engineering,  
Lingaya's University  
Faridabad, Haryana, India

**Dr Saqib Rasool Chaudhry**

Wireless Networks and Communication Centre  
261 Michael Sterling Building  
Brunel University West London, UK, UB8 3PH

**Dr Shishir Kumar**

Department of Computer Science and Engineering,  
Jaypee University of Engineering & Technology  
Raghogarh, MP, India

**Dr P. K. Suri**

Professor  
Department of Computer Science & Applications,  
Kurukshetra University,  
Kurukshetra, India

**Dr Paramjeet Singh**

Associate Professor  
GZS College of Engineering & Technology,  
India

**Dr Shaveta Rani**

Associate Professor  
GZS College of Engineering & Technology,  
India

**Dr. Seema Verma**

Associate Professor,  
Department Of Electronics,  
Banasthali University,  
Rajasthan - 304022, India

**Dr G. Ganesan**

Professor  
Department of Mathematics,  
Adikavi Nannaya University,  
Rajahmundry, A.P, India

**Dr A. V. Senthil Kumar**

Department of MCA,  
Hindusthan College of Arts and Science,  
Coimbatore, Tamilnadu, India

**Dr Jyoteesh Malhotra**

ECE Department,  
Guru Nanak Dev University,  
Jalandhar, Punjab, India

**Dr R. Ponnusamy**

Professor  
Department of Computer Science & Engineering,  
Aarupadai Veedu Institute of Technology,  
Vinayaga Missions University, Chennai, Tamilnadu, India.

**N. Jaisankar**

Assistant Professor  
School of Computing Sciences,  
VIT University  
Vellore, Tamilnadu, India

## IJCSI Reviewers Committee 2011

- Mr. Markus Schatten, University of Zagreb, Faculty of Organization and Informatics, Croatia
- Mr. Vassilis Papataxiarhis, Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, Athens, Greece
- Dr Modestos Stavrakis, University of the Aegean, Greece
- Dr Fadi KHALIL, LAAS -- CNRS Laboratory, France
- Dr Dimitar Trajanov, Faculty of Electrical Engineering and Information technologies, ss. Cyril and Methodius Univesity - Skopje, Macedonia
- Dr Jinping Yuan, College of Information System and Management, National Univ. of Defense Tech., China
- Dr Alexis Lazanas, Ministry of Education, Greece
- Dr Stavroula Mougiakakou, University of Bern, ARTORG Center for Biomedical Engineering Research, Switzerland
- Dr Cyril de Runz, CReSTIC-SIC, IUT de Reims, University of Reims, France
- Mr. Pramodkumar P. Gupta, Dept of Bioinformatics, Dr D Y Patil University, India
- Dr Alireza Fereidunian, School of ECE, University of Tehran, Iran
- Mr. Fred Viezens, Otto-Von-Guericke-University Magdeburg, Germany
- Dr. Richard G. Bush, Lawrence Technological University, United States
- Dr. Ola Osunkoya, Information Security Architect, USA
- Mr. Kotsokostas N. Antonios, TEI Piraeus, Hellas
- Prof Steven Totosy de Zepetnek, U of Halle-Wittenberg & Purdue U & National Sun Yat-sen U, Germany, USA, Taiwan
- Mr. M Arif Siddiqui, Najran University, Saudi Arabia
- Ms. Ilknur Icke, The Graduate Center, City University of New York, USA
- Prof Miroslav Baca, Faculty of Organization and Informatics, University of Zagreb, Croatia
- Dr. Elvia Ruiz Beltrán, Instituto Tecnológico de Aguascalientes, Mexico
- Mr. Moustafa Banbouk, Engineer du Telecom, UAE
- Mr. Kevin P. Monaghan, Wayne State University, Detroit, Michigan, USA
- Ms. Moira Stephens, University of Sydney, Australia
- Ms. Maryam Feily, National Advanced IPv6 Centre of Excellence (NAV6) , Universiti Sains Malaysia (USM), Malaysia
- Dr. Constantine YIALOURIS, Informatics Laboratory Agricultural University of Athens, Greece
- Mrs. Angeles Abella, U. de Montreal, Canada
- Dr. Patrizio Arrigo, CNR ISMAC, Italy
- Mr. Anirban Mukhopadhyay, B.P.Poddar Institute of Management & Technology, India
- Mr. Dinesh Kumar, DAV Institute of Engineering & Technology, India
- Mr. Jorge L. Hernandez-Ardieta, INDRA SISTEMAS / University Carlos III of Madrid, Spain
- Mr. AliReza Shahrestani, University of Malaya (UM), National Advanced IPv6 Centre of Excellence (NAv6), Malaysia
- Mr. Blagoj Ristevski, Faculty of Administration and Information Systems Management - Bitola, Republic of Macedonia
- Mr. Mauricio Egidio Cantão, Department of Computer Science / University of São Paulo, Brazil
- Mr. Jules Ruis, Fractal Consultancy, The Netherlands

- Mr. Mohammad Iftekhar Husain, University at Buffalo, USA
- Dr. Deepak Laxmi Narasimha, Department of **Software** Engineering, Faculty of Computer Science and Information Technology, University of Malaya, Malaysia
- Dr. Paola Di Maio, DMEM University of Strathclyde, UK
- Dr. Bhanu Pratap Singh, Institute of Instrumentation Engineering, Kurukshetra University Kurukshetra, India
- Mr. Sana Ullah, Inha University, South Korea
- Mr. Cornelis Pieter Pieters, Condast, The Netherlands
- Dr. Amogh Kavimandan, The MathWorks Inc., USA
- Dr. Zhinan Zhou, Samsung Telecommunications America, USA
- Mr. Alberto de Santos Sierra, Universidad Politécnica de Madrid, Spain
- Dr. Md. Atiqur Rahman Ahad, Department of Applied Physics, Electronics & Communication Engineering (APECE), University of Dhaka, Bangladesh
- Dr. Charalampos Bratsas, Lab of Medical Informatics, Medical Faculty, Aristotle University, Thessaloniki, Greece
- Ms. Alexia Dini Kounoudes, Cyprus University of Technology, Cyprus
- Dr. Jorge A. Ruiz-Vanoye, Universidad Juárez Autónoma de Tabasco, Mexico
- Dr. Alejandro Fuentes Penna, Universidad Popular Autónoma del Estado de Puebla, México
- Dr. Ocotlán Díaz-Parra, Universidad Juárez Autónoma de Tabasco, México
- Mrs. Nantia Iakovidou, Aristotle University of Thessaloniki, Greece
- Mr. Vinay Chopra, DAV Institute of Engineering & Technology, Jalandhar
- Ms. Carmen Lastres, Universidad Politécnica de Madrid - Centre for Smart Environments, Spain
- Dr. Sanja Lazarova-Molnar, United Arab Emirates University, UAE
- Mr. Srikrishna Nudurumati, Imaging & Printing Group R&D Hub, Hewlett-Packard, India
- Dr. Olivier Nocent, CReSTIC/SIC, University of Reims, France
- Mr. Burak Cizmeci, Isik University, Turkey
- Dr. Carlos Jaime Barrios Hernandez, LIG (Laboratory Of Informatics of Grenoble), France
- Mr. Md. Rabiul Islam, Rajshahi university of Engineering & Technology (RUET), Bangladesh
- Dr. LAKHOUA Mohamed Najeh, ISSAT - Laboratory of Analysis and Control of Systems, Tunisia
- Dr. Alessandro Lavacchi, Department of Chemistry - University of Firenze, Italy
- Mr. Mungwe, University of Oldenburg, Germany
- Mr. Somnath Tagore, Dr D Y Patil University, India
- Ms. Xueqin Wang, ATCS, USA
- Dr. Borislav D Dimitrov, Department of General Practice, Royal College of Surgeons in Ireland, Dublin, Ireland
- Dr. Fondjo Fotou Franklin, Langston University, USA
- Dr. Vishal Goyal, Department of Computer Science, Punjabi University, Patiala, India
- Mr. Thomas J. Clancy, ACM, United States
- Dr. Ahmed Nabih Zaki Rashed, Dr. in Electronic Engineering, Faculty of Electronic Engineering, menouf 32951, Electronics and Electrical Communication Engineering Department, Menoufia university, EGYPT, EGYPT
- Dr. Rushed Kanawati, LIPN, France
- Mr. Koteswar Rao, K G Reddy College Of ENGG.&TECH,CHILKUR, RR DIST.,AP, India
- Mr. M. Nagesh Kumar, Department of Electronics and Communication, J.S.S. research foundation, Mysore University, Mysore-6, India

- Dr. Ibrahim Noha, Grenoble Informatics Laboratory, France
- Mr. Muhammad Yasir Qadri, University of Essex, UK
- Mr. Annadurai .P, KMCPGS, Lawspet, Pondicherry, India, (Aff. Pondicherry Univeristy, India)
- Mr. E Munivel , CEDTI (Govt. of India), India
- Dr. Chitra Ganesh Desai, University of Pune, India
- Mr. Syed, Analytical Services & Materials, Inc., USA
- Mrs. Payal N. Raj, Veer South Gujarat University, India
- Mrs. Priti Maheshwary, Maulana Azad National Institute of Technology, Bhopal, India
- Mr. Mahesh Goyani, S.P. University, India, India
- Mr. Vinay Verma, Defence Avionics Research Establishment, DRDO, India
- Dr. George A. Papakostas, Democritus University of Thrace, Greece
- Mr. Abhijit Sanjiv Kulkarni, DARE, DRDO, India
- Mr. Kavi Kumar Khedo, University of Mauritius, Mauritius
- Dr. B. Sivaselvan, Indian **Institute** of Information Technology, Design & Manufacturing, Kancheepuram, IIT Madras Campus, India
- Dr. Partha Pratim Bhattacharya, Greater Kolkata College of Engineering and Management, **West Bengal** University of Technology, India
- Mr. Manish Maheshwari, Makhanlal C University of Journalism & Communication, India
- Dr. Siddhartha Kumar Khaitan, Iowa State University, USA
- Dr. Mandhapati Raju, General Motors Inc, USA
- Dr. M.Iqbal Saripan, Universiti Putra Malaysia, Malaysia
- Mr. Ahmad Shukri Mohd Noor, University Malaysia Terengganu, Malaysia
- Mr. Selvakuberan K, TATA Consultancy Services, India
- Dr. Smita Rajpal, Institute of Technology and Management, Gurgaon, India
- Mr. Rakesh Kachroo, Tata Consultancy Services, India
- Mr. Raman Kumar, National Institute of Technology, Jalandhar, Punjab., India
- Mr. Nitesh Sureja, S.P.University, India
- Dr. M. Emre Celebi, Louisiana State University, Shreveport, USA
- Dr. Aung Kyaw Oo, Defence Services Academy, Myanmar
- Mr. Sanjay P. Patel, Sankalchand Patel College of Engineering, Visnagar, Gujarat, India
- Dr. Pascal Fallavollita, Queens University, Canada
- Mr. Jitendra Agrawal, Rajiv Gandhi Technological University, Bhopal, MP, India
- Mr. Ismael Rafael Ponce Medellín, Cenidet (Centro Nacional de Investigación y Desarrollo Tecnológico), Mexico
- Mr. Supheakmongkol SARIN, Waseda University, Japan
- Mr. Shoukat Ullah, Govt. Post Graduate College Bannu, Pakistan
- Dr. Vivian Augustine, Telecom Zimbabwe, Zimbabwe
- Mrs. Mutalli Vatile, Offshore Business Philipines, Philipines
- Mr. Pankaj Kumar, SAMA, India
- Dr. Himanshu Aggarwal, Punjabi University,Patiala, India
- Dr. Vauvert Guillaume, Europages, France
- Prof Yee Ming Chen, Department of Industrial Engineering and Management, Yuan Ze University, Taiwan
- Dr. Constantino Malagón, Nebrija University, Spain
- Prof Kanwalvir Singh Dhindsa, B.B.S.B.Engg.College, Fatehgarh Sahib (Punjab), India

- Mr. Angkoon Phinyomark, Prince of Singkla University, Thailand
- Ms. Nital H. Mistry, Veer Narmad South Gujarat University, Surat, India
- Dr. M.R.Sumalatha, Anna University, India
- Mr. Somesh Kumar Dewangan, Disha Institute of Management and Technology, India
- Mr. Raman Maini, Punjabi University, Patiala(Punjab)-147002, India
- Dr. Abdelkader Outtagarts, Alcatel-Lucent Bell-Labs, France
- Prof Dr. Abdul Wahid, AKG Engg. College, Ghaziabad, India
- Mr. Prabu Mohandas, Anna University/Adhiyamaan College of Engineering, india
- Dr. Manish Kumar Jindal, Panjab University Regional Centre, Muktsar, India
- Prof Mydhili K Nair, M S Ramaiah Institute of Technnology, Bangalore, India
- Dr. C. Suresh Gnana Dhas, VelTech MultiTech Dr.Rangarajan Dr.Sagunthala Engineering College,Chennai,Tamilnadu, India
- Prof Akash Rajak, Krishna Institute of Engineering and Technology, Ghaziabad, India
- Mr. Ajay Kumar Shrivastava, Krishna Institute of Engineering & Technology, Ghaziabad, India
- Mr. Deo Prakash, SMVD University, Kakryal(J&K), India
- Dr. Vu Thanh Nguyen, University of Information Technology HoChiMinh City, VietNam
- Prof Deo Prakash, SMVD University (A **Technical** University open on I.I.T. Pattern) Kakryal (J&K), India
- Dr. Navneet Agrawal, Dept. of ECE, College of Technology & Engineering, MPUAT, Udaipur 313001 Rajasthan, India
- Mr. Sufal Das, Sikkim Manipal Institute of Technology, India
- Mr. Anil Kumar, Sikkim Manipal Institute of Technology, India
- Dr. B. Prasanalakshmi, King Saud University, Saudi Arabia.
- Dr. K D Verma, S.V. (P.G.) College, Aligarh, India
- Mr. Mohd Nazri Ismail, System and Networking Department, University of Kuala Lumpur (UniKL), Malaysia
- Dr. Nguyen Tuan Dang, University of Information Technology, Vietnam National University Ho Chi Minh city, Vietnam
- Dr. Abdul Aziz, University of Central Punjab, Pakistan
- Dr. P. Vasudeva Reddy, Andhra University, India
- Mrs. Savvas A. Chatzichristofis, Democritus University of Thrace, Greece
- Mr. Marcio Dorn, Federal University of Rio Grande do Sul - UFRGS Institute of Informatics, Brazil
- Mr. Luca Mazzola, University of Lugano, Switzerland
- Mr. Nadeem Mahmood, Department of Computer Science, University of Karachi, Pakistan
- Mr. Hafeez Ullah Amin, Kohat University of Science & Technology, Pakistan
- Dr. Professor Vikram Singh, Ch. Devi Lal University, Sirsa (Haryana), India
- Mr. M. Azath, Calicut/Mets School of Enginerring, India
- Dr. J. Hanumanthappa, DoS in CS, University of Mysore, India
- Dr. Shahanawaj Ahamad, Department of Computer Science, King Saud University, Saudi Arabia
- Dr. K. Duraiswamy, K. S. Rangasamy College of Technology, India
- Prof. Dr Mazlina Esa, Universiti Teknologi Malaysia, Malaysia
- Dr. P. Vasant, Power Control Optimization (Global), Malaysia
- Dr. Taner Tuncer, Firat University, Turkey
- Dr. Norrozila Sulaiman, University Malaysia Pahang, Malaysia
- Prof. S K Gupta, BCET, Guradspur, India

- Dr. Latha Parameswaran, Amrita Vishwa Vidyapeetham, India
- Mr. M. Azath, Anna University, India
- Dr. P. Suresh Varma, Adikavi Nannaya University, India
- Prof. V. N. Kamalesh, JSS Academy of Technical Education, India
- Dr. D Gunaseelan, Ibri College of Technology, Oman
- Mr. Sanjay Kumar Anand, CDAC, India
- Mr. Akshat Verma, CDAC, India
- Mrs. Fazeela Tunnisa, Najran University, Kingdom of Saudi Arabia
- Mr. Hasan Asil, Islamic Azad University Tabriz Branch (Azarshahr), Iran
- Prof. Dr Sajal Kabiraj, Fr. C Rodrigues Institute of Management **Studies** (Affiliated to University of Mumbai, India), India
- Mr. Syed Fawad Mustafa, GAC Center, Shandong University, China
- Dr. Natarajan Meghanathan, Jackson State University, Jackson, MS, USA
- Prof. Selvakani Kandeegan, Francis Xavier Engineering College, India
- Mr. Tohid Sedghi, Urmia University, Iran
- Dr. S. Sasikumar, PSNA College of Engg and Tech, Dindigul, India
- Dr. Anupam Shukla, Indian Institute of Information Technology and Management Gwalior, India
- Mr. Rahul Kala, Indian Institute of Information Technology and Management Gwalior, India
- Dr. A V Nikolov, National University of Lesotho, Lesotho
- Mr. Kamal Sarkar, Department of Computer Science and Engineering, Jadavpur University, India
- Dr. Mokhled S. Altarawneh, Computer Engineering Dept., Faculty of Engineering, Mutah University, Jordan, Jordan
- Prof. Sattar J Aboud, Iraqi Council of Representatives, Iraq-Baghdad
- Dr. Prasant Kumar Pattnaik, Department of CSE, KIST, India
- Dr. Mohammed Amoon, King Saud University, Saudi Arabia
- Dr. Tsvetanka Georgieva, Department of Information Technologies, St. Cyril and St. Methodius University of Veliko Tarnovo, Bulgaria
- Dr. Eva Volna, University of Ostrava, Czech Republic
- Mr. Ujjal Marjit, University of Kalyani, West-Bengal, India
- Dr. Prasant Kumar Pattnaik, KIST, Bhubaneswar, India, India
- Dr. Guezouri Mustapha, Department of Electronics, Faculty of Electrical Engineering, University of Science and Technology (USTO), Oran, Algeria
- Mr. Maniyar Shiraz Ahmed, Najran University, Najran, Saudi Arabia
- Dr. Sreedhar Reddy, JNTU, SSIIETW, Hyderabad, India
- Mr. Bala Dhandayuthapani Veerasamy, Mekelle University, Ethiopia
- Mr. Arash Habibi Lashkari, University of Malaya (UM), Malaysia
- Mr. Rajesh Prasad, LDC Institute of Technical Studies, Allahabad, India
- Ms. Habib Izadkhah, Tabriz University, Iran
- Dr. Lokesh Kumar Sharma, Chhattisgarh Swami Vivekanand Technical University Bhilai, India
- Mr. Kuldeep Yadav, IIT Delhi, India
- Dr. Naoufel Kraiem, Institut Supérieur d'Informatique, Tunisia
- Prof. Frank Ortmeier, Otto-von-Guericke-Universität Magdeburg, Germany
- Mr. Ashraf Aljammal, USM, Malaysia
- Mrs. Amandeep Kaur, Department of Computer Science, Punjabi University, Patiala, Punjab, India
- Mr. Babak Basharirad, University Technology of Malaysia, Malaysia



- Mr. Avinash singh, Kiet Ghaziabad, India
- Dr. Miguel Vargas-Lombardo, Technological University of Panama, Panama
- Dr. Tuncay Sevindik, Firat University, Turkey
- Ms. Pavai Kandavelu, Anna University Chennai, India
- Mr. Ravish Khichar, Global Institute of Technology, India
- Mr Aos Alaa Zaidan Ansaef, Multimedia University, Cyberjaya, Malaysia
- Dr. Awadhesh Kumar Sharma, Dept. of CSE, MMM Engg College, Gorakhpur-273010, UP, India
- Mr. Qasim Siddique, FUIEMS, Pakistan
- Dr. Le Hoang Thai, University of Science, Vietnam National University - Ho Chi Minh City, Vietnam
- Dr. Saravanan C, NIT, Durgapur, India
- Dr. Vijay Kumar Mago, DAV College, Jalandhar, India
- Dr. Do Van Nhon, University of Information Technology, Vietnam
- Mr. Georgios Kioumourtzis, University of Patras, Greece
- Mr. Amol D.Potgantwar, SITRC Nasik, India
- Mr. Lesedi Melton Masisi, Council for Scientific and Industrial Research, South Africa
- Dr. Karthik.S, Department of Computer Science & Engineering, SNS College of Technology, India
- Mr. Nafiz Imtiaz Bin Hamid, Department of Electrical and Electronic Engineering, Islamic University of Technology (IUT), Bangladesh
- Mr. Muhammad Imran Khan, Universiti Teknologi PETRONAS, Malaysia
- Dr. Abdul Kareem M. Radhi, Information Engineering - Nahrin University, Iraq
- Dr. Mohd Nazri Ismail, University of Kuala Lumpur, Malaysia
- Dr. Manuj Darbari, BBDNITM, Institute of Technology, A-649, Indira Nagar, Lucknow 226016, India
- Ms. Izerrouken, INP-IRIT, France
- Mr. Nitin Ashokrao Naik, Dept. of Computer Science, Yeshwant Mahavidyalaya, Nanded, India
- Mr. Nikhil Raj, National Institute of Technology, Kurukshetra, India
- Prof. Maher Ben Jemaa, National School of Engineers of Sfax, Tunisia
- Prof. Rajeshwar Singh, BRCM College of Engineering and Technology, Bahal Bhiwani, Haryana, India
- Mr. Gaurav Kumar, Department of Computer Applications, Chitkara Institute of Engineering and Technology, Rajpura, Punjab, India
- Mr. Ajeet Kumar Pandey, Indian Institute of Technology, Kharagpur, India
- Mr. Rajiv Phougat, IBM Corporation, USA
- Mrs. Aysha V, College of Applied Science Pattuvam affiliated with Kannur University, India
- Dr. Debotosh Bhattacharjee, Department of Computer Science and Engineering, Jadavpur University, Kolkata-700032, India
- Dr. Neelam Srivastava, Institute of engineering & Technology, Lucknow, India
- Prof. Sweta Verma, Galgotia's College of Engineering & Technology, Greater Noida, India
- Mr. Harminder Singh BIndra, MIMIT, INDIA
- Dr. Lokesh Kumar Sharma, Chhattisgarh Swami Vivekanand Technical University, Bhilai, India
- Mr. Tarun Kumar, U.P. Technical University/Radha Govinend Engg. College, India
- Mr. Tirthraj Rai, Jawahar Lal Nehru University, New Delhi, India
- Mr. Akhilesh Tiwari, Madhav Institute of Technology & Science, India
- Mr. Dakshina Ranjan Kisku, Dr. B. C. Roy Engineering College, WBUT, India
- Ms. Anu Suneja, Maharshi Markandeshwar University, Mullana, Haryana, India
- Mr. Munish Kumar Jindal, Punjabi University Regional Centre, Jaito (Faridkot), India

- Dr. Ashraf Bany Mohammed, Management Information Systems Department, Faculty of Administrative and Financial Sciences, Petra University, Jordan
- Mrs. Jyoti Jain, R.G.P.V. Bhopal, India
- Dr. Lamia Chaari, SFAX University, Tunisia
- Mr. Akhter Raza Syed, Department of Computer Science, University of Karachi, Pakistan
- Prof. Khubaib Ahmed Qureshi, Information Technology Department, HIMS, Hamdard University, Pakistan
- Prof. Boubker Sbihi, Ecole des Sciences de L'Information, Morocco
- Dr. S. M. Riazul Islam, Inha University, South Korea
- Prof. Lokhande S.N., S.R.T.M.University, Nanded (MH), India
- Dr. Vijay H Mankar, Dept. of Electronics, Govt. Polytechnic, Nagpur, India
- Dr. M. Sreedhar Reddy, JNTU, Hyderabad, SSIETW, India
- Mr. Ojesanmi Olusegun, Ajayi Crowther University, Oyo, Nigeria
- Ms. Mamta Juneja, RBIEBT, PTU, India
- Dr. Ekta Walia Bhullar, Maharishi Markandeshwar University, Mullana Ambala (Haryana), India
- Prof. Chandra Mohan, John Bosco Engineering College, India
- Mr. Nitin A. Naik, Yeshwant Mahavidyalaya, Nanded, India
- Mr. Sunil Kashibarao Nayak, Bahirji Smarak Mahavidyalaya, Basmathnagar Dist-Hingoli., India
- Prof. Rakesh.L, Vijetha Institute of Technology, Bangalore, India
- Mr B. M. Patil, Indian Institute of Technology, Roorkee, Uttarakhand, India
- Mr. Thipendra Pal Singh, Sharda University, K.P. III, Greater Noida, Uttar Pradesh, India
- Prof. Chandra Mohan, John Bosco Engg College, India
- Mr. Hadi Saboohi, University of Malaya - Faculty of Computer Science and Information Technology, Malaysia
- Dr. R. Baskaran, Anna University, India
- Dr. Wichian Sittiprapaporn, Mahasarakham University College of Music, Thailand
- Mr. Lai Khin Wee, Universiti Teknologi Malaysia, Malaysia
- Dr. Kamaljit I. Lakhtaria, Atmiya Institute of Technology, India
- Mrs. Inderpreet Kaur, PTU, Jalandhar, India
- Mr. Iqbaldeep Kaur, PTU / RBIEBT, India
- Mrs. Vasudha Bahl, Maharaja Agrasen Institute of Technology, Delhi, India
- Prof. Vinay Uttamrao Kale, P.R.M. Institute of Technology & Research, Badnera, Amravati, Maharashtra, India
- Mr. Suhas J Manangi, Microsoft, India
- Ms. Anna Kuzio, Adam Mickiewicz University, School of English, Poland
- Mr. Vikas Singla, Malout Institute of Management & Information Technology, Malout, Punjab, India, India
- Dr. Dalbir Singh, Faculty of Information Science And Technology, National University of Malaysia, Malaysia
- Dr. Saurabh Mukherjee, PIM, Jiwaji University, Gwalior, M.P, India
- Dr. Debojyoti Mitra, Sir Padampat Singhania University, India
- Prof. Rachit Garg, Department of Computer Science, L K College, India
- Dr. Arun Kumar Gupta, M.S. College, Saharanpur, India
- Dr. Todor Todorov, Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Bulgaria

- Mr. Akhter Raza Syed, University of Karachi, Pakistan
- Mrs. Manjula K A, Kannur University, India
- Prof. M. Saleem Babu, Department of Computer Science and Engineering, Vel Tech University, Chennai, India
- Dr. Rajesh Kumar Tiwari, GLA Institute of Technology, India
- Dr. V. Nagarajan, SMVEC, Pondicherry university, India
- Mr. Rakesh Kumar, Indian Institute of Technology Roorkee, India
- Prof. Amit Verma, PTU/RBIEBT, India
- Mr. Sohan Purohit, University of Massachusetts Lowell, USA
- Mr. Anand Kumar, AMC Engineering College, Bangalore, India
- Dr. Samir Abdelrahman, Computer Science Department, Cairo University, Egypt
- Dr. Rama Prasad V Vaddella, Sree Vidyanikethan Engineering College, India
- Prof. Jyoti Prakash Singh, Academy of Technology, India
- Mr. Peyman Taher, Oklahoma State University, USA
- Dr. S Srinivasan, PDM College of Engineering, India
- Mr. Muhammad Zakarya, CIIT, Pakistan
- Mr. Williamjeet Singh, Chitkara Institute of Engineering and Technology, India
- Mr. G.Jeyakumar, Amrita School of Engineering, India
- Mr. Harmunish Taneja, Maharishi Markandeshwar University, Mullana, Ambala, Haryana, India
- Dr. Sin-Ban Ho, Faculty of IT, Multimedia University, Malaysia
- Mrs. Doreen Hephzibah Miriam, Anna University, Chennai, India
- Mrs. Mitu Dhull, GNKITMS Yamuna Nagar Haryana, India
- Mr. Neetesh Gupta, Technocrats Inst. of Technology, Bhopal, India
- Ms. A. Lavanya, Manipal University, Karnataka, India
- Ms. D. Pravallika, Manipal University, Karnataka, India
- Prof. Ashutosh Kumar Dubey, Assistant Professor, India
- Mr. Ranjit Singh, Apeejay Institute of Management, Jalandhar, India
- Mr. Prasad S.Halgaonkar, MIT, Pune University, India
- Mr. Anand Sharma, MITS, Lakshmanagarh, Sikar (Rajasthan), India
- Mr. Amit Kumar, Jaypee University of Engineering and Technology, India
- Prof. Vasavi Bande, Computer Science and Engineering, Hyderabad Institute of Technology and Management, India
- Dr. Jagdish Lal Raheja, Central Electronics Engineering Research Institute, India
- Mr G. Appasami, Dept. of CSE, Dr. Pauls Engineering College, Anna University - Chennai, India
- Mr Vimal Mishra, U.P. Technical Education, Allahabad, India
- Dr. Arti Arya, PES School of Engineering, Bangalore (under VTU, Belgaum, Karnataka), India
- Mr. Pawan Jindal, J.U.E.T. Guna, M.P., India
- Prof. Santhosh.P.Mathew, Saintgits College of Engineering, Kottayam, India
- Dr. P. K. Suri, Department of Computer Science & Applications, Kurukshetra University, Kurukshetra, India
- Dr. Syed Akhter Hossain, Daffodil International University, Bangladesh
- Mr. Nasim Qaisar, Federal Urdu Univetrstity of Arts , Science and Technology, Pakistan
- Mr. Mohit Jain, Maharaja Surajmal Institute of Technology (Affiliated to Guru Gobind Singh Indraprastha University, New Delhi), India
- Dr. Shaveta Rani, GZS College of Engineering & Technology, India

- Dr. Paramjeet Singh, GZS College of Engineering & Technology, India
- Prof. T Venkat Narayana Rao, Department of CSE, Hyderabad Institute of Technology and Management , India
- Mr. Vikas Gupta, CDLM Government Engineering College, Panniwala Mota, India
- Dr Juan José Martínez Castillo, University of Yacambu, Venezuela
- Mr Kunwar S. Vaisla, Department of Computer Science & Engineering, BCT Kumaon Engineering College, India
- Prof. Manpreet Singh, M. M. Engg. College, M. M. University, Haryana, India
- Mr. Syed Imran, University College Cork, Ireland
- Dr. Namfon Assawamekin, University of the Thai Chamber of Commerce, Thailand
- Dr. Shahaboddin Shamshirband, Islamic Azad University, Iran
- Dr. Mohamed Ali Mahjoub, University of Monastir, Tunisia
- Mr. Adis Medic, Infosys Ltd, Bosnia and Herzegovina
- Mr Swarup Roy, Department of Information Technology, North Eastern Hill University, Umshing, Shillong 793022, Meghalaya, India
- Mr. Suresh Kallam, East China University of Technology, Nanchang, China
- Dr. Mohammed Ali Hussain, Sai Madhavi Institute of Science & Technology, Rajahmundry, India
- Mr. Vikas Gupta, Adesh Institute of Engineering & Technology, India
- Dr. Anuraag Awasthi, JV Womens University, Jaipur, India
- Dr. Dr. Mathura Prasad Thapliyal, Department of Computer Science, HNB Garhwal University (Central University), Srinagar (Garhwal), India
- Mr. Md. Rajibul Islam, Ibnu Sina Institute, University Technology Malaysia, Malaysia
- Mr. Adnan Qureshi, University of Jinan, Shandong, P.R.China, P.R.China
- Dr. Jatinderkumar R. Saini, Narmada College of Computer Application, India
- Mr. Mueen Uddin, Universiti Teknologi Malaysia, Malaysia
- Mr. S. Albert Alexander, Kongu Engineering College, India
- Dr. Shaidah Jusoh, Zarqa Private University, Jordan
- Dr. Dushmanta Mallick, KMBB College of Engineering and Technology, India
- Mr. Santhosh Krishna B.V, Hindustan University, India
- Dr. Tariq Ahamad Ahanger, Kausar College Of Computer Sciences, India
- Dr. Chi Lin, Dalian University of Technology, China
- Prof. VIJENDRA BABU.D, ECE Department, Aarupadai Veedu Institute of Technology, Vinayaka Missions University, India
- Mr. Raj Gaurang Tiwari, Gautam Budh Technical University, India
- Mrs. Jeysree J, SRM University, India
- Dr. C S Reddy, VIT University, India
- Mr. Amit Wason, Rayat-Bahra Institute of Engineering & Bio-Technology, Kharar, India
- Mr. Yousef Naeemi, Mehr Alborz University, Iran
- Mr. Muhammad Shuaib Qureshi, Iqra National University, Peshawar, Pakistan, Pakistan
- Dr Pranam Paul, Narula Institute of Technology Agarpara. Kolkata: 700109; West Bengal, India
- Dr. G. M. Nasira, Sasurie College of Engineering, (Affiliated to Anna University of Technology Coimbatore), India
- Dr. Manasawee Kaenampornpan, Mahasarakham University, Thailand
- Mrs. Iti Mathur, Banasthali University, India
- Mr. Avanish Kumar Singh, RRIMT, NH-24, B.K.T., Lucknow, U.P., India

- Dr. Panagiotis Michailidis, University of Western Macedonia, Greece
- Mr. Amir Seyed Danesh, University of Malaya, Malaysia
- Dr. Terry Walcott, E-Promag Consultancy Group, United Kingdom
- Mr. Farhat Amine, High Institute of Management of Tunis, Tunisia
- Mr. Ali Waqar Azim, COMSATS Institute of Information Technology, Pakistan
- Mr. Zeeshan Qamar, COMSATS Institute of Information Technology, Pakistan
- Dr. Samsudin Wahab, MARA University of Technology, Malaysia
- Mr. Ashikali M. Hasan, CelNet Security, India

# **TABLE OF CONTENTS**

<b>1. The Use of Design Patterns in a Location-Based GPS Application</b> <b>David Gillibrand and Khawar Hameed</b>	<b>1-6</b>
<b>2. An Agent-based Strategy for Deploying Analysis Models into Specification and Design for Distributed APS Systems</b> <b>Luis Antonio de Santa-Eulalia, Sophie D Amours and Jean-Marc Frayret</b>	<b>7-18</b>
<b>3. Facial Expression Classification Based on Multi Artificial Neural Network and Two Dimensional Principal Component Analysis</b> <b>Le Hoang Thai, Tat Quang Phat and Tran Son Hai</b>	<b>19-26</b>
<b>4. Withdrawn</b>	
<b>5. PM2PLS-An Integration of Proxy Mobile IPv6 and MPLS</b> <b>Carlos A Astudillo, Oscar J Calderon and Jesus H Ortiz</b>	<b>38-46</b>
<b>6. Language Identification of Web Pages Based on Improved N-gram Algorithm</b> <b>Chew Yew Choong, Yoshiki Mikami and Robin Lee Nagano</b>	<b>47-58</b>
<b>7. Determining Covers in Combinational Circuits</b> <b>Ljubomir Cvetkovic and Darko Drazic</b>	<b>59-63</b>
<b>8. Higher Order Programming to Mine Knowledge for a Modern Medical Expert System</b> <b>Nittaya Kerdprasop and Kittisak Kerdprasop</b>	<b>64-72</b>
<b>9. A New Proxy Blind Signature Scheme based on ECDLP</b> <b>Daniyal M Alghazzawi, Trigui Mohamed Salim and Syed Hamid Hasan</b>	<b>73-79</b>
<b>10. Web Based Application for Reading Comprehension Skills</b> <b>Samir Zidat and Mahieddine Djoudi</b>	<b>80-87</b>
<b>11. Active Fault Tolerant Control-FTC-Design for Takagi-Sugeno Fuzzy Systems with Weighting Functions Depending on the FTC</b> <b>Atef Khedher, Kamel Ben Othman and Mohamed Benrejeb</b>	<b>88-96</b>
<b>12. Efficient Spatial Data mining using Integrated Genetic Algorithm and ACO</b> <b>K Sankar and V Vankatachalam</b>	<b>97-105</b>
<b>13. Electronic Seal Stamping Based on Group Signature</b> <b>Girija Srikanth</b>	<b>106-112</b>
<b>14. Arithmetic and Frequency Filtering Methods of Pixel-Based Image Fusion Techniques</b> <b>Firouz Abdullah Al-Wassai, N. V. Kalyankar and Ali A Al-Zuky</b>	<b>113-122</b>

<b>15. Using Fuzzy Decision-Making in E-tourism Industry: A Case Study of Shiraz city E-tourism</b> <b>Zohreh Hamedi and Shahram Jafari</b>	<b>123-127</b>
<b>16. A Reliable routing algorithm for Mobile Adhoc Networks based on fuzzy logic</b> <b>Arash Dana, Golnoosh Ghalavand, Azadeh Ghalavand and Fardad Farokhi</b>	<b>128-133</b>
<b>17. A Knowledge Driven Computational Visual Attention Model</b> <b>Joseph Amudha, K P Soman and S Padmakar Reddy</b>	<b>134-140</b>
<b>18. A Frame Work for Frequent Pattern Mining Using Dynamic Function</b> <b>Sunil Joshi, R S Jadon and R C Jain</b>	<b>141-146</b>
<b>19. Decision Support System for Medical Diagnosis Using Data Mining</b> <b>D Senthil Kumar, G Sathyadevi and S Sivanesh</b>	<b>147-153</b>
<b>20. Internet and political communication - Macedonian case</b> <b>Sali Emruli and Miroslav Baca</b>	<b>154-163</b>
<b>21. A Framework for Modelling Software Requirements</b> <b>Dhirendra Pandey, Ugrasen Suman and A K Ramani</b>	<b>164-171</b>
<b>22. 3D Model Retrieval Based on Semantic and Shape Indexes</b> <b>My Abdellah Kassimi and Omar El Beqqali</b>	<b>172-181</b>
<b>23. A Thought Structure for Complex Systems Modeling Based on Modern Cognitive Perspectives</b> <b>Kamal Mirzaie, Mehdi N Fesharaki and Amir Daneshgar</b>	<b>182-187</b>
<b>24. Identification of Priestley-Taylor transpiration Parameters used in TSEB model by Genetic Algorithm</b> <b>Abdelhaq Mouda and Nouredine Alaa</b>	<b>188-197</b>
<b>25. An Approach to Cost Effective Regression Testing in Black-Box Testing Environment</b> <b>Ananda Rao Akepogu and Kiran Kumar J</b>	<b>198-208</b>
<b>26. Normalized Distance Measure-A Measure for Evaluating MLIR Merging Mechanisms</b> <b>Chetana Sidige, Sujatha Pothula, Raju Korra, Madarapu Naresh Kumar and Mukesh Kumar</b>	<b>209-214</b>
<b>27. Brain Extraction and Fuzzy Tissue Segmentation in Cerebral 2D T1-Weighed Magnetic Resonance Images</b> <b>Bouchaib Cherradi, Omar Bouattane, Mohamed Youssfi and Abdelhadi Raihani</b>	<b>215-223</b>
<b>28. A New Round Robin Based Scheduling Algorithm for Operating Systems-Dynamic Quantum Using the Mean Average</b> <b>Abbas Noon, Ali Kalakech and Seifedine Kadry</b>	<b>224-229</b>

<b>29. A Multi-Modal Recognition System Using Face and Speech</b> <b>Samir Akrouf, Belayadi Yahia, Mostefai Messaoud and Youssef Chahir</b>	<b>230-236</b>
<b>30. A Temporal Neuro-Fuzzy Monitoring System to Manufacturing Systems</b> <b>Rafik Mahdaoui, Mouss Leila-Hayet, Mohamed Djamel Mouss and Ouahiba Chouhal</b>	<b>237-246</b>
<b>31. An Efficient Stream Cipher Algorithm for Data Encryption</b> <b>Majid Bakhtiari and Mohd Aizaini Maarof</b>	<b>247-253</b>
<b>32. Rectangular Patch Antenna Performances Improvement Employing Slotted Rectangular shaped for WLAN Applications</b> <b>Mouloud Challal, Arab Azrar and Mokrane Dehmas</b>	<b>254-258</b>
<b>33. Semantic annotation of requirements for automatic UML class diagram generation</b> <b>Soumaya Amdouni, Soumaya Amdouni, Wahiba Ben Abdessalem Karaa and Sondes Bouabid</b>	<b>259-264</b>
<b>34. Blind speech separation based on undecimated wavelet packet-perceptual filterbanks and independent component analysis</b> <b>Ibrahim Missaoui and Zied Lachiri</b>	<b>265-272</b>
<b>35. A Neural Network Model for Construction Projects Site Overhead Cost Estimating in Egypt</b> <b>Ismaail ElSawy, Hossam Hosny and Mohammed Abdel Razek</b>	<b>273-283</b>
<b>36. Time of Matching Reduction and Improvement of Sub-Optimal Image Segmentation for Iris Recognition</b> <b>R M Farouk and G F Elhadi</b>	<b>284-295</b>
<b>37. Recurrent Neural Networks Design by Means of Multi-Objective Genetic Algorithm</b> <b>Hanen Chihi and Najet Arous</b>	<b>296-302</b>
<b>38. Selective Acknowledgement Scheme to Mitigate Routing Misbehavior in Mobile Ad Hoc Network</b> <b>Nimitr Suanmali, Kamalrulnizam Abu Bakar and Suardinata</b>	<b>303-307</b>
<b>39. An Analytical Framework for Multi-Document Summarization</b> <b>J Jayabharathy, S Kanmani and Buvana</b>	<b>308-314</b>
<b>40. Improving Web Page Readability by Plain Language</b> <b>Walayat Hussain, Osama Sohaib and Arif Ali</b>	<b>315-319</b>
<b>41. 2-Jump DNA Search Multiple Pattern Matching Algorithm</b> <b>Raju Bhukya and D V L N Somayajulu</b>	<b>320-329</b>
<b>42. Data Structure and Algorithm for Combination Tree To Generate Test Case</b> <b>Ravi Prakash Verma, Bal Gopal and Md Rizwan Beg</b>	<b>330-333</b>



<b>43. Generation of test cases from software requirements using combination trees</b> <b>Ravi Prakash Verma, Bal Gopal and Md Rizwan Beg</b>	<b>334-340</b>
<b>44. Evolutionary Biclustering of Clickstream Data</b> <b>R Rathipriya, K Thangavel and J Bagyamani</b>	<b>341-347</b>
<b>45. Transmission Power Level Selection Method Based On Binary Search Algorithm for HiLOW</b> <b>Lingeswari V Chandra, Selvakumar Manickam, Kok-Soon Chai and Sureswaran Ramadass</b>	<b>348-353</b>
<b>46. Setting up of an Open Source based Private Cloud</b> <b>G R Karpagam and J Parkavi</b>	<b>354-359</b>
<b>47. Real-Time Strategy Experience Exchanger Model Real-See</b> <b>Mostafa Aref, Magdy Zakaria and Shahenda Sarhan</b>	<b>360-368</b>
<b>48. Sensitivity Analysis of TSEB Model by One-Factor-At-A-Time in irrigated olive orchard</b> <b>Abdelhaq Mouda and Nouredine Alaa</b>	<b>369-377</b>
<b>49. Power Efficient Higher Order Sliding Mode Control of SR Motor for Speed Control Applications</b> <b>Muhammad Rafiq, Saeed-ur-Rehman, Fazal-ur-Rehman and Qarab Raza</b>	<b>378-387</b>
<b>50. Semantic Search in Wiki using HTML5 Microdata for Semantic Annotation</b> <b>P Pabitha, K R Vignesh Nandha Kumar, N Pandurangan, R Vijayakumar and M Rajaram</b>	<b>388-394</b>
<b>51. Formal Verification of Finger Print ATM Transaction through Real Time Constraint Notation RTCN</b> <b>Vivek Kumar Singh, Tripathi S.P, R P Agarwal and Singh J.B.</b>	<b>395-400</b>
<b>52. Self-Destructible Concentrated P2P Botnet</b> <b>Mukesh Kumar, Sujatha Pothula, P Manikandan, Madarapu Naresh Kumar, Chetana Sidige and Sunil Kumar Verma</b>	<b>401-406</b>
<b>53. Fast Overflow Detection in Moduli Set</b> <b>Mehrin Rouhifar, Mehdi Hosseinzadeh, Saeid Bahanfar and Mohammad Teshnehlab</b>	<b>407-414</b>
<b>54. A Novel Feature Selection method for Fault Detection and Diagnosis of Control Valves</b> <b>Binoy B Nair, M T Vamsi Preetam, Vandana R Panicker, V Grishma Kumar and A Tharanya</b>	<b>415-421</b>
<b>55. A Survey on Data Mining and Pattern Recognition Techniques for Soil Data Mining</b> <b>D Ashok Kumar and N Kannathasan</b>	<b>422-428</b>

<b>56. Markov Model for Reliable Packet Delivery in Wireless Sensor Networks</b> <b>Vijay Kumar, R B Patel, Manpreet Singh and Rohit Vaid</b>	<b>429-432</b>
<b>57. Comparative Study of VoIP over WiMax and WiFi</b> <b>M Atif Qureshi, Arjumand Younus, Muhammad Saeed, Farhan Ahmed Siddiqui, Nasir Touheed and M Shahid Qureshi</b>	<b>433-437</b>
<b>58. IBook-Interactive and Semantic Multimedia Content Generation for eLearning</b> <b>Arjumand Younus, M Atif Qureshi, Muhammad Saeed, Syed Asim Ali, Nasir Touheed and M Shahid Qureshi</b>	<b>438-443</b>
<b>59. Applying RFID Technology to construct an Elegant Hospital Environment</b> <b>A Anny Leema and M Hemalatha</b>	<b>444-448</b>
<b>60. Image Compression Using Wavelet Transform Based on the Lifting Scheme and its Implementation</b> <b>A Alice Blessie, J Nalini and S C Ramesh</b>	<b>449-453</b>
<b>61. Incorporating Agent Technology for Enhancing the Effectiveness of E-learning System</b> <b>N Sivakumar, K Vivekanandan, B Arthi, S Sandhya and Veenas Katta</b>	<b>454-461</b>
<b>62. Linear Network Coding on Multi-Mesh of Trees using All to All Broadcast</b> <b>Nitin Rakesh and Vipin Tyagi</b>	<b>462-471</b>
<b>63. Minimization of Call Blocking Probability by Using an Adaptive Heterogeneous Channel Allocation Scheme for Next Generation Wireless Handoff Systems</b> <b>Debabrata Sarddar, Arnab Raha, Shubhajeet Chatterjee, Ramesh Jana, Shaik Sahil Babu, Prabir Kr Naskar, Utpal Biswas and Mrinal Kanti Naskar</b>	<b>472-477</b>
<b>64. On-Demand Multicasting in Ad-hoc Networks-Performance Evaluation of AODV, ODMRP and FSR</b> <b>M Rajendiran</b>	<b>478-482</b>
<b>65. Enhanced Stereo Matching Technique using Image Gradient for Improved Search Time</b> <b>Pratibha Vellanki and Madhuri Khambete</b>	<b>483-486</b>
<b>66. Analyzing the Impact of Scalability on QoS-aware Routing for MANETs</b> <b>Rajneesh Kumar Gujral and Manpreet Singh</b>	<b>487-495</b>
<b>67. Improving Data Association Based on Finding Optimum Innovation Applied to Nearest Neighbor for Multi-Target Tracking in Dense Clutter Environment</b> <b>E M Saad, El Bardawiny, H I Ali and N M Shawky</b>	<b>496-507</b>
<b>68. An Efficient Quality of Service Based Routing Protocol for Mobile Ad Hoc Networks</b> <b>Tapan Kumar Godder, M. M Hossain, M Mahbubur Rahman and Md. Sipon Mia</b>	<b>508-514</b>

<b>69. SEWOS-Bringing Semantics into Web operating System</b> <b>A.M. Riad, Hamdy K Elminir, Mohamed Abu ElSoud and Sahar F Sabbeh</b>	<b>515-521</b>
<b>70. Segmenting and Hiding Data Randomly Based on Index Channel</b> <b>Emad T Khalaf and Norrozila Sulaiman</b>	<b>522-529</b>
<b>71. Data-Acquisition Data Analysis and Prediction Model for Share Market</b> <b>Harsh Shah and Sukhada Bhingarkar</b>	<b>530-534</b>
<b>72. Fast Handoff Implementation by using Curve Fitting Equation With Help of GPS</b> <b>Debabrata Sarddar, Shubhajeet Chatterjee, Ramesh Jana, Shaik Sahil Babu, Hari Narayan Khan, Utpal Biswas and Mrinal Kanti Naskar</b>	<b>535-542</b>
<b>73. Visual Cryptography Scheme for Color Image Using Random Number with Enveloping by Digital Watermarking</b> <b>Shyamalendu Kandar, Arnab Maiti and Bibhas Chandra Dhara</b>	<b>543-549</b>
<b>74. Computation of Multiple Paths in MANETs Using Node Disjoint Method</b> <b>M Nagaratna, P V S Srinivas, V Kamakshi Prasad and C Raghavendra Rao</b>	<b>550-554</b>
<b>75. WLAN Security-Active Attack of WLAN Secure Network</b> <b>Anil Kumar Singh and Bharat Mishra</b>	<b>555-559</b>
<b>76. Mining databases on World Wide Web</b> <b>Manali Gupta, Vivek Tomar, Jaya Verma and Sudeepa Roy</b>	<b>560-564</b>
<b>77. Performance Analysis of IEEE 802.11 Non-Saturated DCF</b> <b>Bhanu Prakash Battula, R Satya Prasad and Mohammed Moulana</b>	<b>565-568</b>
<b>78. Enhancing the Capability of N-Dimension Self-Organizing Petrinet using Neuro-Genetic Approach</b> <b>Manuj Darbari, Rishi Asthana, Hasan Ahmed and Neelu Jyoti Ahuja</b>	<b>569-571</b>
<b>79. Vulnerabilities of Electronics Communication: solution mechanism through script</b> <b>Arun Kumar Singh, Pooja Tewari, Shefalika Ghosh Samaddar and Arun K Misra</b>	<b>572-582</b>
<b>80. Image Registration in Digital Images for Variability in VEP</b> <b>N Sivanandan and N J R Muniraj</b>	<b>583-587</b>
<b>81. WiMAX-Worldwide Interoperability for Microwave Access-A Broadband Wireless Product in Emerging</b> <b>Komal Chandra Joshi and M P Thapliyal</b>	<b>588-591</b>
<b>82. Simulation and Optimization of MQW based optical modulator for on chip optical interconnect</b> <b>Sumita Mishra, Naresh K Chaudhary and Kalyan Singh</b>	<b>592-596</b>

**83. Determination of the Complex Dielectric Permittivity Industrial Materials of the Adhesive Products for the Modeling of an Electromagnetic Field at the Level of a Glue Joint** **597-601**  
**Mahmoud Abbas and Mohammad Ayache**

**84. Power Aware Routing in Wireless Sensor Network** **602-610**  
**Rajesh Kumar Sahoo, Satyabrata Das, Durga Prasad Mohapatra and Manas Ranjan Patra**

# The Use of Design Patterns in a Location-Based GPS Application

David Gillibrand and Khawar Hameed

Staffordshire University,  
The Octagon, Beaconside, Stafford, ST18 0AD

## Abstract

The development of location-based systems and applications presents a number of challenges – including those of designing and developing for a range of heterogeneous mobile device types, the associated spectrum of programming languages, and the incorporation of spatial concepts into applied software solutions. This paper addresses these challenges by presenting a harmonised approach to the construction of GPS location-based applications that is based on Design Patterns. The context of location-based systems is presented, followed by several design patterns - including the Observer and Bridge Design Patterns, which are described and applied to the application. Finally the benefits of using Design Patterns in this framework-oriented approach are discussed and future related work in the area of systems design for mobile applications is outlined.

**Keywords:** *Reusable software, Object Orientation, Design Patterns, Mobile Applications, Location-Based Systems.*

## 1. Introduction

The idea of design patterns came from a building architect Christopher Alexander who wrote a book “A Pattern Language: Towns, Buildings, Construction” in 1977 [1]. The idea of what a pattern is, is summed up in the following quote by Alexander: “Each pattern describes a problem which occurs over and over again in our environment, and then describes the core of the solution to that problem, in such a way that you can use this solution a million times over, without ever doing it the same way twice.” Design Patterns have been used in object-oriented design since the widely acclaimed book *Design Patterns: Elements of Reusable Object-Oriented Software* by Gamma, Helm, Johnson and Vlissides [2] commonly known as the Gang of Four. It contains a description of the concepts of patterns, plus a catalog of 23 design patterns with their full documentation. They took Alexanders’ idea of a pattern and applied it to software Engineering. Essentially a software design pattern is a piece of literature that describes a well tried out solution to a given problem in a given context. As part of that description the essential elements are the pattern name, problem description, problem solution and any consequences of using the pattern. The benefit of using

patterns is to make the software more reusable - by building in quality attributes of software such as extensibility, maintainability and readability. The principle of design patterns therefore provides a basis for the development of specific schema based on systemic rigour at a low level of abstraction in systems modeling essentially providing a framework to encapsulate what are potentially complex application domains and the associated interactions between system components. These affordances provide the motivation for our paper that of conceptualising, capturing and modelling location-based systems and dynamic spatial data associated with the design and development of location-based applications -such as those used in navigation systems, using the relatively formal construct of design patterns.

## 2. Location-Based Systems and GPS

Having introduced the subject of design patterns and the motivation, the application domain of location-based systems is presented. This provides the basis for the subsequent articulation of design patterns and aims to demonstrate the transformation of a somewhat conceptual and abstract notion of space into an applied schema constructed using design patterns. These provide the foundation for programmers to construct code that underpins location-based applications.

Location-based systems are those which exploit and leverage the concept of mobility in context of local or remote environmental conditions and factors, and are founded on the core principle of *anyplace* as the driving rationale. Essentially, location-based systems deliver information that is relevant to users in context of their location at any particular point in time and where the focus or contextualisation of information and services is governed by location [3][4]. Furthermore, this information and the associated services can be defined as triggered or user-requested [4]. Developments in location-based systems have been driven by regulatory requirements such as international legislation and a growing awareness of the commercial opportunities facilitated by exploiting the technical ability to provide

value-added information and enhanced experience to mobile consumers, and through emerging demand levels, typical services and associated business models [5]. Location-based systems are strongly coupled to the concept of context within mobile computing systems and form a special class of context-aware systems [6]. Location is a determinant in that it contributes significantly to the universe of discourse created - essentially, all activities are hosted within a particular environmental location and context. A key characteristic of location-based systems is the changing physical location of the mobile user, which may be continual - such as when in a moving vehicle or walking, or periodic - where there are periods of short-term or transient residency of the user in a location.

Kakihara and Sorensen [7] discuss the view of spatial mobility as one dimension of mobility that is, 'the most immediate aspect of mobility' in that the physical locational space provides the immersive context for objects within that space. This discussion further articulates three composite aspects - that of the mobility of objects, the mobility of symbols, and the mobility of space itself. The dimensional aspects of location lead to a potentially more complex universe of discourse comprising the determination of object positioning within space (for example using co-ordinate geometry) where location identification is not only based on triangulation of co-ordinates - where each co-ordinate represents a particular dimension, but also based on time - where objects move through space and time. In this case, objects can be deemed to possess an orthogonal property where *different locations in time* exist for those objects.

The transformation of these spatial concepts into real-world deployments of location-based systems is also evident. Within the public sector there is acknowledgement of the unique features and advantages of mobile technologies to enhance engagement between governmental institutions and the citizens they serve through the development of innovative location-based services and new methods of interaction [8]. Specific examples of mobile location-based applications include those concerned with supporting front-line emergency services for public security and safety [6] with a range of associated improvement and efficiency gains being reported.

Private sector interest in mobile location-based systems is underpinned by new commercial and revenue-generating opportunities evidenced primarily by numerous consumer-oriented applications in a number of categories including navigation and travel, social networking, leisure and entertainment. For example, in September 2009 Apple,

through its mobile application distribution channel - the 'App Store', had twenty different categories of mobile applications with 268 applications within the 'travel' category, and a large number of applications across all categories that exploited the user's location profile as part of the application configuration.

Supporting technology for location-based systems typically includes mobile network platforms for determining location including wide area systems such as Cell Identification in mobile radio networks, Global Positioning Systems (GPS), and Broadband Satellites [5] and more localised sensor technologies such as WiFi (802.11), Bluetooth, and Radio Frequency Identification (RFID) [9]. Spatial databases provide the core repository infrastructure to host multi-dimensional data, with associated data models and query capabilities that enable location-based queries to be satisfied. The end-to-end delivery of mobile location-based systems includes a number of stakeholders, each of which is critical to the operation of the complete system. These include mobile network operators, content providers and aggregators, technology infrastructure providers, application service providers, and device manufacturers. As the potential scope and opportunities offered by mobile location-based systems increase, there is a risk of increasing complexity leading to evaluation of suitable business models and frameworks and components that address the overall aggregation of services [10][11]. Whilst appreciating this increasing complexity at higher levels of abstraction in location-based systems, we posit that an equal focus and effort on the use of design patterns to formalize and structure the lower-level construction of such systems is of merit.

The remainder of this section introduces the example of a GPS application as a component of a location-based system. This illustration is subsequently developed and serves a vehicle for the articulation of the associated design patterns.

The GPS application consists of reading data from a GPS receiver which constantly sends a stream of \$GPRMC sentences to a GPS class. An example of a sentence is: \$GPRMC, 140036,A, 5226.5059, N, 00207.6806, W,2.0, 064.64, 120710,001.0,E\*34 where 194322 is the time of fix (14:00:36 UTC), A is a navigation receiver warning (A = OK, V = warning), 5226.5059,N is Latitude (52 deg. 26.5059 min North), 00207.6806,W is Longitude (002 deg. 07.6806 min West), 2.0 is Speed over ground (Knots), 064.64 is Course Made Good( degrees), 120710 is Date of fix (12 July 2010), 001.0,E is the Magnetic variation (1.0 deg East), \*34 is the mandatory checksum.



The application then continually reads, parses and stores the sentences as records in a buffer. Those records are then available for the application to read. The application displays three views, a text view which as well as displaying the basic information in the GPS sentence also displays the distance travelled and average speed, a compass which uses the course made good part of the sentence (the user needs to be travelling at about 3knots for this to display a meaningful value) and a breadcrumb trail which shows the trail as well as minimum height, maximum height and the ascent (the difference between them) see Figure 1.

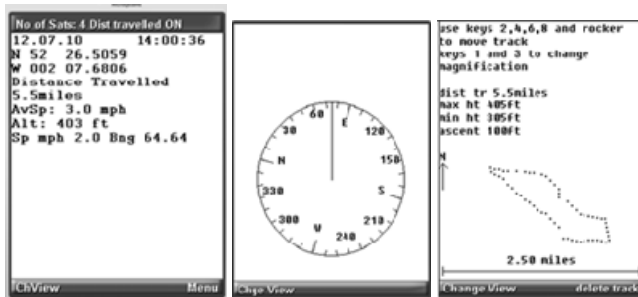


Fig. 1. Different views from a GPS application

### 3. Design Patterns Used in a GPS Application

The principle design patterns used are the Observer, Strategy and Bridge design pattern. What follows is a description of those patterns followed by an explanation of how those patterns can be applied to the GPS application.

#### 3.1 Observer Design Pattern

The intent of the pattern as described by Gamma et al. is to define a one-to-many dependency between objects so that when one object state changes all the other objects are notified and change state accordingly. The observer pattern can be applied in any of the following situations: When an abstraction has two aspects, one dependent on the others, encapsulating these aspects in separate objects lets you vary and reuse them independently; when a change to one object requires changing others, and you don't know how many objects need to be changed; When an object should be able to notify others without making assumptions about who these objects are [2]. Figure 2 shows a class diagram that represents the general case of the observer design pattern.

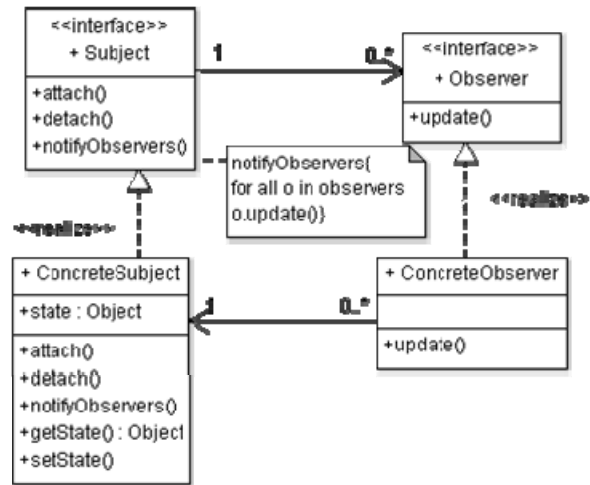


Fig. 2. The Observer Design Pattern

The ConcreteSubject class contains the data (state) with associated get and set methods and a list of observers. There are methods to add or delete an observer, (attach and detach). The notifyObservers method iterates through the observers list in the ConcreteSubject class invoking the update method in each observer object. The update method in the Observer object then gets the state (or data) of the Subject object. The Observer pattern can be varied with respect to the update protocol. The Pull model protocol (which we've just described) can be implemented in java by sending a changeEvent object to the observers (views) every time the data or state is changed in the Subject Object. On receiving this object the views obtain the latest data from the subject by invoking the update method. Alternatively there is the Push model protocol - when data is changed, the Subject sends a message to the Observers saying that the data has changed and also sends an additional argument that describes the state change. This comes in as an argument to the update method, so the observer has already got the latest data without having to invoke the getState method of the Subject.

#### 3.2 Strategy Design Pattern

The Strategy Pattern “defines a family of algorithms, encapsulates each one, and makes them interchangeable. The Strategy pattern lets the algorithm vary independently from clients that use it” [12]. Instead of using case statements to differentiate between different algorithms, a more flexible design is to encapsulate each transformation algorithm as a separate class. The class diagram for Strategy is shown in Figure 3.

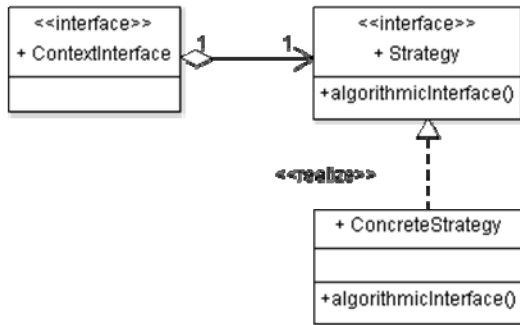


Fig. 3. The Strategy Design Pattern

### 3.3 Bridge Design Pattern

A more flexible alternative to the strategy design pattern is the Bridge design pattern which allows you to vary the abstractions as well as the implementation by placing them into two different class hierarchies - see Figure 4.

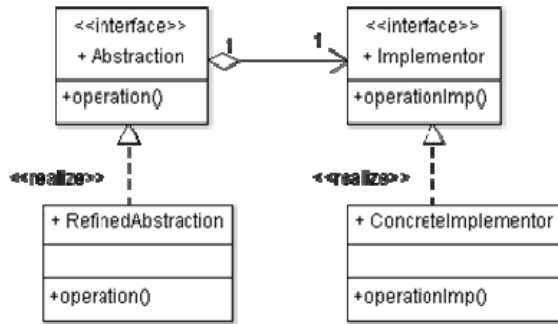


Fig 4. The Bridge Design Pattern

## 4. Applying Patterns to a GPS Application

### 4.1 Application of the Observer Design Pattern

The GPS application lends itself to the observer design pattern. The Subject or the data of the application is encapsulated by a Record class. The Record class has attributes that reflect the information contained in the GPS sentence. The three views (Figure 1) need to be updated every time the Record object changes its values. Figure 5 shows how the observer design pattern has been incorporated into the GPS application.

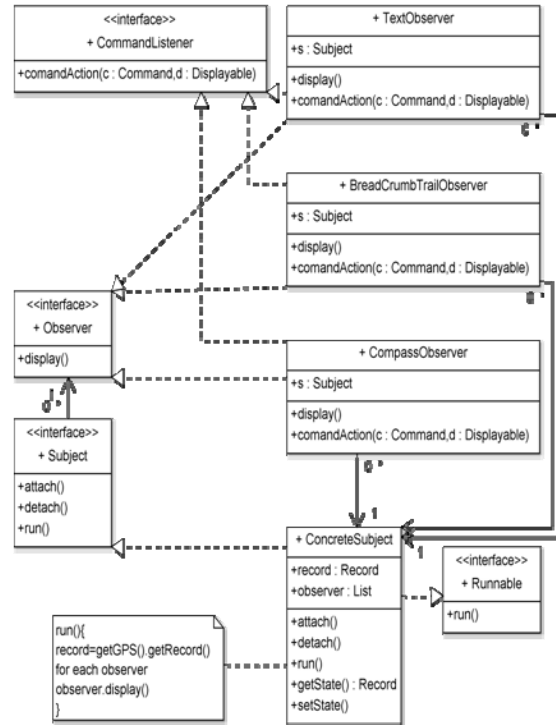


Fig. 5. The Observer Design Pattern Applied to a GPS Application

The ConcreteSubject has a List of the three observers (the views as shown in Figure 1). It also contains the state (record). This record is constantly updated from the GPSreceiver. The views or observers need to be constantly refreshed with the current value of the record. The ConcreteSubject invokes its own thread of execution (it implements the Runnable interface) with the run method. The run method takes the place of the notifyObservers method in the general case (Figure 2). The run method iterates through the three observers and invokes their display method. The observers have an object reference to the Subject passed to them in their constructor method and are then able to invoke the getState method of the Subject. This is essentially using the Pull model protocol. The use of the Observer design pattern allows you to easily extend the application with further views if required and very little modification to the existing code making it a modular and robust design. It decouples the data from the views. The Observer design pattern is the principle one used in MVC (model view controller) architecture. The model is the data of the application, the views are different views of the data and are responsible for drawing that data on the screen and the control is responsible for handling the user input and then updating the model or the view. In our



application the function of the control is split between the Subject which gets the record from a GPS class (record=getGPS().getRecord()) and the Observers which implement the CommandListener interface which can then receive user input commands .

The application also has the ability to change the position format. It can be changed to the OSGrid reference system which is the British national grid system, instead of lat and long, in which case the position format is displayed in terms of northings and eastings as found on an Ordnance Survey map - see figure 6.



Fig. 6 A view showing the British National Grid reference system

## 4.2 Application of the Bridge Design Pattern

The calculation that transforms from a lat/long coordinate system to the OSgrid coordinate system is quite complicated and, in the future, when extending the application it might be deemed appropriate to include other geographical grid systems for transformation. One way of implementing such a functional request might be to use a case statement and passing in a value which reflects the grid system to be chosen. However, this design is inflexible and would involve a rewrite of the class that contained the case statement for each change made. A better way would be to use the Bridge design pattern. In the GPS application, this is applied by creating a new class for each transformation algorithm - see Figure 7 making that part of the application easily extendable.

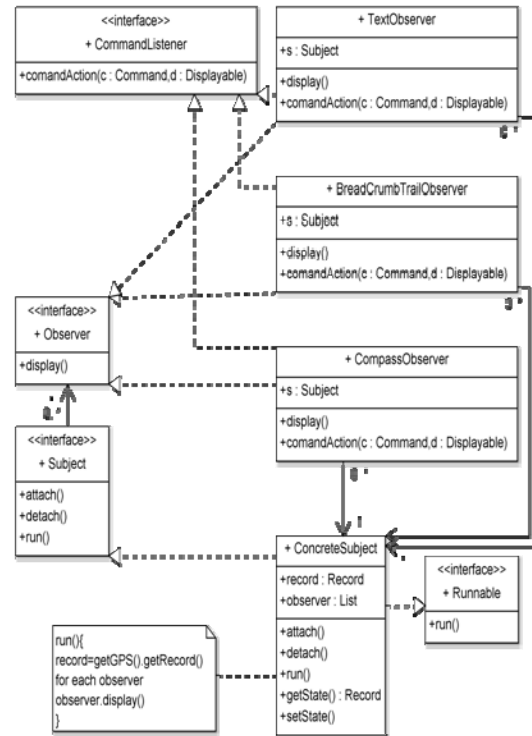


Fig. 7. The Bridge Design Pattern applied to a GPS application

## 5. Conclusions

The use of design patterns brings many benefits in the software development process, especially in terms of code reuse and maintainability. Much of the software that results offers the desirable properties of high cohesion and low coupling. Many of the patterns are well documented and categorised according to different criteria - for example the Gang of Four categorises its patterns by scope (Object or Class) and their purpose (creational, structural or behavioural). The key to successfully using design patterns is to learn and understand the pattern and to classify it in a meaningful way. Once understood, using patterns become second nature and if well documented allows a design level of abstraction to be visible with the added benefits of improved communication between developers.

The use of design patterns to underpin the development of application-specific components of location-based system warrants particular attention. The increasing focus on the development and adoption of location-based systems provided fertile ground for developing schematic and reusable constructs that provide a vehicle for capturing conceptual aspects of location and the translation of these into applied and usable notation for system developers. Our focus on design patterns for location-based systems

has also been incorporated into a Masters level course in Mobile Applications and Systems where students are taught the conceptual basis of mobility and location-based systems, and the associated development of software underpinned by design patterns. The approach to design patterns in this paper can be adopted by both scholars and practitioners in industry. Our future work in this area is concerned with development of schematic constructs to model the multi-dimensional aspects of mobility – those of time, space, and context and to position these at different levels of abstraction within the system development process - such as design patterns at lower levels of abstraction, and enterprise architecture-based constructs at higher levels of abstraction. In doing so, we aim to focus on a systemic and structured approach to the development of mobile applications and systems.

## References

- [1] C. Alexander Pattern Language: Towns, Buildings, Construction. 1977
- [2] Gamma , Helm, Johnson, Vlissides, Design Patterns: Elements of Reusable Object-Oriented Software, Addison-Wesley 1994
- [3] Duri, S., Cole, A., Munson, J. & Christensen, J., 2001, WMC '01: Proceedings of the 1st international workshop on Mobile commerce, An approach to providing a seamless end- user experience for location-aware applications. ACM, pp. 20-5.
- [4] D'Roza, T. & Bilchev, G., 2003, An overview of location-based services, BT Technology Journal, 21(1), pp. 20-7.
- [5] Rao, B. & Minakakis, L., 2003, Evolution of mobile location-based services, Commun. ACM, 46(12), pp. 61-5.
- [6] Streefkerk, J.W., van Esch-Bussemaekers, M.P. & Neerincx, M.A., 2008, MobileHCI '08: Proceedings of the 10th international conference on Human computer interaction with mobile devices and services, Field evaluation of a mobile location-based notification system for police officers. ACM, pp. 101-8.
- [7] Kakihara, M. & Sørensen, C., 2001, Expanding the 'mobility' concept, SIGGROUP Bull., 22(3), pp. 33-7. Trimi, & Sheng 2008
- [8] Trimi, S. & Sheng, H., 2008, Emerging trends in M-government, Commun. ACM, 51(5), pp. 53-8.
- [9] Johnson, S., 2007, A framework for mobile context-aware applications, BT Technology Journal, 25(2), pp. 106-11.
- [10] Aphrodite & Evaggelia, 2001, Business models and transactions in mobile electronic commerce: requirements and properties, Computer Networks, 37(2), pp. 221-36.
- [11] de Reuver, M. & Haaker, T., 2009, Designing viable business models for context-aware mobile services, Telematics and Informatics, 26(3), pp. 240-8.
- [12] Freeman, Freeman, Sierra, Bates Head First Design Patterns O'Reilly 2004

**David Gillibrand** is a Senior Lecturer in the Faculty of Computing, Engineering & Technology at Staffordshire University. His research is in the area of Object-Oriented technologies, Design Patterns, Enterprise Applications, Mobile programming, Databases, and System methods. He has had publications in object-oriented journals and delivered courses in system design to industry.

**Khawar Hameed** is a Principal Lecturer in the Faculty of Computing, Engineering & Technology at Staffordshire University. His research is in the area of mobile and remote working, enterprise mobility, and mobile learning. He has been a key driver in the adoption of mobile computing and technology within the Faculty's portfolio and has helped drive the development of undergraduate and post-graduate degrees in this technology area. He has contributed extensively to the development and delivery of externally funded projects and academic-industrial collaborations in mobile/wireless technology that aim to develop and enhance the collective intellectual capital that supports the growth of mobile and wireless systems as a discipline both within academia and in industry.

# An Agent-based Strategy for Deploying Analysis Models into Specification and Design for Distributed APS Systems

Luis Antonio de Santa-Eulalia<sup>1</sup>, Sophie D'Amours<sup>2</sup> and Jean-Marc Frayret<sup>3</sup>

<sup>1</sup> Téléuq, Université du Québec à Montréal  
Québec City, Québec, Canada

<sup>2</sup> Université Laval  
Québec City, Québec, Canada

<sup>3</sup> École Polytechnique de Montréal  
Montréal, Québec, Canada

## Abstract

Despite the extensive use of the agent technology in the Supply Chain Management field, its integration with Advanced Planning and Scheduling (APS) tools still represents a promising field with several open research questions. Specifically, the literature falls short in providing an integrated framework to analyze, specify, design and implement simulation experiments covering the whole simulation cycle. Thus, this paper proposes an agent-based strategy to convert the 'analysis' models into 'specification' and 'design' models combining two existing methodologies proposed in the literature. The first one is a recent and unique approach dedicated to the 'analysis' of agent-based APS systems. The second one is a well-established methodological framework to 'specify' and 'design' agent-based supply chain systems. The proposed conversion strategy is original and is the first one allowing simulation analysts to integrate the whole simulation development process in the domain of distributed APS.

**Keywords:** *Advanced Planning and Scheduling (APS), Agent-Based Simulation, Methodological Framework, Analysis, Specification and Design, FAMASS.*

## 1. Introduction

Advanced Planning and Scheduling (APS) systems comprise a set of techniques for the supply chain planning over short, intermediate, and long-term time periods. They employ advanced mathematical algorithms or logic to perform optimization or simulation on finite capacity scheduling, sourcing, capital planning, resource planning, forecasting, demand management, and other. APS simultaneously considers a range of constraints and business rules to provide real-time planning and

scheduling, decision support, available-to-promise, and capable-to-promise capabilities. In addition, these systems often generate and evaluate multiple 'what-if' scenarios [1].

The use of these sophisticated optimization approaches in complex real-life supply chain situations has recently become possible mainly due to the increased computing power of companies [2].

Despite the contribution of APS systems to the supply chain planning domain, some criticism exists in this area [3]. Traditional APSs are basically monolithic systems that cannot model and take into account the complex everyday interactions and information exchanges between partners. For example, APS systems are deficient in handling sophisticated interaction mechanisms that allow the implementation of delegation and coordination approaches, which are methodologies based on negotiation, and cooperation strategies [4, 5]. As a result, the focus on relationships in a multi-tier environment has only recently been claimed by the APS community [6].

To cope with this problem, recent advances in supply chain planning have arisen in the area of agent technology. This technology is able to capture the distributed nature of supply chain entities (e.g. customers, manufacturers, logistics operators etc.) and mimic their business behaviours (e.g. making advanced production decisions and negotiating with other supply chain members), thus supporting their collaborative planning process. Because of these abilities, among several others described in the

literature, agent-based supply chain systems have great potential for simulating complex and realistic scenarios [7, 4; 9, 10, 11]. Distributed APS systems employing agent-based technology are referred to in this paper as distributed APS systems [12].

Distributed APS systems are normally developed through the use of modelling and simulation frameworks and, usually, these frameworks provide principles, steps, methods and tools for creating a model. They help people understand the simulation problem to be modelled and translate it into a computing model normally used in simulation experiments in the supply chain planning area.

In order to create such models, these frameworks guide simulation modellers through one or several development steps [13]. The first modelling step is *analysis*, where one defines an abstract description of the modelled supply chain planning system containing functional and non-functional requirements. Next, during *specification*, the information derived from the analysis is translated into a formal model. As the analysis phase does not necessarily allow obtaining a formal model, the specification examines the analysis requirements and builds a model based on a formal approach. After, in the *design* phase one creates a data-processing model that describes the specification model in more detail. In the case of an agent-based system, design models are close to how agents operate. Finally, during *implementation*, the design model is translated into a specific software platform or it is programmed [13].

The problem behind these modelling frameworks is that normally simulation systems are implemented as directed by pre-stated requirements with little explicit focus on system analysis, specification, design and implementation in an integrated manner [14]. According to a recent literature review [15], to the best of our knowledge there are no integrated modelling approaches covering the whole developed process in this area. Moreover, there is one unique ‘analysis’ modelling, the FAMASS (*FORAC Architecture for Modelling Agent-based Simulations for Supply chain planning*) framework, dedicated to the distributed APS domain, and which was proposed by us recently [21, 22, 23].

Despite its contribution to the literature, FAMASS is limited to the identification and mapping of functional requirements of distributed APS simulations, i.e. the ‘analysis’ phase only. If the simulation analysts desire to go further in the modelling process, they have to employ another ‘specification’ and ‘design’ methodology. This can be laborious, since analysts need to thoroughly master FAMASS and another methodology.

In order to facilitate FAMASS analysts in converting their analysis models into specification and design models, this paper proposes an agent-based deployment strategy. This strategy enlarges the FAMASS scope to the other modelling phases, thus covering the entire modelling cycle. By doing so, analysis can go smoother and quicker through this cycle.

To do so, we were inspired by the specification and design principles of the Labarthe et al. [9] framework, a recent and largely cited development in the field of methodological agent-oriented framework for supply chain management simulation. Since the focus of this framework is on supply chain management as a general concept (and not specialized in APS systems), we had to perform some minor adaptations to this approach. Despite these adaptations, the main ideas of Labarthe et al. [9] are explicitly considered in the deployment strategy. The Labarthe et al. framework is adopted here because it covers the specification and design phases properly at the business and agent levels, just as FAMASS does, which facilitates the deployment process.

This deployment strategy demonstrates that the analysis phase of FAMASS can be integrated with other existing approaches specialized in specification and design modelling. Furthermore, it allows us to avoid the research effort needed to develop a totally new specification and design methodology for the domain, although it would be suitable (and even desirable) for future research initiatives.

This paper is organized as follows: a literature review in modelling and simulation for distributed APS systems is presented in Section 2. Section 3 introduces the FAMASS approach, while Section 4 summarizes the Labarthe et al. [9] framework. Next, the deployment process is explained in Section 5. Finally, Section 6 outlines some final remarks and suggests future work.

## 2. Modelling and Simulation Frameworks for distributed APS

The use of agent technology in Supply Chain Management is a fruitful field. From the inaugural work of Fox et al. [16] until today, a large variety of works have appeared to propose different ways of encapsulating supply chain entities and performing simulation experiments.

Two types of modelling approaches can be identified in the literature. The first type proposes generic approaches for modelling agent-based supply chain systems in general terms, while the second type proposes a modelling framework that specifically takes into consideration Advanced Planning and Scheduling (APS) tools when

planning, i.e. the incorporate optimization procedures or finite capacity planning models when performing supply chain planning. APS systems emerged in the last decade to provide a suite of planning and scheduling modules for the firm's internal supply chain, from the raw materials source to the consumers and covering decisions ranging from the strategic to the operational level [17].

In the first type of approach (general agent-based models), examples of relevant contributions include Labarthe et al. [9], Van der Zee, and Van der Vorst [18], MaMA-S [13]. One of the most cited works in the domain is Labarthe et al. [9], which propose a methodological framework for modelling customer-centric supply chains in the context of mass customization. They define a conceptual model for supply chain modelling and show how the multi-agent system can be implemented using predefined agent platforms. Van der Zee and Van der Vorst [18] propose an agent framework derived from an object-oriented approach to explicitly model control structures of supply chains. MaMA-S [13] provides a multi-agent methodology for a distributed industrial system, which is divided into five main phases and two support phases. The authors propose formal methods for the specification, design and implementation phases, but the analysis phase is not tackled by them.

This second type of modelling approach provides more sophisticated models of supply chains by incorporating Advanced Planning and Scheduling routines [12]. These approaches, sometimes called d-APS systems (for distributed APS), are composed of semi-autonomous APS tools, each dedicated to a specialized planning area and that can act together in a collaborative manner employing sophisticated interaction schemas.

Examples of this kind of work are Egri et al. [19], Lendermann et al. [20] and Swaminathan et al. [11]. Egri et al. [19] is a Gaia-based approach for modelling advanced distributed supply chain planning for mass customization. They develop a model for representing roles and interactions of agents based on the SCOR (Supply-Chain Operations Reference) model. Lendermann et al. [20] developed an approach to couple discrete-event simulation and APS for collaborative supply chain optimization, based on the HLA (High Level Architecture) technology for distributed simulation synchronization. Swaminathan et al. [11] provide a supply chain modelling framework containing a library of modular and reusable software components, which represents different kinds of supply chain agents, their constituent control elements and their interaction protocols.

These simulation and modelling approaches have greatly contributed to the domain, however, in spite of these

advances, there exists a relevant gap in this field related to the initial developing step of such simulation systems, the analysis phase [12]. Most of the researched works in the literature suggest approaches for specification and design, and some for implementation, but the analysis phase is not explicitly treated [12, 13, 14, 21]. Most of these works suppose that the analysis phase furnishes the necessary information and concentrate their discussions on further phases, mainly specification and design. The first work dedicated to the analysis of distributed APS systems using the agent-based paradigm is FAMASS [21]. Despite its contribution to the agent-based modelling of distributed APS systems, FAMASS does not cover the specification and design phases of the development process. This is an interesting research gap in the literature. Section 3 details the FAMASS approach for the analysis phase, while Section 4 presents a frequently cited method for specification and design of agent-based supply chain systems from Labarthe et al. [9]. Next, Section 5 combines these two approaches in order to create a deployment strategy to translate analysis models into specification and design.

### 3. The FAMASS Approach

The FAMASS (*FORAC Architecture for Modelling Agent-based Simulation for Supply chain planning*) is the first and unique modelling approach dedicated to the analysis phase of distributed APS simulations [21, 22, 23]. This approach was recently tested in Santa-Eulalia et al. [24].

It is organized into two abstraction levels: Supply chain: refers to the supply chain planning problem, i.e. the business viewpoint; Agent: the supply chain domain problem is translated into an agent-based view (Figure 1).

At these two abstraction levels, four modelling approaches are proposed, namely the General Problem Analysis (GPA), the Distributed Planning Analysis (DPA), the Social Agent Organization Analysis (SAOA) and the Individual Agent Organization Analysis (IAOA), as schematized in Fig. 1.

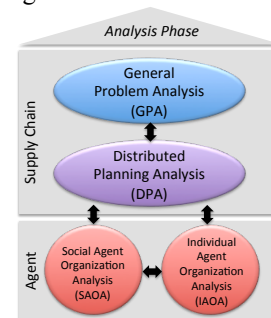


Fig. 1: Four main modelling approaches proposed for analysis of supply chain and agent levels [23].



These four modelling approaches are explained in the following subsections.

### 3.1 General Problem Analysis (GPA)

GPA is the first modelling effort where simulation analysts have to think about the simulation problems. The GPA is based on Santa-Eulalia et al. [12], in which a discussion about the simulation objective and the problem structure is provided.

Basically, the GPA proposes that the simulation analysis has to take two main aspects into consideration: general aspects and experimental aspects. General aspects represent macro definitions of the simulation problem, including the object and environment to be simulated, the simulation questions, hypotheses and objectives. Experimental aspects are related to the design of experiments, where one defines the factors, uncertainties and key performance indicators of the simulation.

These elements refer to the general definition of the simulation problem, according to what is desired to be studied, and it will guide the whole development process.

This general definition is then organized through some formalisms from SysML (Systems Modeling Language) [25]. In this case, some Requirements Diagrams help the analysts organize the GPA. An example of how this can be done is provided in [23].

### 3.2 Distributed Planning Analysis (DPA)

The DPA identifies what the desired supply chain planning entities are, as well as their roles. These entities are identified according to their mission in the supply chain and their planning functions at different decision levels.

To identify the main supply chain planning entities, FAMASS employs the concepts of supply chain integration proposed by Shapiro [26]. The author states that supply chain management refers to integrated planning relying on three basic dimensions: i) *Intertemporal dimension*: refers to different decision levels, i.e. strategic, tactical and operational decision levels; ii) *Functional dimension*: stands for different planning functions in a supply chain, which can be related to procurement, manufacturing, distribution and sales; iii) *Spatial dimension*: refers to the fact that supply chains are composed of geographically dispersed units of analysis.

This gives rise to the notion of a Supply Chain Block. A Supply Chain Block can be defined as a supply chain planning entity, which is a functional unit capable of: performing part of the supply chain planning decisions or their totality; or performing the execution of the supply chain decisions (part of them or their totality). These entities have a certain degree of autonomy and are able to interact with each other. Possible Supply Chain Blocks for covering the integrated supply chain planning dimensions

are proposed in the framework of Fig. 2, which is called the supply chain planning cube.

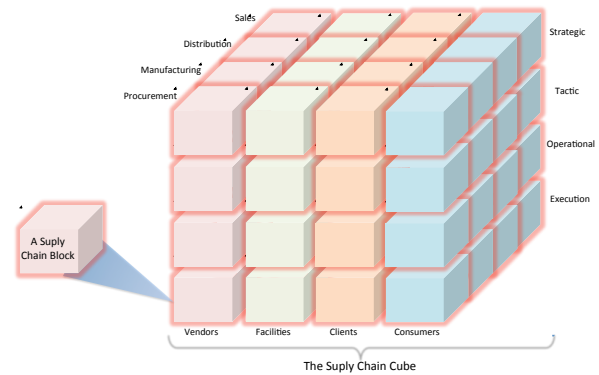


Fig. 2: Supply Chain Planning Cube [23].

A vertical slice of the supply chain planning cube for one spatial unit of analysis (e.g. facilities) is similar to the planning matrix proposed by Meyr and Stadtler [27], except for the execution level. The supply chain planning cube is an evolution of the planning matrix, due to the fact that it represents the possibility of collaboration among different traditional APS systems. It also includes execution entities.

Based on the supply chain cube, one has to perform requirements determination for the simulation aspects. This cube serves as a metamodel to help simulation analysts identify their simulation requirements. For example, the analysts decide which kind of Supply Chain Blocks will be needed in their simulation experiments, providing the basic architectural aspects of the simulation system. Then, their requirements are organized through UML-based use cases and requirements diagrams from SyML. An example of the DPA is provided in Santa-Eulalia et al. [23].

### 3.3 Social Agent Organization Analysis (SAOA)

So far, the concept of Supply Chain Block has been used to represent entities responsible for part of the supply chain planning. Together they compose a population of entities interacting with each other, having a collective co-existence within the planning system. When these entities incorporate attitudes, orientations and behaviours comprising the interests, needs or intentions of other Supply Chain Blocks, they can be seen as social entities. They can exhibit complex actions that take into account the collectivity. A way to represent social entities is to model them as agents, thus creating multi-agent societies.

The general logic indicated that a Supply Chain Block can be directly translated into agents by adding agent abilities to them. This is based on the agentification definition of Shen et al. [28], who explain that the agentification process can be functional-based (i.e. white Supply Chain Block) or physical-based (i.e. gray Supply Chain Block).

However, in some situations a Supply Chain Block can be transformed into more than one agent, for example when specialization is required, in which case a planning agent can be specialized according to certain generic responsibility orientations, such as products, processors, processes or projects, to obtain faster or more precise responses for certain given situations. In other situations, apart from agents proceeding from the supply chain planning cube, different intermediary agents can be created to perform activities related to, e.g. the coordination of the agents' society. In addition, the agentification process can also include the representation of information sources, interfaces and other services.

The importance of this discussion relies on the notion that agentification is the basis for two mutually dependent aspects in agent-based systems which define the metamodel for the SAOA:

- *Social structures*: represent the agent system architecture [24] characterizing the blueprint of relationships, giving a high level view of how groups solve problems and the role each agent plays within the structure. There are diverse types of social structures, such as hierarchical, federated and autonomous.
- *Social protocols*: are agents' abilities concerning social aspects, normally related to cooperation principles (i.e. agents have to cooperate in order to plan the entire supply chain). Diverse abilities can be considered, like communication, grouping and multiplication, coordination, collaboration by sharing tasks and resources and conflict resolution through negotiation and arbitration.

Different social structures and protocols are provided in Santa-Eulalia [22].

Similar to the DPA, these two aspects of the SAOA serve as a metamodel to help simulation analysis identify their requirements for the simulation model. For example, different social protocols can be tested in the simulation. Then, requirements can be organized through agent-based use cases from AUML (Agent Unified Modelling Language) and requirements diagrams from SysML. An example of the SAOA is provided in Santa-Eulalia et al. [23].

### 3.4 Individual Agent Organization Analysis (IAOA)

As mentioned by Ferber [29], the task of assigning roles to every individual agent is normally regarded as the last phase in constructing an organization. The logic is that as soon as one knows what the functions to be assigned are, one defines individual specializations. These local assignments influence social protocols functioning inside their respective social structures. In addition, it also influences the local performance of the supply chain planning entities. This is the main idea of the IAOA.

At the individual level, agents can be organized according to different internal architectures but there is little consensus on how to conceive the internal architectures of agents [30] in the literature. In order to cope with this, the metamodel for the IAOA proposes that whatever the state of mind of an agent is (cognitive, reactive or hybrid), and whatever internal architecture an agent employs, an agent can be described simply according to its 'abilities'. This is the central point when performing simulation. An 'ability' can be defined as the quality of being able to perform an action, or facilitate the action's accomplishment. 'Abilities' allow for the implementation of actions and the determination of the system's behaviour, as well as the determination of its related performance.

Based on this notion, the metamodel defines two elements:

- *The Response Space*: stands for a collection of general abilities available for the agents, including very simple reactive abilities or sophisticated cognitive ones. For example, one agent can have a simple ability to monitor the inventory levels of the supply chain, or a complex ability to perform production planning employing an optimization method.
- *Capacity to Produce an Adapted Response*: represents the aptitude to choose which abilities have to be transformed into actions at a given time to respond to a given situation. This capacity can vary from elementary to complex. The simplest possible capacity is related to a reactive 'if-then' mechanism, where no cognition is necessary. For example, if the inventory level drops to a given threshold, the agent uses its procurement ability to start a procurement action. As the agent becomes more intelligent, more complex responses can be made for some given situations. For example, the linear "if-then" logic can be substituted by more complex approaches based on action optimization and learning.

Based on these two elements of the metamodel, one can carry out requirements determination for the simulation model, selecting the desired requirements in terms of agents' abilities. Similar to the SAOA, the IAOA's requirements are organized through agent-based use cases from AUML and requirements diagrams from SysML [23].

FAMASS is detailed in Santa-Eulalia et al. [21, 22, 23]. An application of this approach is presented in Santa-Eulalia et al. [24].

## 4. Labarthe et al.'s Methodological Framework

The Labarthe et al. [9] framework is schematized in Fig. 3 and is briefly described afterwards.

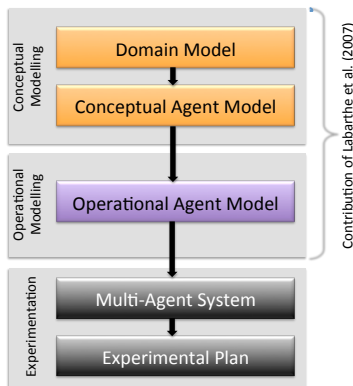


Fig. 3: Summarizing the Labarthes et al. [9] framework.

The authors propose the modelling steps indicated in Fig. 3. Their contribution corresponds to two abstraction levels: conceptual modelling and operational modelling. Conceptual modelling is performed in two steps, the Domain Model and the Conceptual Agent Model.

#### 4.1 Domain Model (DM)

The Domain Model (DM) creates an abstraction of the supply chain. Inspired from the NetMAN approach [31, 32], Labarthe et al. [9] create two sub-models: a Structural Model and a Dynamic Model.

The Structural Model, which is based on responsibility networks [33], defines the structure of the supply chain, i.e. its ‘actors’ and their related responsibilities, and it also depicts the material flows among all ‘actors’. The Dynamic Model complements the Structural Model by defining the behaviour of each ‘actor’ and its related interaction modes.

#### 4.2 Conceptual Agent Model (CAM)

The Conceptual Agent Model (CAM) remodels the Domain Model guided by the agentification process. From the Structural and Dynamic models, a unique agent model is created. A Conceptual Agent Model specifies the ‘agents’, the ‘objects’ transacted between them and the nature of the agent’s interactions (‘physical interactions’ and ‘informational interactions’). In this case, each ‘actor’ specified in the Structural Model produces a specific agent. Also, any activity of an actor generates a specific agent in close interaction with the agent associated to the actor concerned, which is regrouped in the same partition. In addition, any exchange of information from the Dynamic Model generates a message-based informational interaction; and any material flow from the dynamic model leads to a physical type interaction.

After, at the Operational Level, Labarthe et al. [9] proposes the Operational Agent Model (OAM).

#### 4.3 Operational Agent Model (OAM)

The Operational Agent Model (OAM) is based on the Conceptual Agent Model, and it aims to build a computer model of the studied supply chain which will be later implemented on a simulation platform. First, the Operational Agent models the software architecture (at the social level). Next, it models the internal agent architecture (individual level), dealing with knowledge, behaviours and interactions among agents.

After creating the Domain Model, the Conceptual Agent Model and the Operational Agent Model, a Multi-Agent System is implemented at the Exploitation level and a set of Experimental Plans supports the realization of simulation experiments (the black modelling approaches shown in Fig. 3). The author illustrated the Exploitation level through the implementation of a case study in a simulation environment.

This is only a summarized review of Labarthe et al. [9]’s work. For further details about this framework and its applications, the reader is referred to Labarthe et al. [9, 35, 36] and Labarthe [34].

### 5. The Deployment Process

As explained in the introduction, the original framework of Labarthe et al. [9] had to be slightly adapted to be suitable to the distributed APS domain.

The first adaptation occurs at the Domain Modelling. The main reason for not strictly employing the Labarthe et al. [9] Domain Model is because it is based on the responsibility network [33], which uses the definition of centre, i.e. a business entity – a decisional one – linked at the physical level by material flow. Centres do not correspond exactly to our semi-autonomous units, the Supply Chain Blocks (defined in subsection 3.2), which are based on the supply chain cube. We adapted the Labarthe et al. [9] model and thus proposed a modelling approach where the ‘centres’ are substituted by Supply Chain Blocks.

Another relevant difference refers to the fact that we separate the Operating System (i.e. the Execution layer) and the Decision System (i.e. the Strategic, Tactical and Operational layers) in the Domain Model, which is not done in the Labarthe et al. [9] Domain Model. They distinguish these two layers later, in the Operational Agent Model. We decided to separate them earlier because both



systems have to be identified in regard to the supply chain cube introduced in subsection 3.2. If we did not consider entities of the Operating System at this step, the Domain Model would be incomplete for a distributed APS, according to the definition of the supply chain cube.

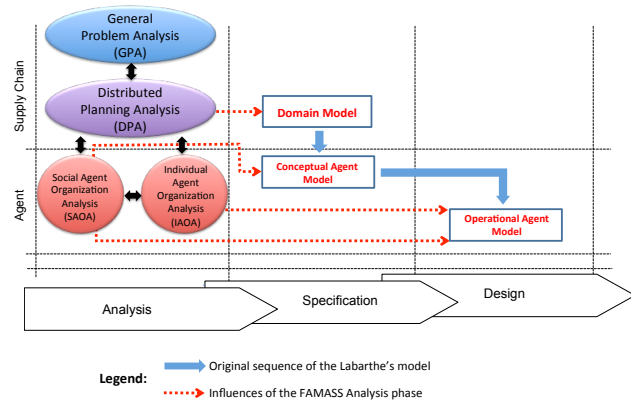


Fig. 4: Deploying process.

Fig. 4 depicts the general idea of the deployment process. From the analysis phase, the Distributed Planning Analysis models are the basis for the creation of the Domain Model. The Domain Model represents the supply chain under study and how advanced planning decisions are articulated. Next, the Conceptual Agent Model is naturally created from the Domain Model, but the Social Agent Organization Analysis is also used as an important reference. The Social Agent Organization Analysis provides the Social Structures for the Conceptual Agent Model and it reflects the agentification process used during the Social Agent Organization Analysis. Finally, the Operational Agent Model is created from the Conceptual Agent Model. However, relevant information about social protocols requirements comes from the Social Agent Organization Analysis, while requirements concerning the agents' abilities come from the Internal Agent Organization Analysis.

It is interesting to note, in Fig. 4, that the Domain Model and Conceptual Agent Model roughly correspond to the specification phase, while the Operational Agent Model can be considered equivalent to the design phase. The Domain Model and Conceptual Agent Model are the first formal models to describe the supply chain and the agent domain. The Operational Agent Model is closely related to how agents operate.

To sum up, FAMASS proposes a set of abstract notions for distributed APS systems, while Labarthe et al. [9] provide a formal and detailed description of how the system should work.

The following subsection discusses the Domain Model generation.

### 5.1 Domain Model (DM)

The objective of the Domain Model is to identify what is to be modelled in the supply chain. As seen in Fig. 4, the Distributed Problem Analysis (DPA) can be translated directly into the Domain Model.

Table 1 and Table 2 provide a translation strategy to create FAMASS Structural and Dynamic Models based on Labarthe et al. [9].

Table 1: Structural Model.

Element	Labarthe et al.	FAMASS Counterpart
<b>Central elements</b>	<p><b>Main element:</b> a network of Centres [33] (roles and responsibilities) and their interactions.</p> <p><b>Roles:</b> Processor, producer, assembler, fulfiller, distributor, retailer, transporters, customer. Roles define the nature of the responsibility set.</p> <p><b>Responsibilities:</b> examples, packing, grouping, sales, etc.</p> <p><b>Organizational level:</b> supply chain, enterprise, business unit, cells, resources.</p>	<p><b>Main element:</b> a network of Supply Chain Blocks and their interactions (interactions are simple representations of Supply Chain Block's relations). A Supply Chain Block is used instead of centres.</p> <p><b>Roles:</b> From the "spatial" axis of the supply chain cube (subsection 3.2), we identify the Supply Chain Blocks and their roles: vendors, facilities, clients and consumers.</p> <p><b>Responsibilities:</b> one can identify responsibilities from the 'functional' axis of the supply chain cube (subsection 3.2): procurement, manufacturing, distribution and sales.</p> <p><b>Organizational levels:</b> strategic, tactical, operational, execution for vendors, facilities, clients and consumers (i.e., the intertemporal axis).</p>
<b>Modelling formalism</b>	<p>Responsibility networks of Montreuil and Lefrançois [33].</p>	<p>Class diagrams and class tables (from AUML – Agent Unified Modelling Language). The concept is the same as for responsibility network, but it is represented using AUML formalisms. Centres are classes; roles are roles in each class; responsibilities are operations in each class; organizational levels are stereotypes of the classes; business processes are operations in each class.</p>
<b>Modelling process</b>	<p>Identify decision elements of the supply chain and the physical interactions among them.</p>	<p>We identify the elements from the execution and decision systems and we add only the physical interactions. Informational interactions are added in the dynamic model (later on in the modelling process).</p>

Table 2: Dynamic Model.

Element	Labarthe et al.	FAMASS Counterpart
<b>Central elements</b>	Describes (in time) the system behaviour and the elements that compose it. Uses the responsibility network to recognize [33] the coordination modes by identifying the physical and informational relations used according to the environmental stimulus.	Describes the same elements, but with the possibility to add more information based on different experimental definitions, i.e. different configurations of the Supply Chain Blocks, and different performance indicators and uncertainties.
<b>Modelling formalism</b>	NetMan [31, 32] approach plus a representation of the decoupling point position. The decoupling point position is mentioned here because it is an important issue in the Labarthe et al. [9] framework.	Class diagrams and class tables (AUML). All flows are represented by arrows. The decoupling point is represented in the class name. Centre models are represented by arrows as well. Stock holding (raw material, work-in-process or final products) is represented in the operations of each class.
<b>Modelling process</b>	Apart from the physical flow identified previously, the modelling process describes the informational flow exchanged according to the dynamics of the environment.  Four informational flow types for coordination are identified: i) needs expression; ii) offers expression; iii) information about coordination; and iv) information sharing by models exchanges. In addition, the decoupling point is positioned and inventories are mapped (raw material, work-in-process and final product).  It identifies two models (for models exchange): the network model and the centre model.	The same flows are identified, as well as inventory positions and decoupling point position. They are described in the class tables.

The most important difference between Labarthe et al. [9] and FAMASS is the use of centre for the former and the use of Supply Chain Block for the latter. Supply Chain Block is used instead of centres in FAMASS because decision entities are central elements. Labarthe [34, p.119] explains that a centre represents a decision process, but centre definitions are closely associated to physical entities of the execution system, i.e. there is a direct relation between a centre and an entity of the execution system. Later in the Labarthe et al. [9] modelling process, the decision system is introduced more formally in the Operational Agent Model. We separate the decision system from the execution system in the Domain Model, since we know that they are relevant for experimental

definitions in distributed APS systems. Another difference is related to the fact that we employ a unique modelling formalism based on an AUML approach, coherent with the analysis phase of FAMASS, which employs only UML-inspired formalisms.

The next sub-section transforms the Domain Model into a Conceptual Agent Model.

## 5.2 Conceptual Agent Model (CAM)

The Conceptual Agent Model represents the agentification process of the Labarthe et al. [9] approach. The agentification process defines the agent society based on the Domain Model, i.e. which agents are created from the centres (in our case, Supply Chain Block) and how they are organized. Labarthe et al. [9] propose rules for creating agents (i.e., each centre becomes an actor-agent and each centre activity becomes an activity-agent). As discussed before, FAMASS converts each Supply Chain Block into an agent. It also verifies whether some agents are extinguished (e.g. merged with another agent) or whether new agents are introduced (e.g. a mediator). This information is obtained during the Social Agent Organization Analysis (SAOA).

As indicated in Fig. 4, the Conceptual Agent Model is generated from the Domain Model and the SAOA (in this case, the social structures). Using Labarthe et al. [9] rules, the Domain Model provides the basic classes' definition and, using the SAOA, it can be verified if new agent classes are derived from the Domain Model and if different social structures have to be tested and considered in the Conceptual Agent Model. Social Protocols from SAOA are not used in Conceptual Agent Modelling.

The Strategy for creating a Conceptual Agent Model is shown in Table 3.

Table 3: Conceptual Agent Models.

Element	Labarthe et al.	FAMASS Counterpart
Central elements	<p><b>Actor-agent:</b> centre.</p> <p><b>Activity-agent:</b> represents a process of transformation, distribution, or stock keeping.</p> <p><b>Object:</b> products.</p> <p><b>Informational interaction:</b> same as in Domain Model.</p> <p><b>Physical interaction:</b> same as in Domain Model.</p>	<p><b>Actor-agent:</b> agents representing an organizational unit of the supply chain (i.e. vendors, facilities, clients or customers), related to the 'spatial' axis. Actor-agents group several other agents, the activity-agents.</p> <p><b>Activity-agent:</b> agents from the decision system, representing the processes of procurement, manufacturing distribution and sales. These agents are at three different decision levels and they are related to the 'functional' axis.</p> <p><b>Objects:</b> defined products. This is the first time products are specified.</p> <p><b>Information interactions:</b> they come from the Domain Model.</p> <p><b>Physical Interactions:</b> they come from the Domain Model.</p>
Modelling formalism	<p>A graphical modelling formalism [34] that models the two types of agents and their interactions. The CAM model is derived from the DM model.</p>	<p>Adapted class diagrams, tables and package diagrams. The adaptation of the class diagrams refers to the insertion of objects (products), represented by simple square boxes in the link between two classes.</p>
Modelling process	<ol style="list-style-type: none"> <li><b>From centre to actor-agent:</b> each centre creates an actor-agent.</li> <li><b>Physical interactions between actor-agents:</b> physical flow is specified by an arrow linking agents and indicating their respective exchanged objects.</li> <li><b>Informational interactions between actor-agents:</b> similar to 2, but for information flow.</li> <li><b>Organizational frontiers definition:</b> establishes the organization frontiers for the actor-agents and places the physical flows between the organizations.</li> <li><b>Definition of the activity-agents:</b> each activity of a centre is transformed into an activity-agent.</li> <li><b>Physical interactions between activity-agents:</b> specify the physical flow between the activity-agents and their related objects exchanged.</li> <li><b>Informational interactions between activity-agents:</b> same as 6, plus the interaction between actor-agents and activity-agents.</li> </ol>	<p><b>Similar process, with the following differences:</b></p> <ul style="list-style-type: none"> <li>- Actor-agents and activity-agents: in the classes, use role definitions to indicate if it is an actor-agent or an activity-agent;</li> <li>- Interactions: links between classes.</li> </ul>

It is important to note that an actor-agent coordinates a population of other activity-agents in the Labarthe et al. [9] approach. In the case of FAMASS, we decided to use the notion of actor-agent only as an aggregation of agents inside the same organization using a package diagram.

The next sub-section transforms the Conceptual Agent Model into an Operational Agent Model.

### 5.3 Operational Agent Model (OAM)

According to Labarthe [34], the OAM represents implementable models. These models involve a choice between two different agent architectures, i.e. the cognitive and the reactive architectures. We believe that most of the time it is not possible to completely distinguish cognitive agents from deliberative agents, meaning that normally agents can be seen as a hybrid state within the cognitive-reactive continuum. In Labarthe et al. [9]'s work, agents from the decision system assume a cognitive agent architecture, composing a cognitive agent society. Based on this society, the author then creates a reactive society responsible for the transformation process (execution system), linked with the cognitive society.

As we believe that the agents from the decision system can also assume reactive behaviours (see subsection 3.2), we prefer not to use this agent architecture notation for the Operational Agent Model. Instead, we create two societies (decision agents and execution agents) from the Conceptual Agent Model and start to define all agents' behaviours and agents' protocols in detail, as done by Labarthe et al. [9], which is not contradictory to Labarthe et al.'s [9] work. As explained before, instead of separating into decision and execution societies at the Operational Agent Model, our approach does it at the beginning of the specification phase, i.e. at the Domain Model.

In sum, our Operational Agent Model is generated from the Conceptual Agent Model, the Social Agent Organization Analysis and the Internal Agent Organization Analysis, as illustrated in Fig. 5.

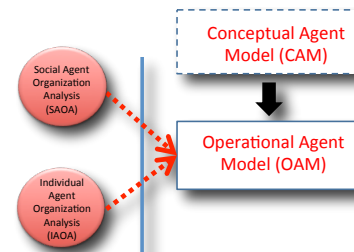


Fig. 5: Creating an Operational Agent Model.

From the Conceptual Agent Model we represent two societies, the decision agents and the execution agents. This is the starting point of the Operational Agent Model. After, we obtain requirements about agent protocols from the Social Agent Organization Analysis, and we obtain requirements about agent abilities from the Internal Agent Organization Analysis.

Table 4 summarizes the deployment strategy for the Operational Agent Model.

Table 4: Operational Agent Models.

Element	Labarthe et al.	FAMASS Counterpart
Central elements	<p><b>Multi-agent system architecture:</b> a cognitive and a reactive agent society are represented. A cognitive agent, together with its corresponding reactive agent, form the 'agent-actor'. It is a generic architecture to represent entities capable of taking their own decisions and acting accordingly.</p> <p><b>Specification of the software agent:</b> knowledge, behaviour and interactions of each agent are defined. For the behaviours, the following entities are defined: a) external event: concerning the communication aspect with external entities of the multi-agent system; b) internal event: concerning internal activities of an agent; c) passive state: waiting state; d) active state, being an elementary action or a composite action.</p>	<p><b>Multi-agent system architecture:</b> cognitive agents are seen as decision agents (from the decision system); reactive agents are represented by execution agents (from the execution system).</p> <p><b>Specification of the software agent:</b> same elements, i.e. knowledge, behaviour and interactions.</p>
Modelling formalism	<p>For the multi-agent system architecture, Labarthe [34] proposes his own graphical modelling formalism. For the specification of the software agent for cognitive behaviours, the Agent Behaviour Representation (ABR) formalism [37] is used. For reactive agent behaviours, AUML formalisms are used, specifically state charts. For interactions, protocol diagrams from AUML are used.</p>	<p>We used only adapted diagrams from AUML. For behaviours and knowledge representation, we employ Activity Diagrams. For interactions, we use Protocol Diagrams.</p>
Modelling process	<ol style="list-style-type: none"> <li>1. Create a society of cognitive agents. Incorporate the informational flow.</li> <li>2. Create a society of reactive agents. Incorporate the physical flow and the related exchanged objects (products).</li> <li>3. Define the responsibility links between cognitive and reactive agents.</li> <li>5. Specify agent behaviour of the cognitive society using the Agent Behaviour Representation (ABR) formalism.</li> <li>6. Specify agent's behaviour of the reactive society using statecharts.</li> <li>7. Specify agents' interactions through protocol diagrams.</li> </ol>	<p>Same process, but with different formalisms from AUML.</p>

The next sub-section provides some final remarks and conclusions about the proposed deployment strategy.

## 6. Final Remarks and Future Works

This paper presents a conversion strategy from the FAMASS analysis models into specification and design models inspired by the methodological agent-based framework of Labarthe et al. [9]. This strategy facilitates the FAMASS analysts in converting their models and

going faster and smoother through the whole modelling process.

In addition, this deployment strategy demonstrates that the analysis phase of FAMASS can be integrated with other existing approaches specialized in specification and design modelling. With this as an impetus, other methodological frameworks could be inspected in the future so as to verify that FAMASS concepts adhere to other frameworks.

Furthermore, the proposed strategy allows us to avoid the research effort needed to develop a totally new specification and design methodology for the domain, although it would be suitable and desirable for future research initiatives. With regard to this, a forthcoming research effort will work on extending the FAMASS analysis approach, so as to cover the whole FAMASS life-cycle from analysis to simulation. In this way the proposed deploying strategy launches the basis for this FAMASS-extended version of a complete architecture to deal with agent-based simulations in the context of distributed APS systems. Future versions of the FAMASS approach are to be published shortly.

## References

- [1] APICS, "The Association of Operations Management - Online Dictionary", Retrieved June, 2008, from www.apics.org.
- [2] K. Kumar, "Technology for supporting supply chain management", Communications of the ACM, Vol. 44, No. 6, 2001, pp. 58-61.
- [3] M. Van Eck, Advanced planning and scheduling: is logistics everything? Working Paper, Vrije Universiteit Amsterdam, Amsterdam, 2003.
- [4] J.-M. Frayret, S. D'Amours, A. Rousseau, S. Harvey, S., and J. Gaudreault, "Agent-based supply chain planning in the forest products industry", International Journal of Flexible Manufacturing Systems, Vol. 19, No. 4, 2007, pp. 358-391.
- [5] A. L. Azevedo, C. Toscano, J. P. Sousa, and A. L. Soares, "An advanced agent-based order planning system for dynamic networked enterprises", Production Planning & Control, Vol. 15, No. 2, 2004, pp. 133-144.
- [6] L. Cecere, "A changing technology landscape", Supply Chain Management Review, Vol. 10, No. 1, 2006.
- [7] J.-H. Lee, and C.-O. Kim, "Multi-agent systems applications in manufacturing systems and supply chain management: a review paper" International Journal of Production Research, Vol. 46, No. 1, 2008, pp. 233-265.
- [8] W. Shen, Q. Hao, H. L. Yoon, and D. H. Norrie, "Applications of agent-based systems in intelligent manufacturing: An updated review", Advanced Engineering Informatics, Vol. 20, No. 4, 2006, pp. 415-431.
- [9] O. Labarthe, B. Espinasse, A. Ferrarini, A., and B. Montreuil, "Toward a methodological framework for agent-based modelling and simulation of supply chain in a



- mass customization context", *Simulation Modelling Practice and Theory*, Vol. 15, No. 2, 2007, pp. 113-136.
- [10] H. Baumgaertel, and U. John, "Combining agent-based supply net simulation and constraint technology for highly efficient simulation of supply networks using APS systems", in 2003 Winter Simulation Conference, 2003.
- [11] J. M. Swaminathan, S. F. Smith, and N. M. Sadeh, "Modeling supply chain dynamics: a multiagent approach", *Decision Sciences*, Vol. 29, No. 3, 1998, pp. 607-632.
- [12] L. A. Santa-Eulalia, S. D'Amours, and J.-M. Frayret, "Essay on conceptual modeling, analysis and illustration of agent-based simulations for distributed supply chain planning", *INFOR Information Systems and Operations Research Journal*, Vol. 46, No. 2, 2008, pp. 97-116.
- [13] S. Galland, F. Grimaud, P. Beaune, and J. P. Campagne, "MAMA-S: an introduction to a methodological approach for the simulation of distributed industrial systems", *International Journal of Production Economics*, No. 85, 2003, pp. 11-31.
- [14] R. Govindu, and R.B. Chinnam, "A software agent-component based framework for multi-agent supply chain modelling and simulation", *International Journal of Modelling and Simulation*, Vol. 30, No. 2, 2010.
- [15] L. A. Santa-Eulalia, G. Halladjian, S. D'Amours, and J.-M. Frayret, "Integrated methodological frameworks for modelling agent-based APS systems: a systematic literature review", Working Paper CIRRELT-2011-50, CIRRELT – Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation, 2011, available at [www.cirrelt.ca](http://www.cirrelt.ca).
- [16] M. S. Fox, J. F. Chionglo, and M. Barbuceanu, "The integrated supply chain management system", Internal Report - Department of Industrial Engineering, University of Toronto, Canada, from [www.eil.utoronto.ca/iscm-descr.html](http://www.eil.utoronto.ca/iscm-descr.html), 1993.
- [17] H. Stadtler, and C. Kilger, *Supply chain management and advanced planning: concepts, models, software and case studies*, Berlin: Springer, 2004.
- [18] D. J. Van der Zee, and J.G.A.J. Van der Vorst, "A Modeling framework for supply chain simulation: opportunities for improved decision making", *Decision Sciences*, Vol. 36, No. 1, 2005, pp. 65-95.
- [19] P. Egri, and J. Vancza, "Cooperative planning in the supply network – a multiagent organization model", in CEEMAS 2005 - 4th International Central and Eastern European Conference on Multi-Agent Systems, Budapest, Hungary, Springer Verlag, 2005.
- [20] P. Lendermann, B. P. Gan, and L. F. McGinnis, "Distributed simulation with incorporated APS procedures for high-fidelity supply chain optimization", in 2001 Winter Simulation Conference, Arlington, 2001.
- [21] L. A. Santa-Eulalia, S. D'Amours, and J.-M. Frayret, "Modeling Agent-Based Simulations for Supply Chain Planning: the FAMASS Methodological Framework", in 2010 IEEE International Conference on Systems, Man, and Cybernetics, Special Session on Collaborative Manufacturing and Supply Chains, Istanbul, 10-13 October 2010.
- [22] L. A. Santa-Eulalia, "Agent-based simulations for advanced supply chain planning: a methodological framework for requirements analysis and deployment", Ph.D. Thesis, Faculté des Sciences et Génie, Université Laval, Canada, 2009. 387p.
- [23] L. A. Santa-Eulalia, S. D'Amours, and J.-M. Frayret, "Agent-Based Simulations for Advanced Supply Chain Planning: The FAMASS Methodological Framework for Requirements Analysis and Deployment". Working Paper CIRRELT-2011-22, CIRRELT – Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation, 2011, available at [www.cirrelt.ca](http://www.cirrelt.ca).
- [24] L. A. Santa-Eulalia, D. Aït-Kadi, S. D'Amours, J.-M. Frayret, and S. Lemieux, "Agent-based experimental investigations about the robustness of tactical planning and control policies in a softwood lumber supply chain", *Production Planning & Control, Special Issue on Applied Simulation, Planning and Scheduling Techniques in Industry*, 1366-5871, first published on February 10th 2011 (iFirst).
- [25] OMG, "OMG Systems Modeling Language (OMG SysML™)", Object Management Group Specification Report, June 2010.
- [26] J. F. Shapiro, "Modeling the supply chain", Duxbury: Pacific Grove, 2000.
- [27] H. Meyr, and H. Stadtler, "Types of supply chain", in *Supply chain management and advanced planning: concepts, models, software and case studies*, Berlin: Springer, 2004.
- [28] W. Shen, F. Maturana, and D. H. Norrie, "MetaMorph II: an agent-based architecture for distributed intelligent design and manufacturing", *Journal of Intelligent Manufacturing*, No. 11, 2000, pp. 237-251.
- [29] J. Ferber, *Multi-agent systems: an introduction to distributed artificial intelligence*, Harlow: Addison-Wesley, 1999.
- [30] L. Sanya, and W. Hongwei, "Agent architecture for agent-based supply chain integration & coordination", *Software Engineering Notes*, Vol. 28, No. 4, 2003.
- [31] J.-M. Frayret, S. D'Amours, B. Montreuil, and L. Cloutier, "A network approach to operate agile manufacturing systems", *International Journal of Production Economics*, Vol. 74, No. 1-3, 2001, pp. 239-259.
- [32] B. Montreuil, J.-M. Frayret, and S. D'Amours, "A strategic framework for networked manufacturing", *Computers in Industry*, Vol. 42, No. 2-3, 2000, pp. 299-317.
- [33] B. Montreuil, and P. Lefrançois, "Organizing factories as responsibility networks", *Progress in Material Handling Research*, 1996, pp. 375-411.
- [34] O. Labarthe, "Modélisation et simulation orientées agents de chaînes logistiques dans un contexte de personnalisation de masse : modèles et cadre méthodologique", Ph.D. Thesis, Université Laval (Canada) and Université Paul Cézanne (France), 2006.
- [35] O. Labarthe, B. Montreuil, A. Ferrarini, and B. Espinasse, "Modélisation multi-agents pour la simulation de chaînes logistiques de type personnalisation de masse", in 5e Conférence Francophone de MODélisation et SIMulation: Modélisation et simulation pour l'analyse et l'optimisation des systèmes industriels et logistiques, MOSIM'04. Nantes, 2004.
- [36] O. Labarthe, E. Tranvouez, A. Ferrarini, B. Espinasse, and B. Montreuil, "A heterogeneous multi-agent modelling for

distributed simulation of supply chains” in HoloMAS 2003, pp. 134-145, 2003.

- [37] E. Tranvouez, “IAD et ordonnancement : une approche coopérative du réordonnancement par systèmes multi-agents”, Ph.D. Thesis, Université de Valenciennes et du Hainaut-Cambrésis, 2001.

**Luis Antonio de Santa-Eulalia** is a professor at Téléu-UQAM (Université du Québec à Montréal), Canada, a member of the innovation board of Axia Value Chain (North America division), and a researcher of the NSERC Strategic Research Network on Value Chain Optimization (VCO). He holds a Ph.D. in Industrial Engineering from Université Laval, Canada, an MSc. and BSc. both in Industrial Engineering respectively from the University of São Paulo, Brazil, and Federal University of São Carlos, Brazil. He has worked as a researcher and consultant in the domains of production planning and control, supply chain management, and simulations. His current research interests are related to novel business models and technology for sustainable value chain management.

**Sophie D'Amours** holds a Ph.D. in Applied Mathematics and Industrial Engineering from the École Polytechnique de Montréal, as well as a MBA and a BSc in Mechanical Engineering from Université Laval. She is currently a scientific director of the NSERC Strategic Research Network on Value Chain Optimization (VCO) and she holds a Canada Research Chair in Planning Sustainable Forest Value Networks as well as an NSERC Industrial Chair. She is also Director of the FORAC Research Consortium. Sophie is a full professor at the Faculty of Science and Engineering, Department of Mechanical Engineering, at Université Laval. Her research interests are in supply chain management and planning, web-based applications, and forest sector.

**Jean-Marc Frayret** is Associate Professor at the École Polytechnique de Montréal, Québec, Canada. He holds a Ph.D. in Mechanical and Industrial Engineering from Université Laval, Canada. He is a member of the CIRRELT, a research centre dedicated to the study of network organizations and logistics. He is also a researcher of the NSERC Strategic Research Network on Value Chain Optimization (VCO) and of the FORAC Research Consortium. His research interests include agent-based and distributed manufacturing systems, supply chain management and interfirm collaboration. Dr. Frayret has published several articles in these fields in various journals and international conferences.

# Facial Expression Classification Based on Multi Artificial Neural Network and Two Dimensional Principal Component Analysis

Thai Le<sup>1</sup>, Phat Tat<sup>1</sup> and Hai Tran<sup>2</sup>

<sup>1</sup> Computer Science Department, University of Science, Ho Chi Minh City, Vietnam

<sup>2</sup> Informatics Technology Department, University of Pedagogy, Ho Chi Minh City, Vietnam

## Abstract

Facial expression classification is a kind of image classification and it has received much attention, in recent years. There are many approaches to solve these problems with aiming to increase efficient classification. One of famous suggestions is described as first step, project image to different spaces; second step, in each of these spaces, images are classified into responsive class and the last step, combine the above classified results into the final result. The advantages of this approach are to reflect fulfill and multiform of image classified. In this paper, we use 2D-PCA and its variants to project the pattern or image into different spaces with different grouping strategies. Then we develop a model which combines many Neural Networks applied for the last step. This model evaluates the reliability of each space and gives the final classification conclusion. Our model links many Neural Networks together, so we call it Multi Artificial Neural Network (MANN). We apply our proposal model for 6 basic facial expressions on JAFFE database consisting 213 images posed by 10 Japanese female models.

**Keywords:** Facial Expression, Multi Artificial Neural Network (MANN), 2D-Principal Component Analysis (2D-PCA).

## 1. Introduction

There are many approaches apply for image classification. At the moment, the popular solution for this problem: using K-NN and K-Mean with the different measures, Support Vector Machine (SVM) and Artificial Neural Network (ANN).

K-NN and K-Mean method is very suitable for classification problems, which have small pattern representation space. However, in large pattern representation space, the calculating cost is high.

SVM method applies for pattern classification even with large representation space. In this approach, we need to

define the hyper-plane for classification pattern [1]. For example, if we need to classify the pattern into L classes, SVM methods will need to specify  $1 + 2 + \dots + (L-1) = L(L-1) / 2$  hyper-plane. Thus, the number of hyper-planes will rate with the number of classification classes. This leads to: the time to create the hyper-plane high in case there are several classes (costs calculation).

Besides, in the situation the patterns do not belong to any in the L given classes, SVM methods are not defined [2]. On the other hand, SVM will classify the pattern in a given class based on the calculation parameters. This is a wrong result classification.

One other approach is popular at present is to use Artificial Neural Network for the pattern classification. Artificial Neural Network will be trained with the patterns to find the weight collection for the classification process [3]. This approach overcomes the disadvantage of SVM of using suitable threshold in the classification for outside pattern. If the patterns do not belong any in L given classes, the Artificial Neural Network identify and report results to the outside given classes.

In this paper, we propose the Multi Artificial Neural Network (MANN) model to apply for image classification.

Firstly, images are projected to difference spaces by Two Dimensional Principal Component Analysis (2D-PCA).

Secondly, in each of these spaces, patterns are classified into responsive class using a Neural Network called Sub Neural Network (SNN) of MANN.

Lastly, we use MANN's global frame (GF) consisting some Component Neural Network (CNN) to compose the classified result of all SNN.

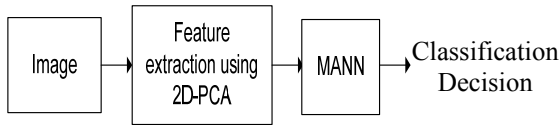


Fig 1. Our Proposal Approach for Image Classification

## 2. Background and Related Work

There are a lot of approaches to classify the image featured by  $m$  vectors  $X = (v_1, v_2, \dots, v_m)$ . Each of patterns is needed to classify in one of  $L$  classes:  $\Omega = \{\Omega_i \mid 1 \leq i \leq L\}$ . This is a general image classification problem [3] with parameters  $(m, L)$ .

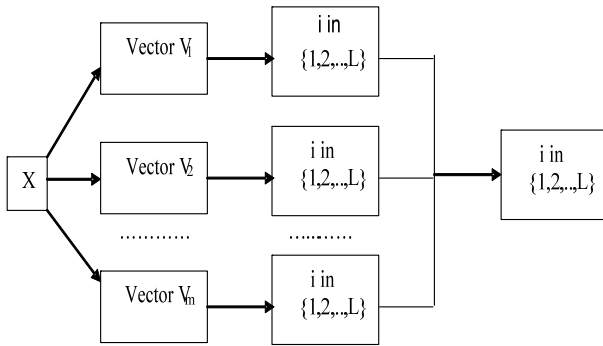


Fig 2. Image with  $m$  feature vectors Classification

First, the extraction stage featured in the image is performed. It could be used wavelet transform, or Principal Component Analysis (PCA). PCA known as one of the well-known approach for facial expression extraction, called "Eigenface" [3]. In traditional PCA, the face images must be converted into 1D vector which has problem with high dimensional vector space.

Then, Yang et al. [12] has proposed an extension of PCA technique for face recognition using gray-level images. 2D-PCA treats image as a matrix and computes directly on the so-called image covariance matrix without image-to-vector transformation. The eigenvector estimates more accurate and computes the corresponding eigenvectors more efficiently than PCA. D. Zhang et al. [13] was proposed a method called Diagonal Principal Component Analysis (DiaPCA), which seeks the optimal projective vectors from diagonal face images and therefore the correlations between variations of rows and those of columns of images can be kept [3]. That is the reason why,

in this paper, we used 2D-PCA (rows, columns and block-based) and DiaPCA (diagonal-based) for extracting facial feature to be the input of Neural Network.

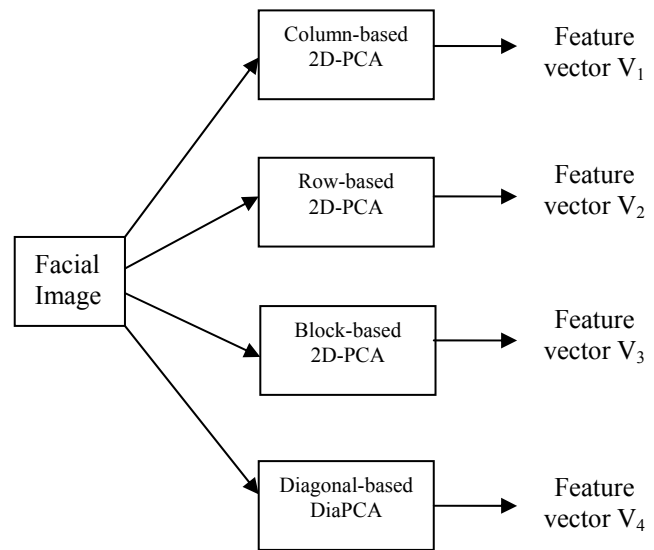


Fig 3. Facial Feature Extraction

Sub-Neural Network will classify the pattern based on the responsive feature. To compose the classified result, we can use the selection method, average combination method or build the reliability coefficients...

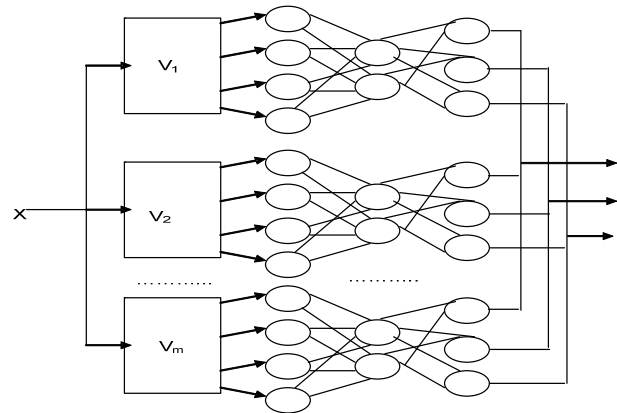


Fig 4. Processing of Sub Neural Networks

The selection method will choose only one of the classified results of a SNN to be the whole system's final conclusion:

$$P(\Omega_i \mid X) = P_k(\Omega_i \mid X) \quad (k=1..m) \quad (1)$$

Where,  $P_k(\Omega_i \mid X)$  is the image  $X$ 's classified result in the  $\Omega_i$  class based on a Sub Neural Network,  $P(\Omega_i \mid X)$  is the



pattern  $X$ 's final classified result in the  $\Omega_i$ . Clearly, this method is subjectivity and omitted information. The average combination method [4] uses the average function for all the classified result of all SNN:

$$P(\Omega_i | X) = \sum_{k=1}^m \frac{1}{m} P_k(\Omega_i | X) \quad (2)$$

This method is not subjectivity but it set equal the importance of all image features.

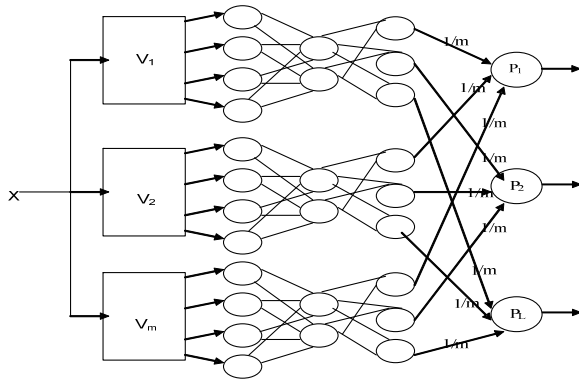


Fig 5. Average combination method

On the other approach is building the reliability coefficients attached on each SNN's output [4], [5]. We can use fuzzy logic, SVM, Hidden Markup Model (HMM) [6]... to build these coefficients:

$$P(\Omega_i | X) = \sum_{k=1}^m r_k P_k(\Omega_i | X) \quad (3)$$

Where,  $r_k$  is the reliability coefficient of the  $k^{\text{th}}$  Sub Neural Network. For example, the following model uses Genetics Algorithm to create these reliability coefficients.

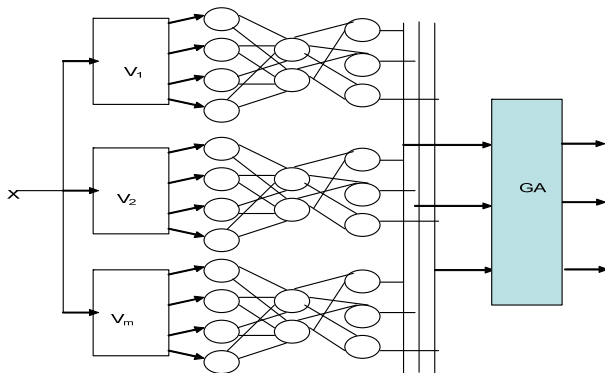


Fig 6. NN\_GA model [4]

In this paper, we propose to use Neural Network technique. In details, we use a global frame consisting of some

CNN(s). The weights of CNN(s) evaluate the importance of SNN(s) like the reliability coefficients. Our model combines many Neural Networks, called Multi Artificial Neural Network (MANN).

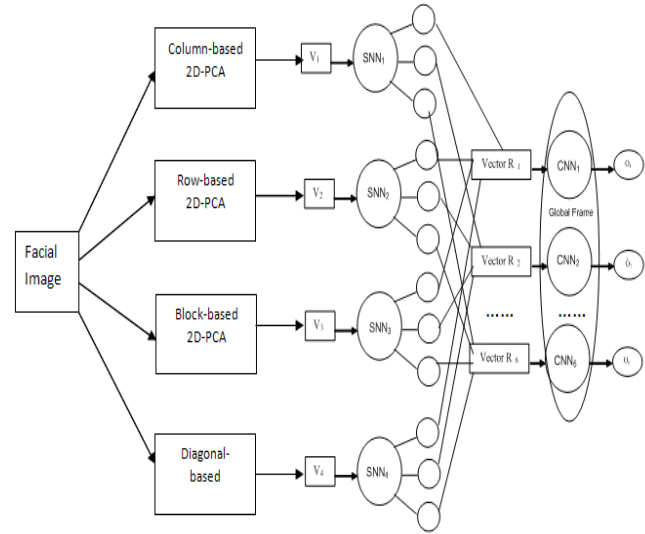


Fig 7. PCA and MANN combination

### 3. Image Feature Extraction using 2D-PCA

#### 3.1 Two Dimensional Principal Component Analysis (2D-PCA)

Assume that the training data set consists of  $N$  face images with size of  $m \times n$ .  $X_1, X_2, \dots, X_N$  are the matrices of sample images. The 2D-PCA proposed by Yang et al. [2] is as follows:

Step 1. Obtain the average image  $\bar{X}$  of all training samples:

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i \quad (4)$$

Step 2. Estimate the image covariance matrix

$$C = \frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^T \times (X_i - \bar{X}) \quad (5)$$

Step 3. Compute  $d$  orthonormal vectors  $W_1, W_2, \dots, W_d$  corresponding to the  $d$  largest eigenvalues of  $C$ .  $W_1, W_2, \dots, W_d$  construct a  $d$ -dimensional projection subspace, which are the  $d$  optimal projection axes.

Step 4. Project  $X_1, X_2, \dots, X_N$  on each vector  $W_1, W_2, \dots, W_d$  to obtain the principal component vectors:

$$F_i^j = A_j W_i, \quad i=1..d; j=1..N \quad (6)$$

Step 5. The reconstructed image of a sample image  $A_j$  is defined as:

$$A_{recs(j)} = \sum_{i=1}^d F_i^j W_i^T \quad (7)$$

### 3.2 DiaPCA

The DiaPCA extract the diagonal feature which reflects variations between rows and columns. For each face image in training set, the corresponding diagonal image is defined as follows:

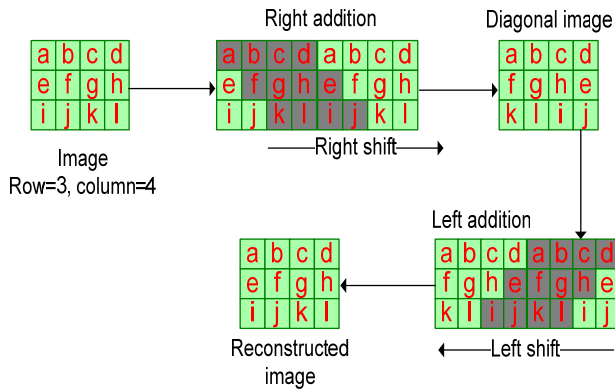


Fig 8. Extract the diagonal feature if rows  $\leq$  columns

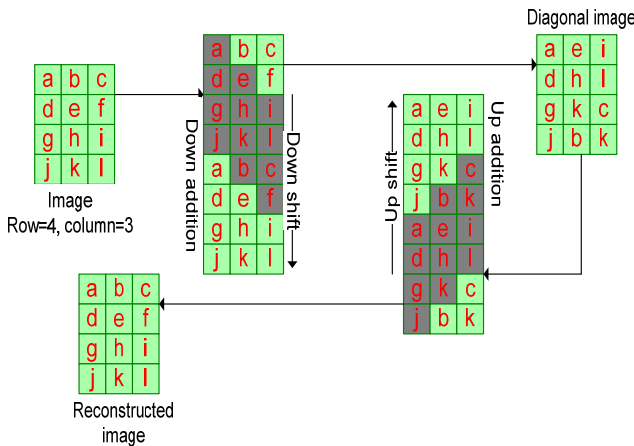


Fig 9. Extract the diagonal feature if rows  $>$  columns

### 3.3 Facial Feature Extraction

Facial feature extraction used 2D-PCA and its variants to project the pattern or image into different spaces with different grouping strategies. A facial image will be projected to 4 presentation spaces by PCA (column-based, row-based, diagonal-based, and block-based). Each of above presentation spaces extracts to the feature vectors.

So a facial image will be presented by  $V_1, V_2, V_3, V_4$ . In particular,  $V_1$  is the feature vector of column-based image,  $V_2$  is the feature vector of row-based image,  $V_3$  is the feature vector of diagonal-based image and  $V_4$  is the feature vector of block-based image.

Feature vectors ( $V_1, V_2, V_3, V_4$ ) presents the difference orientation of original facial image. They are the input to Multi Artificial Neural Network (MANN), which generates the classified result.

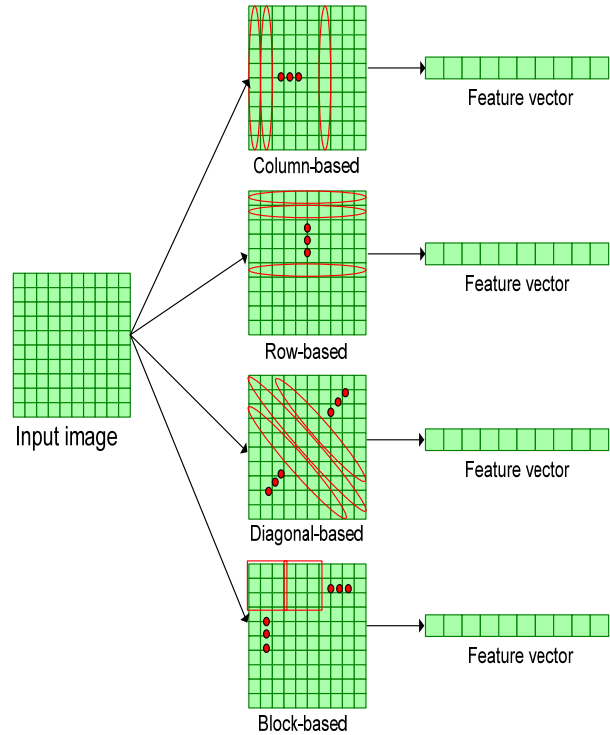


Fig 10. Facial Feature Extraction using 2D-PCA and DiaPCA

## 3. Multi Artificial Neural Network for Image Classification

### 3.1 The MANN structure

Multi Artificial Neural Network (MANN), applying for pattern or image classification with parameters  $(m, L)$ , has  $m$  Sub-Neural Network (SNN) and a global frame (GF) consisting  $L$  Component Neural Network (CNN). In particular,  $m$  is the number of feature vectors of image and  $L$  is the number of classes.

**Definition 1:** SNN is a 3 layers (input, hidden, output) Neural Network. The number input nodes of SNN depend on the dimensions of feature vector. SNN has  $L$  (the

number classes) output nodes. The number of hidden node is experimentally determined. There are  $m$  (the number of feature vectors) SNN(s) in MANN model. The input of the  $i^{\text{th}}$  SNN, symbol is  $SNN_i$ , is the feature vector of an image. The output of  $SNN_i$  is the classified result based on the  $i^{\text{th}}$  feature vector of image.

**Definition 2:** Global frame is frame consisting  $L$  Component Neural Network which compose the output of SNN(s).

**Definition 3:** Collective vector  $k^{\text{th}}$ , symbol  $R_k$  ( $k=1..L$ ), is a vector joining the  $k^{\text{th}}$  output of all SNN. The dimension of collective vector is  $m$  (the number of SNN).

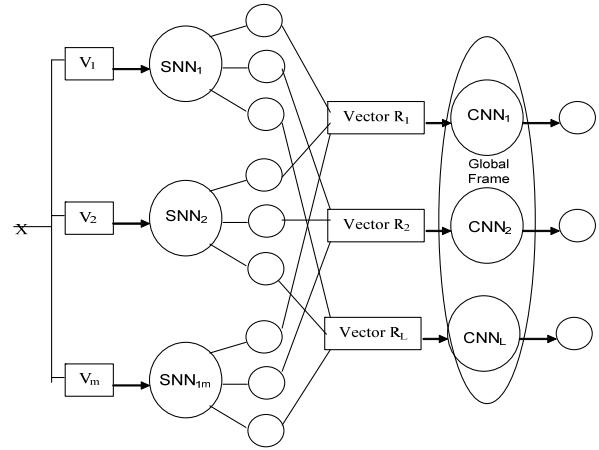


Fig 12. MANN with parameters ( $m, L$ )

### 3.2 The MANN training process

The training process of MANN is separated in two phases. Phase (1) is to train SNN(s) one-by-one called local training. Phase (2) is to train CNN(s) in GF one-by-one called global training.

In local training phase, we will train the  $SNN_1$  first. After that we will train  $SNN_2, SNN_m$ .

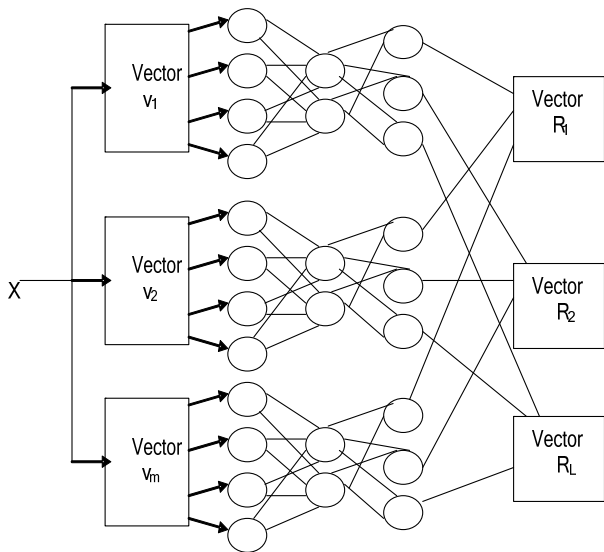


Fig 11. Create collective vector for CNN(s)

**Definition 4:** CNN is a 3 layers (input, hidden, output) Neural Network. CNN has  $m$  (the number of dimensions of collective vector) input nodes, and 1 (the number classes) output nodes. The number of hidden node is experimentally determined. There are  $L$  CNN(s). The output of the  $j^{\text{th}}$  CNN, symbols is  $CNN_j$ , give the probability of  $X$  in the  $j^{\text{th}}$  class.

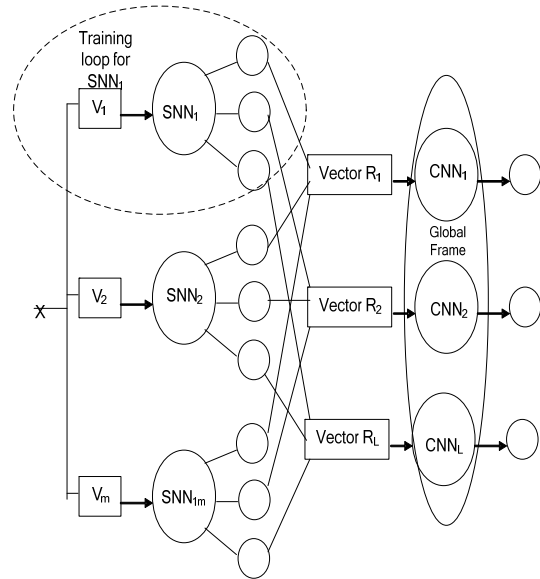


Fig 13. SNN1 local training

In the global training phase, we will train the  $CNN_1$  first. After that we will train  $CNN_2, \dots, CNNL$ .

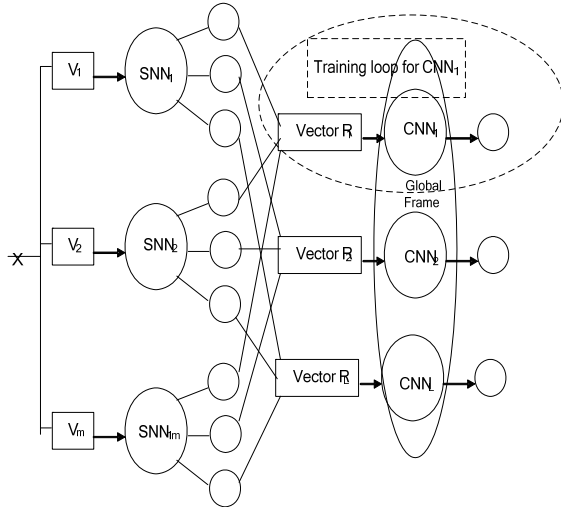


Fig 14. CNN<sub>1</sub> global training

### 3.3 The MANN classification

The classification process of pattern X using MANN is below: firstly, pattern X are extract to m feature vectors. The  $i^{th}$  feature vector is the input of  $SNN_i$  classifying pattern. Join all the  $k^{th}$  output of all SNN to create the  $k^{th}$  ( $k=1..L$ ) collective vector, symbol  $R_k$ .

$R_k$  is the input of  $CNN_k$ . The output of  $CNN_k$  is the  $k^{th}$  output of MANN. It gives us the probability of X in the  $k^{th}$  class. If the  $k^{th}$  output is max in all output of MANN and bigger than the threshold. We conclude pattern X in the  $k^{th}$  class.

### 4. Six Basic Facial Expressions Classification

In the above section, we explain the MANN in the general case with parameters (m, L) apply for image classification. Now we apply MANN model for six basic facial expression classifications. In fact that this is an experimental setup with MANN with (m=4, L=6).

We use an automatic facial feature extraction system using 2D-PCA (column-based, row-based and block based) and DiaPCA (diagonal-based).

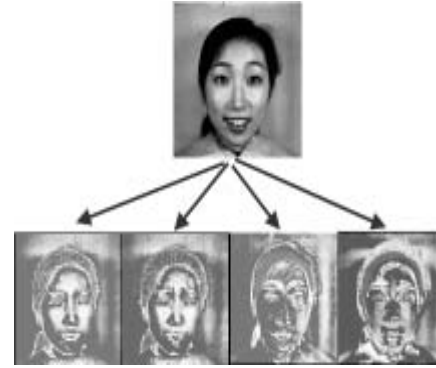


Fig 2. 2D-PCA and DiaPCA

The column-based feature vector is the input for  $SNN_1$ . The row-based feature vector is the input for  $SNN_2$ . The diagonal-based feature vector  $h$  is the input for  $SNN_3$ . The block-based feature vector is the input for  $SNN_4$ . All  $SNN(s)$  are 6 output nodes matching to 6 basic facial expression (happiness, sadness, surprise, anger, disgust, fear) [12]. Our MANN has 6  $CNN(s)$ . They give the probability of the face in six basic facial expressions. It is easy to see that to build MANN model only use Neural Network technology to develop our system.

We apply our proposal model for 6 basic facial expressions on JAFFE database consisting 213 images posed by 10 Japanese female models. The result of our experience sees below:

**Table 1. Facial Expression Classification Precision**

Classification Methods	Precision of classification
$SNN_1$	81%
$SNN_2$	79%
$SNN_3$	86%
$SNN_4$	83%
Average	89%
MANN	93%

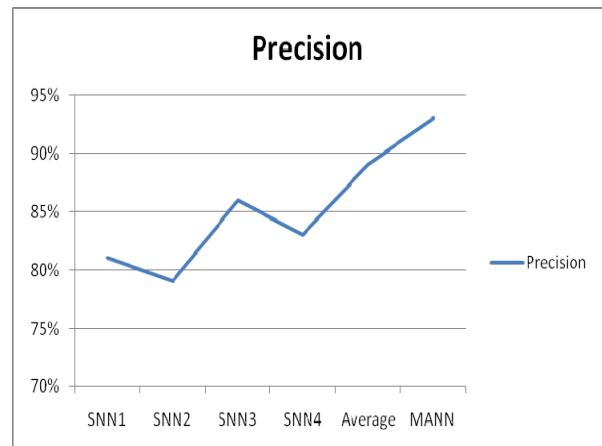


Fig 3. Facial Expression Classification Result

It is a small experimental to check MANN model and need to improve our experimental system. Although the result classification is not high, the improvement of combination result shows the MANN's feasibility such a new method combines.

We need to integrate with another facial feature sequences extraction system to increase the classification precision.

## 5. Conclusion

In this paper, we explain 2D-PCA and DiaPCA for facial feature extraction. These features are the input of our proposal model Multi Artificial Neural Network (MANN) with parameters (m, L). In particular, m is the number of images' feature vectors. L is the number of classes. MANN model has m Sub-Neural Network  $SNN_i$  ( $i=1..m$ ) and a Global Frame (GF) consisting L Components Neural Network  $CNN_j$  ( $j=1..L$ ).

Each of SNN uses to process the responsive feature vector. Each of CNN use to combine the responsive element of SNN's output vector. The weight coefficients in  $CNN_j$  are as the reliability coefficients the SNN(s)' the jth output. It means that the importance of the ever feature vector is determined after the training process. On the other hand, it depends on the image database and the desired classification. This MANN model applies for image classification.

To experience the feasibility of MANN model, in this research, we propose the MANN model with parameters ( $m=4$ ,  $L=3$ ) apply for six basic facial expressions and test on JAFFE database. The experimental result shows that the proposed model improves the classified result compared with the selection and average combination method.

## References

- [1] S. Tong, and E. Chang, "Support vector machine active learning for image retrieval", in the ninth ACM international conference on Multimedia, 2001, pp. 107-118.
- [2] R. Brown, and B. Pham, "Image Mining and Retrieval Using Hierarchical Support Vector Machines", in the 11th International Multimedia Modeling Conference (MMM'05), 2005, Vol. 00, pp. 446-451.
- [3] M. A. Turk and A. P. Penland, "Face recognition using eigenfaces", IEEE Int. Conf. of Computer Vision and Pattern Recognition, 1991, pp. 586-591.
- [4] H.T Le, "Building, Development and Application Some Combination Models of Neural Network (NN), Fuzzy Logic

- (FL) and Genetics Algorithm (GA)", PhD Mathematics Thesis, University of Science, Ho Chi Minh City, Vietnam, 2004.
- [5] H. B. Le, and H. T. Le, "the GA\_NN\_FL associated model for authenticating finger printer", in the Knowledge-Based Intelligent Information & Engineering Systems, Wellington Institute of Technology, New Zealand, 2004.
- [6] A. Ghoshal, P. Ircing, and S. Khudanpur, "Hidden Markov models for automatic annotation and content-based retrieval of images and video", in the 28th annual international ACM SIGIR conference on Research and development in information retrieval, 2005, pp. 544-551.
- [7] Y. Chen, and J. Z. Wang, "A region-based fuzzy feature matching approach to content-based image retrieval", Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2002, pp. 1252-1267
- [8] D. Hoiem, R. Sukthankar, H. Schneiderman, and L. Huston, "Object-based image retrieval using the statistical structure of images", in Computer Vision and Pattern Recognition, IEEE Computer Society Conference, 2004, Vol. 2, pp. II-490-II-497.
- [9] S. Y. Cho, and Z. Chi, "Genetic Evolution Processing of Data Structure for Image Classification", in IEEE Transaction on Knowledge and Data Engineering Conference, 2005, Vol 17, No 2, pp. 216-231
- [10] C. M. Bishop, "Pattern Recognition and Machine Learning", Springer: Press, 2006.
- [11] M. J. Lyons, J. Budynek, and S. Akamatsu, "Automatic Classification of Single Facial Images", in IEEE Transactions on Pattern Analysis and Machine Intelligence, 1999, Vol. 21, pp.1357-1362
- [12] J. Yang, D. Zhang, A. F. Frangi, and J.-y. Yang, "Two-dimensional PCA: a new approach to appearance-based face representation and recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, Vol 26, pp. 131-137, 2004
- [13] D. Zhang, Z.-H. Zhou, and S. Chen, "Diagonal principal component analysis for face recognition", Pattern Recognition, 2006, Vol. 39, pp. 140-142.

**Dr Le Hoang Thai** received B.S degree and M.S degree in Computer Science from Hanoi University of Technology, Vietnam, in 1995 and 1997. He received Ph.D. degree in Computer Science from Ho Chi Minh University of Sciences, Vietnam, in 2004. Since 1999, he has been a lecturer at Faculty of Information Technology, Ho Chi Minh University of Natural Sciences, Vietnam. His research interests include soft computing pattern recognition, image processing, biometric and computer vision. Dr. Le Hoang Thai is co-author over twenty five papers in international journals and international conferences.

**Tat Quang Phat** received B.S degree from Binh Duong University,

Vietnam, in 2007. He is currently pursuing the M.S degree in Computer Science Ho Chi Minh University of Science.

**Tran Son Hai** is a member of IACSIT and received B.S degree and M.S degree in Ho Chi Minh University of Natural Sciences, Vietnam in 2003 and 2007. From 2007-2010, he has been a lecturer at Faculty of Mathematics and Computer Science in University of Pedagogy, Ho Chi Minh city, Vietnam. Since 2010, he has been the dean of Information System department of Informatics Technology Faculty and a member of Science committee of Informatics Technology Faculty. His research interests include soft computing pattern recognition, and computer vision. Mr. Tran Son Hai is co-author of four papers in the international conferences and national conferences.



# PM<sup>2</sup>PLS: An Integration of Proxy Mobile IPv6 and MPLS

Carlos A. Astudillo<sup>1</sup>, Oscar J. Calderón<sup>1</sup> and Jesús H. Ortiz<sup>2</sup>

<sup>1</sup> New Technologies in Telecommunications R&D Group, Department of Telecommunications, University of Cauca, Popayán, 19003, Colombia

<sup>2</sup> School of Computer Engineering, University of Castilla y la Mancha Ciudad Real, Spain

## Abstract

This paper proposes a handover scheme supporting Multi-Protocol Label Switching (MPLS) in a Proxy Mobile IPv6 (PMIPv6) domain that improves the mobility and gives Quality of Service (QoS) and Traffic Engineering (TE) capabilities in wireless access networks. The proposed scheme takes advantages of both PMIPv6 and MPLS. PMIPv6 was designed to provide Network-based Localized Mobility Management (NETLMM) support to a Mobile Node (MN); therefore, the MN does not perform any mobility related signaling, while MPLS is used as an alternative tunneling technology between the Mobile Access Gateway (MAG) and the Local Mobility Anchor (LMA) replacing the IP-in-IP tunnels with Label Switched Path (LSP) tunnels. It can also be integrated with other QoS architectures such as Differentiated Services (DiffServ) and/or Integrated Services (IntServ). In this study, we used MATLAB to perform an analysis to evaluate the impact of introducing MPLS technology in PMIPv6 domain based on handover latency, operational overhead and packet loss during the handover. This was compared with PMIPv6, and a PMIPv6/MPLS integration. We proved that the proposed scheme can give better performance than other schemes.

**Keywords:** Localized Mobility Management, MPLS, PMIPv6, PMIPv6/MPLS, PM<sup>2</sup>PLS.

## 1. Introduction

Some host-based mobility management protocols such as Mobile IPv6 (MIPv6) [1] and its extensions (i.e. Hierarchical Mobile IPv6 (HMIPv6) [2] and Fast Handover in Mobile IPv6 (FMIPv6) [3]) have been standardized by the Internet Engineering Task Force (IETF) for Internet mobility support, but they have not widely deployed in real implementations [4]. One of the most important obstacles in order to deploy mobility protocols is the modification that must be done in the terminal (Mobile Host - MH). Proxy Mobile IPv6 has been proposed by the IETF NETLMM working group as a network-based mobility management protocol [5]. It allows the communication between the Mobile Node and the Correspondent Node (CN) while MN moves without

its participation in any mobility signaling. On the other hand, Multiprotocol Label Switching is a forwarding technology that supports Quality of Service and Traffic Engineering capabilities in IP networks [6]. Furthermore, it provides fast and efficient forwarding by using labels swapping instead of IP forwarding.

MPLS is being used by most network operators to carry IP traffic. Introduce network-based mobility capabilities in MPLS networks can be useful [7].

There are few works that have handled the integration of PMIPv6 and MPLS. Recently, an IETF Internet Draft proposed MPLS tunnels (LSP tunnels) as an alternative to IP-in-IP tunnel between Local Mobility Anchor (LMA) and Mobile Access Gateway (MAG) [7]. The draft specifies two different labels: a classic MPLS label and Virtual Pipe (VP) labels as a way to differentiate traffic in the same tunnel. The authors focus on the management of VP labels rather than classic MPLS labels. The authors assume that there are LSPs established between the MAG and the LMA and use two labels for each packet; both labels are pushed by the Label Edge Router (LER).

But, as mentioned in [8], the use of VP label is not strictly necessary because this label is only used to eliminate the necessity of the LMA to look up the network layer header in order to send packets to the CN. It adds 4 overhead bytes (VP label size) to the LSP tunnel (8 overhead bytes in total). Reference [8] makes a study of PMIPv6/MPLS on Wireless Mesh Network (WMN) with and without VP labels in terms of handover delay and operation overhead. Reference [9] makes a study in an Aeronautical Telecommunication Network (ATN) and uses VP labels in the same way of [7]. Reference [10] makes a quantitative and qualitative analysis of the PMIP/MPLS integration and other schemes, but they do not give details about design considerations, label management or architecture operation.

This work proposes an integration of PMIPv6 and MPLS called PM<sup>2</sup>PLS. The integration is done in an overlay way [11] and the relationship between binding updates and LSPs setup is sequential. We do not consider necessary to use VP label since this label only divided traffic from



different operators (its use is optional). We use Resource Reservation Protocol – Traffic Engineering (RSVP-TE) [12] as label distribution protocol to establish a “bidirectional LSP” between the LMA and the MAG. Since a LSP in MPLS is unidirectional, we call “bidirectional LSP” to two LSP that do not necessarily follow the same upstream and downstream path but that the ingress Label Switch Router (LSR) in the LSP upstream is the egress LSR in the LSP downstream and vice versa. In future works, we want to integrate PM<sup>2</sup>PLS and QoS architectures such as IntServ and/or DiffServ in order to assure QoS in a mobility enabled MPLS access network where the MN is not based on MIPv6.

The rest of the paper is organized as follows. Section 2 presents an overview about PMIPv6 and MPLS. Section 3 introduces the PMIPv6/MPLS integration called PM<sup>2</sup>PLS. Section 4 shows the performance analysis of PM<sup>2</sup>PLS on 802.11 access network based on handover latency, operational overhead and packet loss during handover. Finally, we conclude in Section 5.

## 2. Background

### 2.1 Proxy Mobile IPv6

PMIPv6 was designed to provide network-based mobility support to a MN in a topologically localized domain [5]; this means that the CN is exempted to participate in any mobility related signaling and all mobility control functions shift to the network. In this context, PMIPv6 defined two new entities called Local Mobility Anchor and Mobile Access Gateway. The function of LMA is to maintain reachability to the MN and it is the topological anchor point for the MN’s home network prefix(es), this entity has a Binding Cache (BC) that links the MN with its current Proxy CoA (MAG’s address). MAG runs in the Access Router (AR) and is responsible for tracking the mobile node’s movements at the access link and for initiating binding registrations to the LMA; it also establishes a bidirectional tunnel with the LMA to enable the MN to use an address from its home network prefix (MN-HNP) and emulates the MN’s home link. This entity has a Binding Update List (BUL) which contains the MNs attached to it, and their corresponding LMAA (LMA’s address). Figure 1 shows a common PMIPv6 scenario with LMAs, MAGs, MNs, CN, tunnels between LMA and MAG and data flow.

In a PMIPv6 domain, the options for establishing the tunnel between LMA and MAG are as follows: IPv6-In-IPv6 [5], Generic Routing Encapsulation (GRE), IPv6-In-IPv4 or IPv4-In-IPv4 [13].

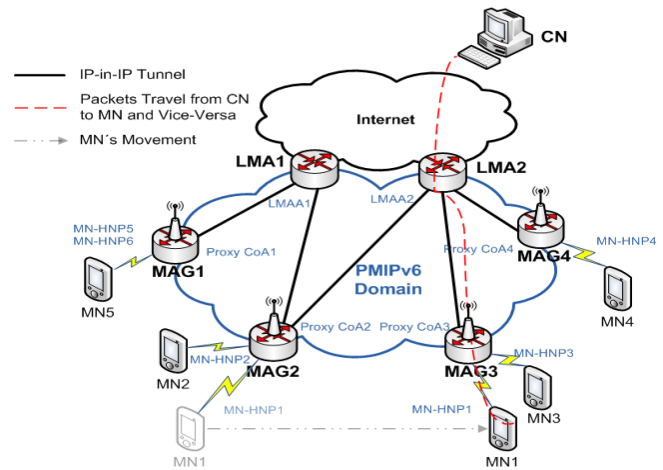


Fig. 1 PMIPv6 scenario.

### 2.1 Multi-Protocol Label Switching

Conventional IP forwarding mechanisms are based on network reachability information. As a packet traverses the network, each router uses the IP header in the packet to obtain the forwarding information. This process is repeated at each router in the path, so the optimal forwarding is calculated again and again. MPLS [6] is a forwarding packets paradigm integrated with network-layer routing. It is based on labels that assign packet flows to a Forwarding Equivalent Class (FEC). FEC has all information about the packet (e.g. destination, precedence, Virtual Private Network (VPN) membership, QoS information, route of the packet, etc.), once a packet is assigned to a FEC no further analysis is done by subsequent routers, all forwarding is driven by the labels. All packets with the same FEC use the same virtual circuit called Label Switched Path (LSP). To deploy MPLS in an IP network, a label header is inserted between layer two and layer three headers as shown in Figure 2. The MPLS header is composed by: 20-bit label field, 3-bit initially defined as EXPerimental and current used as Traffic Class (TC) field [15], 1-bit Bottom of Stack (S) field, and 8-bit Time to Live (TTL) field. MPLS also offers a traffic engineering capabilities that provides better use of the network resources.

MPLS consists of two fundamentals components: The FEC-to-NHLFE mapping (FTN) which forwards unlabeled packets, this function is running in the ingress router (LER, Label Edge Router) and mapping between IP packets and FEC must be performed by the LER. And the Incoming Label Mapping (ILM) that makes a Label-to-NHLFE mapping to forward labeled packets.

The RFC 3031 defines a “LSP Tunnel” as follows: “It is possible to implement a tunnel as a LSP, and use label switching rather than network layer encapsulation to cause

the packet to travel through the tunnel” [6]. The packets that are sent through the LSP tunnel constitute a FEC.

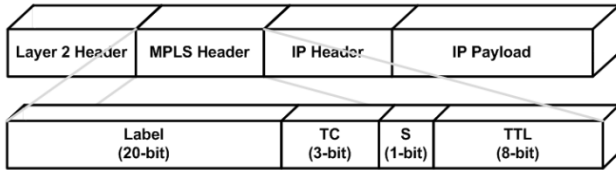


Fig. 2 MPLS header format.

### 3. PMIPv6 and MPLS Integration

We propose a PMIPv6/MPLS architecture called PM<sup>2</sup>PLS. First, we give previous concepts on the integration of MPLS and MIPv6 (and its extensions), then, we describe the design considerations, MAG and LMA operation and finally, the signaling flow between components is described.

#### 3.1 Previous Concepts

Previous works on integrating MIPv6, HMIPv6 and/or FMIPv6 in MPLS networks consider two models for doing that: integrated or overlay [11]. In the integrated model, some processes are united; in the overlay one, processes and information are separated as long as possible. We choose to use the overlay model since it allows an easy integration with current deployed MPLS networks. Another important item in previous integrations is the relationship between binding updates and LSPs setup. There are two proposes. The first one is to make the LSP setup in an encapsulated way [11] which means that the LSP establishment is initialized after a Binding Update (BU) message arrives to the Home Agent (HA), Mobility Anchor Point (MAP) or Regional Gateway (RG) but the Binding Acknowledgment (BA) is sent after a LSP setup process is finished. The other method is called “sequential” where the LSP setup is initialized after a successful binding update process finished [11]. It means that the LSP setup is initialized when a BA message arrives to CN, Foreign Agent (FA) or Access Router (AR). Reference [11] concluded that sequential way has better handover performance than encapsulated one. In our scheme the relationship between binding updates and LSP setup can be viewed as “sequential”, but we optimized the LSP setup since the process is initialized in the LMA after the Proxy Binding Update (PBU) message has been accepted and Proxy Binding Acknowledgment (PBA) message sent, it does not wait for PBA arrives to the MAG since we consider that it is not necessary.

#### 3.2 Design Considerations

We give the design considerations for the PM<sup>2</sup>PLS architecture in this subsection.

- We used LSP tunnels as specified in [6], [12]. The LSP Tunnel must be “bidirectional” between MAG and LMA (two LSP Tunnels established by RSVP-TE, one from LMA to MAG and other between MAG and LMA). Note that the upstream LSP not necessarily follows the same path that downstream LSP. This “bidirectional” LSP Tunnel must be used for forwarding the mobile nodes’ data traffic between MAG and LMA. It can also be used for sending PBU and PBA between MAG and LMA.
- The LSP setup could be pre-established or dynamically assigned. In a dynamic way, the LSP would be setup only once, when the first MN arrives to specific MAG, the follows MNs can use the established LSP, if it is necessary to re-evaluated the LSP capabilities, it should be performed by RSVP-TE techniques. It also improves the Proxy Binding Update and Proxy Binding Acknowledgment messages delivery of sub-sequence location updates.
- The introduction of network-based mobility in MPLS networks should be in an overlay way. It means that data base will not be integrated between PMIPv6 and MPLS. The BC, BUL and the Label Forwarding Information Base (LFIB) should be maintained separately. But a relationship between processes sequence should be performed and the information should be shared.
- The MN should be IPv6-Base. We only consider the use of IPv6 MN-HoA since the process of address configuration in IPv4 is too large, instead IPv6 supports stateless address configuration.
- The Transport Network could be IPv6 or IPv4.
- The traffic in the same MAG is managed for itself.
- The wireless access network that we consider in this study is 802.11. It is necessary to define the Access Network (AN) type because of the analysis that will be described, but it does not imply that others access technologies as Long Term Evolution (LTE), WiMax or 3G Networks couldn’t be used with PM<sup>2</sup>PLS.
- This architecture cannot support multicast traffic.
- Penultimate hop popping is desirable. It should be used, since the packet processing at the last hop (in the MPLS domain) would be optimized. It avoids double processing in the last hop (i.e. MPLS and IP header processing).
- Label merging and aggregation are undesirable. Those constraints allow having unique label per LSP and more than one LSP for the same FEC, respectively (e.g. it is useful when we want to introduce load balancing between the LMA and a specific MAG).

### 3.3 Architecture Components

The architecture components shown in Figure 3 are described. Figure 4 gives the protocol stack of PM<sup>2</sup>PLS entities and the signaling flow between them when a handover occurs is shown in Figure 5.

- MAG/LER: It is an entity which has the MAG (from PMIPv6) and LER (from MPLS) functionality inside its protocol stack.
- LMA/LER: It is an entity which has the LMA (from PMIPv6) and LER (from MPLS) functionality inside its protocol stack.
- LSR: It is a MPLS router as specified in [6].
- MN: It is a mobile node which implements IPv6.
- CN: It is a mobile/fixed node which implements IPv6 or IPv4.

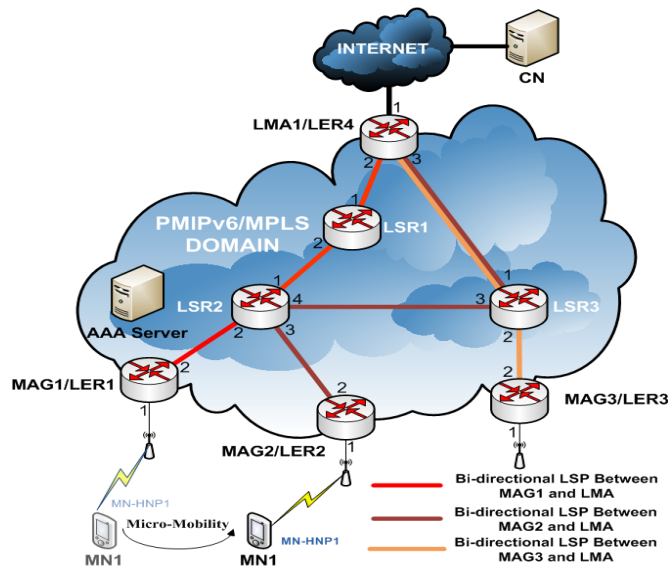


Fig. 3 PM<sup>2</sup>PLS scenario.

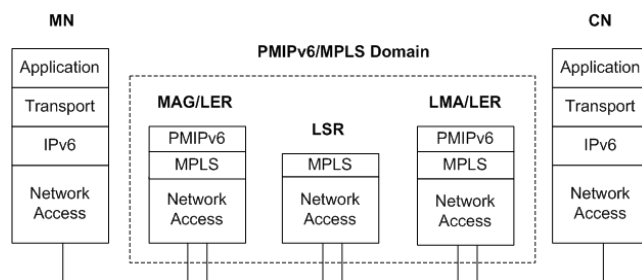


Fig. 4 Protocols stack of PM<sup>2</sup>PLS components

### 3.4 LMA/LER Operation

When a PBU message is received by the LMA, it processes the message as specified in [5], after PBU is accepted and the PBA is sent, immediately the LMA

verifies if it is assigned the MN's PCoA to a FEC (there are LSP tunnel between LMA and MN's MAG). If an entry already exists with the MN-PCoA as FEC, it does not need to setup the LSP, since a LSP Tunnel already exists, If not a RSVP Path message are generated from LMA to MAG to setup the LSP between LMA and MAG. When the LSP setup process is finished (Path and Resv RSVP messages are received and processed) and the LMA had assigned a label to that FEC, it should have a entry in the LFIB with the FEC assign to the tunnel between LMA and MAG. Periodically, the LSP capability should be evaluated in order to assure that the traffic across the LSP is being satisfied.

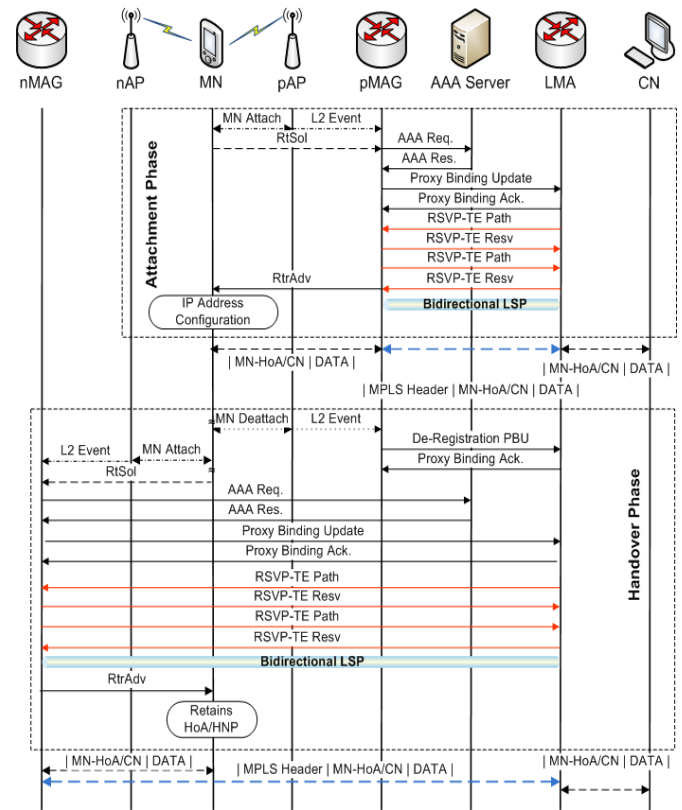


Fig. 5 Signaling flow in PM<sup>2</sup>PLS.

### 3.5 MAG/LER Operation

When a PBA message is received by the MAG with a status field set to zero (accepted), it processes the message in the same way as specified in [5], and then a RSVP Path message is generated from MAG to LMA to setup the LSP between MAG and LMA. If an entry already exists with MN's LMA as a FEC, it does not need to setup the LSP, since it already exists. Periodically, the LSP capability should be evaluated in order to assure that the traffic across the LSP is being satisfied.

### 3.6 Handover Procedure

When roaming for first time in a PMIPv6/MPLS domain, the MN obtains a MN-HoA based on its HNP and keeps it as long as stays in the PMIPv6 domain. This means that the MN only executes the address configuration and Duplicate Address Detection (DAD) once.

The handover process in PM<sup>2</sup>PLS scenario is as follows. When the MN moves from a MAG/LER to another MAG/LER in the same domain, first the MN detaches from a Access Point (AP) in a previous MAG/LER (pMAG/LER) area and attaches to a AP in new MAG/LER (nMAG/LER) area, at this moment nMAG/LER knows the MN-ID and other information by layer 2 procedures (Note that in PMIPv6 it is not necessary to wait for a Router Solicitation message (RtSol), this message can be sent by the MN at any time during the handover process). nMAG/LER performs a MN's authentication, and then sends a PBU to the LMA. Upon receiving the PBU message, the LMA follows the procedure described in section 3.4, it generates a PBA messages and if it is necessary to send RSVP Path message. The MAG on receiving the PBA message follows the procedure described in section 3.5. It updates its Binding Update List and sends a RSVP-Path if it is necessary. Finally, the sends a Router Advertisement (RtrAdv) message containing the MN's HNP, and this will ensure the MN will not detect any change with respect to the layer 3 attachment of its interface (it retains the configured address).

### 3.7 Example of LFIBs in PM<sup>2</sup>PLS Nodes

Based on Figure 3, we give an example of the Label Forwarding Information Base (LFIB) of each node in the PM<sup>2</sup>PLS scenario. In this example, we use penultimate hop popping and assume that the upstream LSP has the same path (the same nodes) of the downstream LSP. We show the content of the LFIB in LMA1/LER4 (Table 1), MAG1/LER1 (Table 2), MAG2/LER2 (Table 3), MAG3/LER3 (Table 4), LSR1 (Table 5), LSR2 (Table 6), and LSR3 (Table 7).

## 4. Performance Analysis

In this section we analyze the performance of PM<sup>2</sup>PLS on 802.11 Wireless LAN (WLAN) access network based on handover delay, attachment delay, operational overhead and packet loss during handover. We compared our proposal with single PMIPv6 and PMIPv6/MPLS in an encapsulated way as proposed in [8].

Table 1: LMA1/LER4's LFIB

FEC	In Label	In IF	Out Label	Out IF
LMA-MAG1	-	-	20	2
LMA-MAG2	-	-	22	3
LMA-MAG3	-	-	27	3

Table 2: MAG1/LER1's LFIB

FEC	In Label	In IF	Out Label	Out IF
MAG1-LMA	-	-	40	2

Table 3: MAG2/LER2's LFIB

FEC	In Label	In IF	Out Label	Out IF
MAG2-LMA	-	-	55	2

Table 4: MAG3/LER3's LFIB

FEC	In Label	In IF	Out Label	Out IF
MAG3-LMA	-	-	60	2

Table 5: LSR1's LFIB

FEC	In Label	In IF	Out Label	Out IF
LMA-MAG1	20	1	15	2
MAG1-LMA	35	2	-	1

Table 6: LSR2's LFIB

FEC	In Label	In IF	Out Label	Out IF
LMA-MAG1	15	1	-	2
MAG1-LMA	40	2	35	1
LMA-MAG2	32	4	-	3
MAG2-LMA	55	3	50	4

Table 7: LSR3's LFIB

FEC	In Label	In IF	Out Label	Out IF
LMA-MAG2	22	1	32	3
MAG2-LMA	50	3	-	1
LMA-MAG3	27	1	-	2
MAG3-LMA	60	2	-	1

### 4.1 Handover Process in 802.11

In order to study the handover performance of PM<sup>2</sup>PLS, we consider an 802.11 WLAN access to calculate the L2 handover delay (that is when a MN attaches to a new Access Point (AP)). During the handover at layer two, the station cannot communicate with its current AP. The IEEE 802.11 handover procedure involves at least three entities: the Station (MN in PM<sup>2</sup>PLS), the Old AP and the New AP. It is executed in three phases: Scanning (Active or



Passive), Authentication and Re-association as shown in Figure 6 [16]. The scanning phase in a handover process is attributed to mobility, when signal strength and the signal-to-noise ratio are degraded the handover starts. At this point, the client cannot communicate with its current AP and it initializes the scanning phase. There are two methods in this phase: Active and Passive. In the passive method the station only waits to hear periodic beacons transmitted by neighbour APs in the new channel, in the active one, the station also sends probe message on each channel in its list and receives response of APs in its coverage range. When the station finds a new AP, it sends an authentication message, and once authenticated can send the re-association message. In this last phase includes the IAPP (Inter Access Point Protocol) [17] procedure to transfer context between Old AP and New AP.

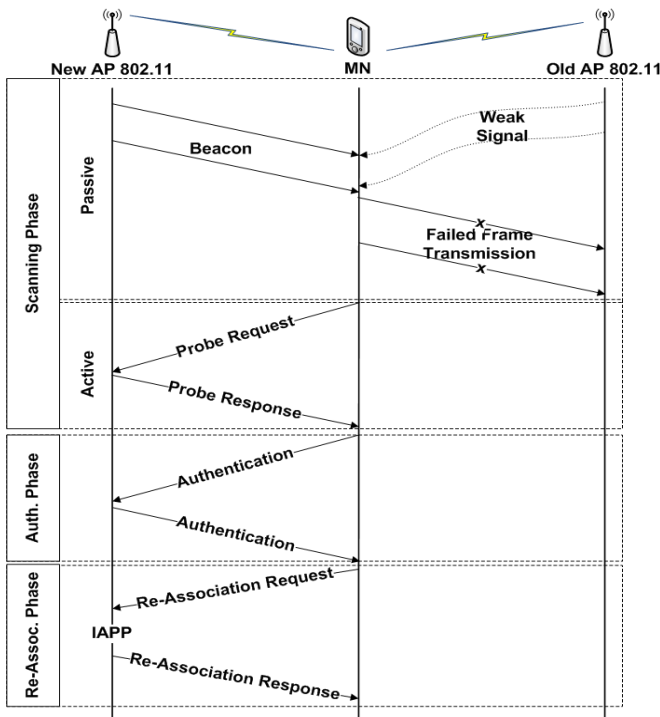


Fig. 6 802.11 handover process

## 4.2 Total Handover Delay

In this subsection we analyze the delay performance of the handover process for our PMIPv6/MPLS integration. The impact of handover on ongoing sessions is commonly characterized by handover delay, especially when we work with real time applications (e.g. Voice over IP, Video over Demand or IPTV) which are sensitive to packet delay and have important requirements of interruption time. For convenience, we define the parameters described in Table 8.

Table 8: Parameter descriptions/settings

Parameter	Description	Value
$\alpha_{RP}$	IP router processing time.	0.2 ms
$\alpha_{AAA-Server}$	Processing time of AAA Server.	0.1 ms
$t_{x,y}$	Time required for a message to pass through links from node x to node y.	N/A
$t_{WL}$	Wireless link delay.	10 ms [4]
$t_{Scanning}$	Delay due to scanning phase of 802.11.	100 ms [16]
$T_{REG}$	Registration or binding update delay.	N/A
$t_{PBU}$	Time of Proxy Binding Update message	N/A
$t_{PBA}$	Time of Proxy Binding Acknowledgment message	N/A
$T_{MD}$	Mobility detection delay.	0 ms
$T_{L3HO}$	L3 handover delay.	N/A
$T_{L2HO}$	L2 handover delay.	115 ms [4]
$T_{HO}$	Total handover delay.	N/A
$T_{Bi-LSP-Setup}$	Delay due to bidirectional LSP setup.	N/A
$t_{Authentication}$	Delay due to 802.11 authentication phase.	5 ms [16]
$t_{Association}$	Delay due to 802.11 association phase.	10 ms [16]
$t_{AP-MAG}$	The delay between the AP and the MAG.	2 ms [4]
$t_{AAA-Resp}$	Delay due to AAA response message.	1 ms
$t_{AAA-Req.}$	Delay due to AAA request message.	1 ms
$T_{AAA}$	Delay due to AAA procedure.	3 ms [4]
$n, m$	Number of hops between MAG-LMA and LMA-MAG respectively.	1-15
$\beta_{RP}$	LSR processing time.	0.1 ms
$\beta_{MAG}$	Processing time of MAG/LER router.	0.2 ms
$\beta_{LMA}$	Processing time of LMA/LER router.	0.5 ms
$\alpha_{MAG}$	Processing time of MAG router.	0.2 ms [18]
$\alpha_{LMA}$	Processing time of LMA router.	0.5 ms [18]
$D_{U1}$	Upstream delay propagation in link 1.	2 ms
$D_{Dk}$	Downstream delay propagation in link k.	2 ms
$\lambda_{PR}$	Send packet ratio	170 packets/sec [19]

The general equation of the total handover delay in a Mobile IP protocols can be expressed as:

$$T_{HO} = T_{L2HO} + T_{MD} + T_{L3HO} \quad (1)$$

$T_{MD}$  is the interval from when an MN finishes Layer 2 handover to when it begins Layer 3 handover. In PM<sup>2</sup>PLS as in PMIPv6, as soon the MN is detected by the MAG with a L2 trigger, the L3 handover is initialized, so  $T_{MD}$  can be considered zero.

$T_{L3HO}$  in PM<sup>2</sup>PLS when a bidirectional LSP exists between MAG and LMA can be expressed as:

$$T_{L3HO} = T_{AAA} + T_{REG} + T_{RA} \quad (2)$$

where the AAA process delay is as follows:

$$T_{AAA} = t_{AAA-Req.} + t_{AAA-Resp.} + \alpha_{AAA-Server}, \quad (3)$$

the binding update delay can be expressed as:

$$T_{REG} = t_{PBU} + t_{PBA} + \beta_{LMA} + \beta_{MAG} \quad (4)$$

where

$$t_{PBU} = t_{MAG,LMA} + (n) \beta_{RP} \quad (5)$$

$$t_{MAG,LMA} = \sum_{k=1}^n D_{Dk} \quad (6)$$

$$t_{PBA} = t_{LMA,MAG} + (m) \beta_{RP} \quad (7)$$

$$t_{LMA,MAG} = \sum_{l=1}^m D_{Ul} \quad (8)$$

finally,

$$T_{REG} = \sum_{k=1}^n D_{Dk} + \sum_{l=1}^m D_{Ul} + (n+m) \beta_{RP} + \beta_{LMA} + \beta_{MAG}. \quad (9)$$

When a bidirectional LSP is not established between MAG and LMA  $T_{L3HO}$  can be calculated as follows:

$$T_{L3HO} = T_{AAA} + T_{REG} + T_{Bi-LSP-Setup} + T_{RA}, \quad (10)$$

where  $T_{AAA}$  is the same as in (3),  $T_{RA}$  is the same as in (16), and from (9)  $T_{REG}$  can be expressed as:

$$T_{REG} = \sum_{k=1}^n D_{Dk} + \sum_{l=1}^m D_{Ul} + (n+m) \alpha_{RP} + \alpha_{LMA} + \alpha_{MAG}. \quad (11)$$

The latency introduced by LSP setup between the LMA and the MAG and vice versa ( $T_{Bi-LSP-Setup}$ ) in  $PM^2PLS$  can be expressed as the delay of one LSP setup, since the LMA initializes LSP setup between LMA and MAG after accepting PBU and sending PBA to the MAG (The LMA does not need to wait nothing else). When PBA arrives to the MAG, it initializes the LSP setup with LMA. We assume that when a LSP setup between MAG and LMA finishes, the LSP between LMA and MAG is already established, since it initialized before MAG to LMA LSP:

$$T_{Bi-LSP-Setup} = t_{RSVP-Resv} + t_{RSVP-Path} \quad (12)$$

where

$$t_{RSVP-Resv} = t_{MAG,LMA} + (n) \alpha_{RP}, \quad (13)$$

$$t_{RSVP-Path} = t_{LMA,MAG} + (m) \alpha_{RP}, \quad (14)$$

$t_{MAG,LMA}$  and  $t_{LMA,MAG}$  are as in (6) and (8) respectively. Finally,  $T_{Bi-LSP-Setup}$  can be expressed as:

$$T_{Bi-LSP-Setup} = \sum_{k=1}^n D_{Dk} + \sum_{l=1}^m D_{Ul} + (n+m) \alpha_{RP}. \quad (15)$$

The delay by router advertisement message can be expressed as:

$$T_{RA} = t_{AP-MAG} + t_{WL}. \quad (16)$$

The L2 handover delay in an 802.11 WLAN access network can be expressed as:

$$T_{L2HO} = t_{Scanning} + t_{Authentication} + t_{Association} \quad (17)$$

$T_{L3HO}$  in PMIPv6 is as in (2), with  $T_{AAA}$  as in (3),  $T_{REG}$  as in (11) and  $T_{RA}$  as in (16). As mentioned above during a PMIPv6 handover is not executed neither Movement Detection (MD) nor Address Configuration (Included DAD).

### 4.3 Packet Loss During Handover

Packet Loss (PL) is defined as the sum of lost packets per MN during a handover. With (20) we can calculate the PL in a handover for a given MN.

$$PL_{PM^2PLS} = T_{PM^2PLS HO} * \lambda_{PR} \quad (20)$$

### 4.4 Operational Overhead

The operational overhead of  $PM^2PLS$  is 4 bytes per packet (MPLS header size).  $PM^2PLS$  reduces significantly the operational overhead with respect to PMIPv6 which has an operational overhead of 40 bytes when uses IPv4 or IPv6 in IPv6 encapsulation (over IPv6 Transport Network), 20 bytes of overhead when uses IPv4 or IPv6 in IPv4 encapsulation (over IPv4 Transport Network), 44 bytes when uses GRE tunnel over TN IPv6, or 24 bytes when uses GRE tunnel over IPv4 TN. A comparison of operational overhead between above schemes is summarized in Table 9.

Table 9: Operational Overhead

Scheme and Tunneling Mechanism	Overhead per Packet	Description
PMIPv6 with IPv6 in IPv6 Tunnel	40	IPv6 header
PMIPv6 with IPv4 in IPv6 Tunnel	40	IPv6 header
PMIPv6 with IPv6 in IPv4 Tunnel	20	IPv4 header
PMIPv6 with IPv4 in IPv4 Tunnel	20	IPv4 header
PMIPv6 with GRE encapsulation (over TN IPv6)	44	IPv6 header + GRE header
PMIPv6 with GRE encapsulation (over TN IPv4)	24	IPv4 header + GRE header
PMIPv6/MPLS with VP Label (over TN IPv4 or IPv6)	8	2 MPLS headers
$PM^2PLS$ (over TN IPv4 or IPv6)	4	MPLS headers

### 4.5 Simulation Results

We compared  $PM^2PLS$ , PMIPv6 [5] and PMIPv6/MPLS as proposed in [8]. We use typical values for parameters involved in above equations as shown in Table 8. Figure 6 shows the impact of hops between the MAG and the LMA in the handover delay. It can be observed that the handover delay increases with the number of hops. PMIPv6/MPLS is the scheme most affected by the number of hops because it integrates the LSP setup in encapsulated way and does not optimize this process. PMIPv6 and  $PM^2PLS$  with a bidirectional LSP established between new MAP

and LMA shown a comparable performance with slightly better response of PM<sup>2</sup>PLS when the number of hops increase because binding update messages (i.e. PBU and PBA) are sent through bidirectional LSP established between the MAG and the LMA instead of using IP forwarding. Figure 7 shows the total packet loss during handover for above schemes. Since packet loss during handover is proportional to the handover latency, PM<sup>2</sup>PLS also have the lowest packet loss ratio between compared schemes. For doing the packet loss simulation we consider a flow of VoIP [19].

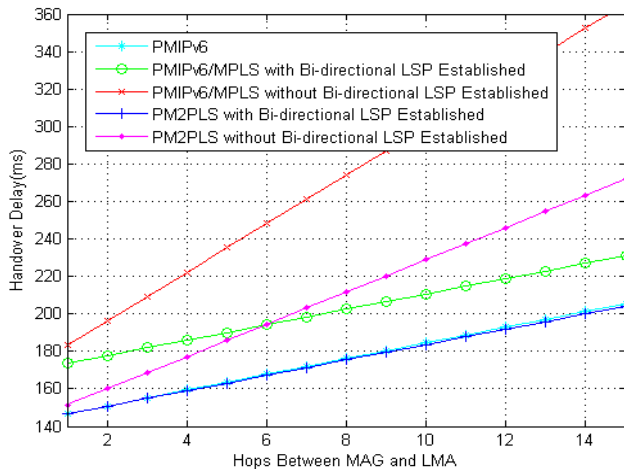


Fig. 7 802.11 handover process

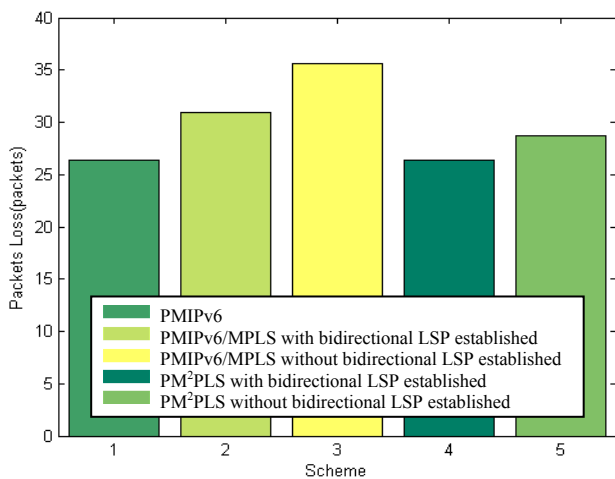


Fig. 8 Packet loss of PMIPv6, PMIPv6/MPLS, and PM2PLS during a handover.

## Conclusions

We proposed an integration of MPLS and PMIPv6 called PM<sup>2</sup>PLS which optimizes the bidirectional LSP setup by integrating binding updates and bidirectional LSP setup in

an optimized sequential way; we also used the LSP established between the MAG and the LMA for sending PBU and PBA messages when it exists. We compared the performance of PM<sup>2</sup>PLS with single PMIPv6 and PMIPv6/MPLS as specified in [8]. We demonstrated that PM<sup>2</sup>PLS has a lower handover delay than PMIPv6/MPLS, and slightly lower than the one of PMIPv6. The operational overhead in MPLS-based schemes is lower than single PMIPv6 schemes since uses LSPs instead of IP tunnelling. With MPLS integrated in a PMIPv6 domain, the access network can use intrinsic Quality of Service and Traffic Engineering capabilities of MPLS. It also allows the future use of DiffServ and/or IntServ in a PMIPv6/MPLS domain.

## Acknowledgments

This work was sponsored by the Colombian Institute of Science and Technology (COLCIENCIAS), <http://www.colciencias.gov.co/> through the national program Young Researchers and Innovators "Virginia Gutiérrez de Pineda". We would like to thank MSc. Victor M. Quintero for his useful comments in the preliminary version of this paper published in the NTMS 2011.

## References

- [1] Johnson D., Perkins C., and Arkko J., "Mobility Support in IPv6," IETF RFC 3775 (Proposed Standard), June 2004.
- [2] Soliman H., Castellucia C., ElMalki K., and Bellier L., "Hierarchical Mobile IPv6 (HMIPv6) Mobility Management," IETF RFC 5380 (Proposed Standard), October 2008.
- [3] Koodli R., "Mobile IPv6 Fast Handovers," IETF RFC 5568 (Proposed Standard), July 2009.
- [4] Kong K.-S., Lee W., Han Y.-H., Shin M.-k., and You H., "Mobility Management for All-IP Mobile Networks: Mobile IPv6 vs. Proxy Mobile IPv6," *IEEE Wireless Communications*, pp. 36-45, April 2008.
- [5] Gundavelli S., Leung K., Devarapalli V., and Chowdh K., "Proxy Mobile IPv6," IETF RFC 5213 (Proposed Standard), August 2008.
- [6] Rosen E., Viswanathan A., and Callon R., "Multiprotocol Label Switching Architecture," IETF RFC 3031 (Proposed Standard), January 2001.
- [7] Xia F. and Sarikaya B., "MPLS Tunnel Support for Proxy Mobile IPv6," IETF Draft, October 25, 2008.
- [8] Garroppo R., Giordano S., and Tavanti L., "Network-based micro-mobility in wireless mesh networks: is MPLS convenient," in *Proceedings of Global Communications Conference (GLOBECOM)*, December 2009.
- [9] Liang J. Z., Zhang X., and Li Q., "A Mobility Management Based on Proxy MIPv6 and MPLS in Aeronautical Telecommunications Network," in *Proceedings of 2009 First International Conference on Information Science and Engineering (ICISE)*, 2009, pp. 2452-2455.



- [10] Carmona-Murillo J., González-Sánchez J. L., and Cortés-Polo D., "Mobility management in MPLS-based access networks. An analytical study," in *Proceedings of IX Workshop in MPLS/GMPLS networks*, July 2009.
- [11] Vassiliou V., "Design Considerations for Introducing Micromobility in MPLS," in *Proceedings of the 11th IEEE Symposium on Computers and Communications (ISCC'06)*, June 2006.
- [12] Awduche D., et al., "RSVP-TE: Extensions to RSVP for LSP Tunnels," IETF RFC 3209 (Proposed Standard), December 2001.
- [13] Muhanna A., Khalil M., Gundavelli S., and Leung K., "Generic Routing Encapsulation (GRE) Key Option for Proxy Mobile IPv6," IETF RFC 5845 (Proposed Standard), June 2010.
- [14] Wikikawa R. and Gunsavelli S., "IPv4 support for proxy Mobile IPv6," IETF Draft, May 2008.
- [15] Andersson L. and Asati R., "Multiprotocol Label Switching (MPLS) Label Stack Entry: "EXP" Field Renamed to "Traffic Class" Field," IETF RFC 5462 (Proposed Standard), February 2009.
- [16] Mishara A., Shin M., and Arbaugh W., "An Empirical Analysis of the IEEE 802.11 MAC Layer Handoff Process," *ACM SIGCOMM Computer Communication Review*, vol. 33, no. 2, pp. 93-102, April 2003.
- [17] IEEE Trial-Use Recommended Practice for Multi-Vendor Access Point Interoperability Via an Inter-Access Point Protocol Across Distribution Systems Supporting IEEE 802.11 Operation, IEEE Std 802.11F-2003.
- [18] Diab A., Mitschele-Thiel A., Getov K., and Blume O., "Analysis of Proxy MIPv6 Performance compared to Fast MIPv6," in *Proceedings of 33rd IEEE Conference on Local Computer Networks 2008 (LCN 2008)*, pp. 579-580, 2008.
- [19] Hasib M., "Analysis of Packet Loss Probing in Packet Networks," Doctoral Thesis, Queen Mary, University of London, June 2006.

**Carlos A. Astudillo** received his B.Sc. degree in Electronics and Telecommunications Engineering from the University of Cauca, Popayán, Colombia, in 2009. In 2010, he got a scholarship from the national program Young Researcher and Innovators "Virginia Gutiérrez de Pineda" of the Colombian Institute of Science and Technology - COLCIENCIAS. He is member of the New Technologies in Telecommunications R&D Group in the same University. Currently, he is a master student in Computer Science at State University of Campinas, Campinas, Brazil. His research interests are Mobility and Quality of Service in Wired/Wireless Networks and Policy-Based Network Management.

**Oscar J. Calderón** received his B.Sc. degree in Electronics and Telecommunications Engineering from the University of Cauca, Popayán, Colombia in 1996. He holds a specialist degree in Telematics Networks and Services (1999) and the Diploma of Advanced Studies (DEA) from the Polytechnic University of Catalonia (2005), Spain. He is full-professor and head of the Department of Telecommunications in the University of Cauca. He is member of the New Technologies in Telecommunications R&D Group in the same University. His research interests are Quality of Service in IP Networks, NGN.

**Jesús H. Ortiz** received his BSc. in Mathematical from the Santiago de Cali University, Colombia, Bsc. in Electrical Engineering from the University of Valle, Colombia and his PhD degree in Computer Engineering from the University of Castilla y la Mancha, Spain, in 1988, 1992 and 1998 respectively. Currently, he is assistant professor in the Universidad of Castilla y la Mancha, Spain in the area of Computer and Mobile Networks. He is reviewer and/or editor of several journals such as IAJIT, IJRRCS, IJCNIS, JSAT, and ELSEVIER.

# Language Identification of Web Pages Based on Improved N-gram Algorithm

Yew Choong Chew<sup>1</sup>, Yoshiki Mikami<sup>2</sup>, Robin Lee Nagano<sup>3</sup>

<sup>1</sup> Information Science and Control Engineering, Nagaoka University of Technology  
Nagaoka, Niigata 940-2188, Japan

<sup>2</sup> Information Science and Control Engineering, Nagaoka University of Technology  
Nagaoka, Niigata 940-2188, Japan

<sup>3</sup> Foreign Language Education Center, University of Miskolc  
Miskolc, Egyetemvaros H3515 Hungary

## Abstract

Language identification of written text in the domain of Latin-script based languages is a well-studied research field. However, new challenges arise when it is applied to non-Latin-script based languages, especially for Asian languages' web pages. The objective of this paper is to propose and evaluate the effectiveness of adapting Universal Declaration of Human Rights and Biblical texts as a training corpus, together with two new heuristics to improve an n-gram based language identification algorithm for Asian languages. Extension of the training corpus produced improved accuracy. Improvement was also achieved by using byte-sequence based HTML parser and a HTML character entities converter. The performance of the algorithm was evaluated based on a written text corpus of 1,660 web pages, spanning 182 languages from Asia, Africa, the Americas, Europe and Oceania. Experimental result showed that the algorithm achieved a language identification accuracy rate of 94.04%.

**Keywords:** Asian Language, Byte-Sequences, HTML Character Entities, N-gram, Non-Latin-Script, Language Identification.

## 1. Introduction

With the explosion of multi-lingual data on the Internet, the need and demand for an effective automated language identifier for web pages is further increased. Wikipedia, a rapidly growing multilingual Web-based encyclopedia on the Internet, can serve as a measure of the multilingualism of the Internet. We can see that the number of web pages and languages (both Latin-script and non-Latin-script based) has increased tremendously in recent years, as shown in Figure 1.

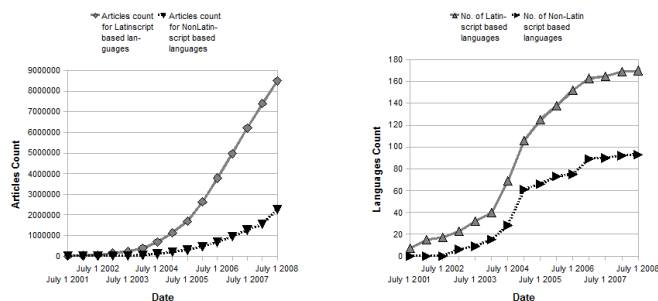


Figure 1 Articles count and number of languages (Latin-script and non-Latin-script based) on Wikipedia's language projects, 2001 to 2008.

### 1.1 Unreliable HTML and XML's Language Attribute

The Hyper Text Markup Language (HTML) is the standard encoding scheme used to create and format a web page. In the latest HTML 4.01 specification, there is a *lang* attribute that defined to specify the base language of text in a web page. Similarly, the Extensible Markup Language (XML) 1.0 specification includes a special attribute named *xml:lang* that may be inserted into documents to specify the language used in the contents. However, the reality remains that many web pages do not make use of this attribute or, even worse, use it incorrectly and provide misleading information.

Using the validation corpus in this study as a sample, we found that only 698 web pages out of 1,660 contain *lang* attribute, as shown in Table 1. When *lang* attribute is available, it does not always indicate the correct language of a web page. Table 1 shows that 72.49% of web pages with *lang* attribute produced correct language indication. Overall, only 30.48% of web pages in our sample

produced correct language identification result from *lang* attribute. Therefore, we are left with deducing information from the text to determine the language of a given web page. This is the domain of language identification.

Table 1 Number of web pages with *lang* attribute and percentage of correct language identification using *lang* attribute as indicator, based on validation corpus of this study.

	Correct Pages	Total Pages	Percent Correct
Web pages with <i>lang</i> attribute	506	698	72.49%
Web pages without <i>lang</i> attribute	0	962	0.00%
Total	506	1660	30.48%

## 1.2 Language Identification

Language identification is the fundamental requirement prior to any language based processing. For example, in a fully automatic machine translation system, language identification is needed to detect the source language correctly before the source text can be translated to another language. Many studies of language identification on written text exists, for example, [Gold 1967] [William B. Cavnar 1994] [Dunning 1994] [Clive Souter 1994] [Michael John Martino 2001] [Izumi Suzuki 2002] [ÖLVECKÝ 2005] [Bruno Martins 2005], just to name a few.

A comparative study on language identification methods for written text was reported in [Lena Grothe 2008]. Their paper compares three different approaches to generate language models and five different methods for language classification.

The first approach generates language model based on "short words". It uses only words up to a specific length to construct the language model. The idea behind this approach is that language specific common words having mostly only marginal length. [Grefenstette 1995] tokenized and extracted all words with a length up to five characters that occurred at least three times from one million characters of text for ten European languages. [Prager 1999] used still shorter words four or fewer characters, for thirteen Western European languages.

The second approach generates language model is based on "frequent words". It uses a specified number of the most frequent words occurring in a text to construct the language model. For instance, the most frequent one hundred words were used in [Clive Souter 1994] and [Michael John Martino 2001], while [Eugene Ludovik 1999] used the most frequent one thousand words.

The third approach generates a language model based on "n-gram". An n-gram is a subsequence of N items from a given sequence. [William B. Cavnar 1994] [Grefenstette 1995] [Prager 1999] used a character-sequence based n-gram method, while [Dunning 1994] used a byte-sequence based n-gram method.

The generated language model is used as the input for language classification method. Many language classification methods had been proposed before, these include Ad-Hoc Ranking [William B. Cavnar 1994], Markov Chains in combination with Bayesian Decision Rules [Dunning 1994], Relative Entropy [Penelope Sibun 1996], Vector Space Model [Prager 1999] and Monte Carlo sampling [Poutsma 2001].

Table 2 shows the information of five selected studies. Previous studies reported excellent results on a few selected Latin-script based languages. Japanese and Russian are the only two exceptional here. The Japanese language, written with the Japanese logographs and syllabaries, and Russian, written in the Cyrillic script, can be easily distinguished from the Latin-script based languages, and also from each other. However, the performance of language identification on non-Latin-script based languages remains unknown.

Most studies in Table 2 are focusing on plain text content. There is only two previous study evaluate its language identification algorithm against web page. Although the proposed heuristics work well on Latin-script based web page, they might not able to effectively handling the non-Latin-script based web page. Usually, non-Latin-script has different bits setting, while many non-Latin-scripts in Asia are encoded in legacy fonts. Besides, none of the studies mentioned about HTML entities, which indeed is commonly used in non-Latin-script based web page.

As previous studies are focusing on Latin-script based languages, most of them adopted a training corpus with limited number of Latin-script based languages only. Thus, our research aims to improve language identification on a broader range of languages, especially for non-Latin-script and added support for web page content. The initial target is set at the 185 languages given in ISO 639-1.

## 1.3 Hyper Text Markup Language and HTML Parser

[Penelope Sibun 1996] states that language identification is a straightforward task. We argue that their claim is only true for language identification on Latin-script based plain text document. Web pages are different from plain text documents since they contain the HTML tags that are used to publish the document on the Web. In order to correctly identify the language of a web page, a HTML parser is

Table 2 Five selected language identification studies on written text with information of languages coverage, training corpus, validation corpus and accuracy of identification.

Research	Language Coverage	Training Corpus	Validation Corpus	Percent Correct
[William B. Cavnar 1994]	English, Portuguese, French, German, Italian, Spanish, Dutch, Polish	Unspecified	3713 text sample from soc.culture newsgroup	99.8%
[Dunning 1994]	Dutch, Polish	A set of text samples from Consortium for Lexical Research	Another set of text samples from Consortium for Lexical Research	99.9%
[Clive Souter 1994]	Dutch/Friesian, English, French, Gaelic, German, Italian, Portuguese, Serbo-Croat, Spanish	A set of text samples from Oxford Text Archive, each is 100 kilobytes	Another set of text samples from Oxford Text Archive	94.0%
[Poutsma 2001]	Danish, Dutch, English, French, German, Italian, Norwegian, Portuguese, Spanish, Swedish	90% of text samples from European Corpus Initiative Multilingual Corpus	10% of text samples from European Corpus Initiative Multilingual Corpus	Result in chart format
[Bruno Martins 2005]	Danish, Dutch, English, Finnish, French, German, Italian, Japanese, Portuguese, Russian, Spanish, Swedish	Text samples of 23 languages collected from newsgroups and the Web	Web pages of 12 languages collected from newsgroups and the Web	91.25%

needed in order to remove the HTML tags and to extract the text content for language identification.

An HTML parser usually processes text based on character sequences. The HTML parser read the content of a web page into character sequences, and then marked the blocks of HTML tags and the blocks of text content. At this stage, the HTML parser uses a character encoding scheme to encode the text. HTML parser usually depends on a few methods (describes in subsection Character and Byte-sequence based HTML Parser) to determine the correct character encoding scheme to be used. If no valid character encoding is detected, the parser will apply a predefined default encoding.

Today, a common approach is to use UTF-8 (a variable-length character encoding for Unicode) as the default encoding, as the first 128 characters of Unicode map directly to their ASCII correspondents. However, using UTF-8 encoding on non-Latin-script based web pages might cause the application to apply a wrong character encoding scheme and thus return an encoded text that is different from its web origin.

Using the validation corpus of this study as an example, we found that 191 web pages were with doubtful character

encoding information. Table 3 shows an example of text rendered by wrongly character encoding. The authors only show one example as the reason for wrong character encoding is identical.

#### 1.4 Unicode and HTML Character Entities

Unicode is a computing industry standard that allowing computers to represent and manipulate text expressed in most of the world's writing systems. The Unicode Consortium has the ambitious goal of eventually replacing existing character encoding schemes with Unicode, as many of the existing schemes are limited in size and scope. Unicode characters can be directly input into a web page if the user's system supports them. If not, HTML character entities provide an alternate way of entering Unicode characters into a web page.

There are two types of HTML character entities. The first type is called character entity references, which take the form *&EntityName;*. An example is *&copy;* for the copyright symbol. The second type is referred as numeric character references, which takes the form *&#N;*, where *N* is either a decimal number (base 10) or a hexadecimal number for the Unicode code point. When *N* represents a hexadecimal number, it must be prefixed by *x*. An

Table 3 Text rendered and language identification results on a selected web page with misleading *charset* information.

Web Page	HTML Parser (Character-sequence based)			Web Origin	
	Detected Charset	Text Rendered	Identified As	Text Rendered	Identified As
chinese-05-news.htm	No Match, use default UTF-8	????	English, Latin, Latin1	杭州旅遊	Chinese, Simplified Chinese, GB2312

examples of these entities is  $\&\#21644;$  (base 10) or  $\&\#x548c;$  (base 16) for the Chinese and also Japanese character "和".

Using HTML character entities, any system is able to input Unicode characters into a web page. However, this causes a problem for language identification as the language property is now represented by label and numeric references. In order to identify the language of an HTML character-entity-encoded web page, we propose a HTML character entity converter to translate such entities to the byte sequences of its corresponding Unicode code point.

## 1.5 Organization of this paper

The remaining of this paper is ordered in the following structure. The authors review related works in the next section. In Methodology section, the authors describe the language identification process and the new heuristics. In Data and Experiments section, the authors explain the nature and preparation of training and validation corpus; followed by description on how the experiments are setup and the purposes of them. In the Result and Discussion section, the authors present the results from the experiments. In the last section, the authors draw conclusions and propose a few areas for future work.

## 2. Related Work

### 2.1 Martin Algorithm

In [Bruno Martins 2005], the authors discussed the problem of automatically identifying the language of a given web page. They claimed that web page is generally contained more spelling errors, multilingual and short text, therefore, it is harder for language identification on the web pages. They adapted the well-known n-gram based algorithm from [William B. Cavnar 1994], complemented it with a more efficient similarity measure [Lin 1998] and heuristics to better handle the web pages. The heuristics included the following six steps:

- i. Extract the text, the markup information, and meta-data.
- ii. Use meta-data information, if available and valid.
- iii. Filter common or automatically generated strings. For example, "This page uses frames".
- iv. Weight n-grams according to HTML markup. For example, n-grams in the title section have more weight than n-grams in meta-data section.
- v. Handle situations when there is insufficient data. When a web page has less than 40 characters, the system reports "unknown language".

- vi. Handle multilingualism and the "hard to decide" cases. When a document cannot be clearly classified to one language, the system will re-apply the algorithm, and weight the largest text block as three times more important than the rest.

In the experiment, they constructed 23 different language models from textual information extracted from newsgroups and the Web. They tested the algorithm using testing data in 12 different languages, namely Danish, Dutch, English, Finnish, French, German, Italian, Japanese, Portuguese, Russian, Spanish and Swedish, respectively. The total number of documents for testing is 6,000, with 500 documents for each language. The testing data were crawled from on-line newspapers and Web portals. Overall, the best identification result returned accuracy of 91.25%, which was lower than other researches on text document. The authors believe that this is due to the much noisier nature of the text in web page.

### 2.2 Suzuki Algorithm

In [Izumi Suzuki 2002], the method is different from conventional n gram based methods in the way that its threshold for any categories is uniquely predetermined. For every identification task on target text, the method must be able to respond to either "correct answer" or "unable to detect". The authors used two predetermined values to decide which answer should respond to a language identification task. The two predetermined values are UB (closer to the value 1) and LB (not close to the value 1), with a standard value of 0.95 and 0.92, respectively. The basic unit used in this algorithm is trigram. However, the authors refer to it as a 3-byte shift-codon.

In order to detect the correct language of a target text, the algorithm will generate a list of shift-codons from the target text. The target's shift-codons will then compare to the list of shift-codons in training texts. If one of the matching rates is greater than UB, while the rest is less than LB, the algorithm will report that a "correct answer" has been found. The language of the training text with matching rate greater than UB is assumed to be language of the target text. By this method, the algorithm correctly identified all test data of English, German, Portuguese and Romanian languages. However, it failed to correctly identify the Spanish test data.

## 3. Methodology



The general paradigm of language identification can be divided into two stages. First, a set of language model is generated from a training corpus during the training phase. Second, the system constructs a language model from the target document and compares it to all trained language models, in order to identify the language of the target document during the identification phase. The algorithm used in this study adopted this general paradigm; however, it contains two new heuristics to properly handle web pages. The first heuristic is to remove HTML tags in byte-sequence stream. The second heuristics is to translate HTML character entities to byte sequences of their Unicode code point. The algorithm only takes text and HTML documents as valid input. The overall system flow of language identification process is shown in Figure 2.

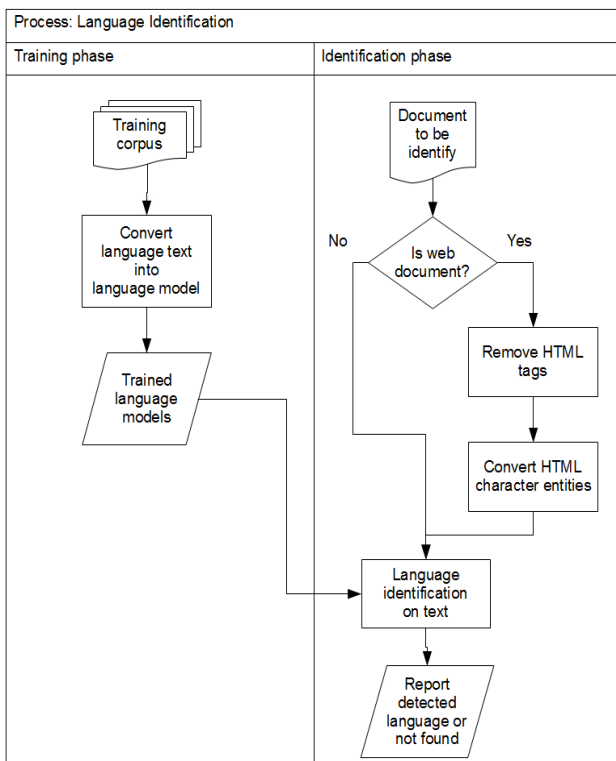


Figure 2 System flowchart for language identification process in this paper.

### 3.1 Byte-sequence based n-gram algorithm

An n-gram is a sub sequence of  $N$  items from a longer sequence. An n-gram order 1 (i.e.  $N=1$ ) is referred to as a monogram; n-gram order 2 as a bi-gram and n-gram order 3 as a trigram. Any other is generally referred to as "N-gram". This paper adapted the n-gram based algorithm proposed by [Izumi Suzuki 2002]. The algorithm generates

language model from text document into trigrams of byte sequences. For example, the trigrams for the Japanese word "こんにちわ" (or 82 B1 82 F1 82 C9 82 BF 82 CD in the Shift-JIS character encoding scheme) are highlighted as follows:

**82 B1 82 F1** 82 C9 82 BF 82 CD  
 82 **B1 82 F1** 82 C9 82 BF 82 CD  
 82 B1 **82 F1 82 C9** 82 BF 82 CD  
 82 B1 82 **F1 82 C9** 82 BF 82 CD  
 82 B1 82 F1 **82 C9 82 BF** 82 CD  
 82 B1 82 F1 82 **C9 82 BF** 82 CD  
 82 B1 82 F1 82 C9 **82 BF 82 CD**  
 82 B1 82 F1 82 C9 82 **BF 82 CD**

The language classification method is based on trigram frequency. The trigram distribution vector of training document has no frequency information. Only the target document has a frequency-weighted vector. In order to detect the correct language of a target document, the algorithm will generate a list of byte-sequence based trigrams from the target document, together with the frequency information of each trigram. The target document's trigrams will then be compared to the list of byte-sequence based trigrams in every training language model. If a target's trigram matches a trigram in the training language model, its frequency value is added to the matching counter. After all trigrams from target document have been compared to trigrams in training language model, the matching rate is calculated by dividing the final matching counter by the total number of target's trigrams.

The matching process for detecting a language can be summarizing as below:

- i. Let  $N$  be the number of trigrams in target document.
- ii. All the trigrams from the target document  $u_1, u_2, \dots, u_N$  are listed. Let  $u_j$  be the  $j^{\text{th}}$  trigram in the target language model.
- iii. Let  $T_i$  be the  $i^{\text{th}}$  language model in the training corpus.  $R_i$  (or  $R$ -values) is calculated from every  $i^{\text{th}}$  language model using equation (1), where  $R_i$  is the rate at which the set of trigrams in  $i^{\text{th}}$  language model of the training corpus appears in the target document.

$$R_i = \sum_{j=1}^n \frac{f(u_j)}{n}, \quad \text{where } f(u_j) = \begin{cases} 1 & \text{if } u_j \in T_i \\ 0 & \text{Otherwise} \end{cases} \quad (1)$$

### 3.2 Character and Byte-sequence based HTML Parser

In order to correctly process a web page, a HTML parser must ascertain what character encoding scheme is used to encode the content. This section describes how to detect

Table 4 Possible scenarios of character encoding scheme determination.

Encoding in HTTP content-type	Override HTTP server-side encoding	Encoding in XML declaration	Encoding in HTML meta charset element	Default encoding by User's application	Result of character encoding detection
Correct	No	Any	Any	Any	Correct
Wrong	No	Any	Any	Any	Wrong
Any	Yes	Correct	Any	Any	Correct
Any	Yes	Wrong	Any	Any	Wrong
Any	Yes	Missing	Correct	Any	Correct
Any	Yes	Missing	Wrong	Any	Wrong
Any	Yes	Missing	Missing	Correct	Correct

the character encoding in Hypertext Transfer Protocol (HTTP) header, XML or HTML.

When a web page is transmitted via the HTTP, the Web server will sent the character encoding in the content-type field of the HTTP header, such as *content-type :text/html; charset=UTF-8*. The character encoding can also be declared within the web page itself. For XML, the declaration is at the beginning of the markup, for instance, *<?xml version="1.0" encoding="utf-8"?>* for HTML, the declaration is within the *<meta>* element, such as *<meta http-equiv="content-type" content="text/html; charset=UTF-8">*. If there is no valid character encoding information detected, a predefined character encoding scheme will be invoked. The default character encoding scheme varies depending on the localization of the application. In the case of conflict between multiple encoding declarations, precedence rules apply to determine which declaration shall be used. The precedence is as follows, with HTTP content-type being the highest priority:

- i. HTTP content-type
- ii. XML declaration
- iii. HTML Meta charset element

Since information in the HTTP header overrides information in the web page, it is therefore important to ensure that the character encoding sent by the Web server is correct. However, in order to serve file or files using a different encoding than that specified in the Web server's default encoding, most Web serves allow the user to override the default encoding defined in HTTP content-type. Table 4 illustrates all possible scenarios of character encoding scheme determination.

Table 4 shows that misleading and missing character encoding information would probably lead to the wrong result. Therefore, it is quite possible that a character-sequence based HTML parser might apply an incorrect character encoding scheme to web pages without valid

character encoding information, especially on non-Latin-script web pages.

The HTML parser implemented in this paper is unique in that it processes the content of a web page based on byte sequences, thus avoiding the above mentioned problem. By using byte sequences, it eliminates the need to detect and apply character encoding scheme on the content extracted from the web page. The HTML parser parses the web page in a linear fashion. It searches for HTML tags from the beginning to the end of page. It looks for valid HTML start and end tags and marks all blocks of HTML tags. The parser removes all detected HTML blocks and return remaining content in byte sequences for language identification. The parser searches in sequence of bytes instead of characters. For example, in order to determine the locations of *<body>* and *</body>* tags in a web page, the parser searches for 3C 62 6F 64 79 3E and 3C 2F 62 6F 64 79 3E, respectively. The parser keeps a list of byte-sequence based HTML tags and uses them to remove HTML tag's blocks from the target web page.

### 3.3 HTML Character Entity Converter

The HTML character entity converter is designed to translate HTML entities to corresponding byte sequences of Unicode's code point. The converter is able to handle both character entity references and numeric character references. There are 252 character entity references defined in HTML version 4, which act as mnemonic aliases for certain characters. Our converter maintains a mapping table between the 252 character entity references and their represented byte sequences in hexadecimal number. When a character entity reference is detected by the converter, it replaces the entity with its associated byte sequences.

For numeric character references, the converter performs a real time decoding process on it. The converter will convert the character reference from decimal (base 10)



number to byte sequences if it detects the following pattern: character ampersand (&), followed by character number sign (#), followed by one or more decimal digits (zero through nine), and lastly followed by character semicolon (;). For example, `&#65;` (representing the Latin capital letter A).

Similarly, the converter will convert the character reference from hexadecimal (base 16) number to byte sequences if it detects the following pattern: character ampersand (&), followed by character number sign (#), followed by character (x), followed by one or more hexadecimal digits (which are zero through nine, Latin capital letter A through F, and Latin small letter a through f), and lastly followed by character semicolon (;). For example, `&#x41;` (again representing the Latin capital letter A).

Table 5 shows the byte sequences output by the HTML character entities converter, using an ampersand sign (&), a Greek small letter beta ( $\beta$ ) and a Chinese character "平" as examples. These examples are carefully selected to show the different ways of conversion based on different number of byte order in UTF-8.

#### 4. Data and Experiments

There are two sets of data used in this study. The first set is the training corpus, which contains training data used to train the language models. The second set is the validation corpus, which is a collection of web pages used as target documents in the experiments.

##### 4.1 Training Corpus

In this paper, the authors prepared two sets of training data. The first set of training data is constructed from 565 Universal Declaration of Human Rights (UDHR) texts collected from the Office of the High Commissioner for Human Rights (OHCHR) web site and Language

Observatory Project (LOP). UDHR was selected as it is the most translated document in the world, according to the Guinness Book of Records.

The OHCHR web site contained 394 translations in various languages. However, 80 of them are in Portable Document Format (PDF). As a result, only 314 languages were collected from OHCHR. The LOP contributed 18 new languages. The total size of the first set of training data is 15,241,782 bytes. Individual file size ranged from 4,012 to 55,059 bytes. From here onward this set of training data will be referred to as training corpus A.

The second set of training data, training corpus B, increases the number of languages by 33. It contains 65 (some are same language but in different encoding schemes) Biblical texts collected from the United Bible Societies (UBS). All files have similar content, but written in different languages, scripts and encodings. The total size of the second set of training data is 1,232,322 bytes. Individual file size ranged from 613 to 54,896 bytes.

Most languages have more than one training file in the training corpora. This is because the same language can be written in different scripts and encodings. For example, the Chinese language has five training files in training corpus A. The five training files by language\_script\_encoding are: Chinese\_Simplified\_EUC-CN, Chinese\_Simplified\_HZ, Chinese\_Simplified\_UTF8, Chinese\_Traditional\_BIG5 and Chinese\_Traditional\_UTF8. Likewise, a language might be covered by texts in training corpus A and B.

Table 6 shows the number of languages, scripts, encodings and user-defined fonts of the training corpora, sorted according to geographical regions. The column header (A  $\cup$  B) represents the distinct number of languages, scripts, encodings and fonts in the corpora.

From Table 6, we can observe that the Asian region is more diversity in its written languages. Asia has the highest number of scripts (writing systems), character

Table 5 Example to show output of HTML character entities converter, based on three different types of HTML entities and each using different byte order.

Character	Character Entity References	Numeric Character References	Unicode Code Point	UTF-8 Byte Order			Output in Byte Sequences
				Byte-1	Byte-2	Byte-3	
&	<code>&amp;amp;</code>	<code>&amp;#38;</code>	U+0026	0xxxxxxx			U+0026 -> 00100110 -> 0x26
$\beta$	<code>&amp;beta;</code>	<code>&amp;#946;</code>	U+03B2	110yyyxx	10xxxxxx		U+03B2 -> 1100111010110010 -> 0xCEB2
平		<code>&amp;#x5e73;</code>	U+5E73	1110yyyy	10yyyyxx	10xxxxxx	U+5E73 -> 111001011011100110110011 -> 0xE5B9B3

Table 6 Number of languages, scripts, encodings and user-defined font's information in training corpus A and B, sorted according to geographical region.

Training Corpus	Language			Script			Encoding			Font		
	A	B	A∪B	A	B	A∪B	A	B	A∪B	A	B	A∪B
Africa	90	10	97	4	2	4	2	1	2	1	0	1
Asia	79	27	92	28	17	32	13	4	14	17	6	23
Caribbean	5	1	6	4	1	4	2	1	2	3	0	3
Central America	7	0	7	1	0	1	2	0	2	0	0	0
Europe	64	16	72	4	3	5	6	3	6	1	0	1
Int. Aux. Language(IAL)	3	1	4	1	1	1	3	1	3	0	0	0
Middle East	1	0	1	1	0	1	2	0	2	0	0	0
North America	20	1	21	2	1	2	2	1	2	1	0	1
Pacific Ocean	16	3	18	1	1	1	2	1	2	0	0	0
South America	47	0	47	1	0	1	2	0	2	0	0	3
Unique count			365			40			19			29

encoding schemes and user-defined fonts. Each of these factors makes language identification difficult. In the case of user-defined fonts, many of them do not comply with international standards, hence making language identification an even more challenging task.

#### 4.2 Validation Corpus

The validation corpus is comprised of texts from web pages. The authors predefined three primary sources to search for web pages in different languages. These sources are Wikipedia, the iLoveLanguages gateway and online news/media portals. The source referred here is not necessarily a single web site. For example, a web portal might contain, or link to, many web sites. Table 7 shows more detailed information on each source.

The rule for selection is to collect one web page per web site. The authors believe that in general a web site will apply the same character encoding scheme to the web

pages it hosts. Thus, it would be redundant to collect more than one page from the same web site. For each language, we collected a maximum of 20 web pages. Popular languages like Arabic (ar), Chinese (zh), and English (en) are easy to find, while less popular languages, like Fula (ff), Limburgish (li), or Sanskrit (sa) are very difficult to find.

The authors' initial target was to cover all of the 185 languages listed in ISO 369-1. However, three languages, namely Kanuri (kr), Luba-Katanga (lu) and South Ndebele (nr) could not be found from the sources, nor by using search engines on the Web. As a result, the final validation corpus used in the experiments contained 182 languages. There are 1,660 web pages in the validation corpus, occupying 76,149,358 bytes of storage. The authors did not normalize the size of collected web pages as the wide variation reflects the real situation on the Web.

Each web page in the validation corpus has its filename in

Table 7 Information of defined Web's sources for collecting web pages for the validation corpus.

Web Site	Validation corpus			
	No. of Pages	Total Size (bytes)	Min. (bytes)	Max. (bytes)
Wikipedia	171	7,511,972	601	146,131
iLoveLanguages	103	790,934	3,634	18,445
BBC	34	396,292	2,990	61,190
China Radio	13	1,891,896	9,419	222,526
Deutsche Welle	14	1,164,620	5,957	87,907
The Voice of Russia	26	1,832,797	39,198	103,251
Voice of America	22	1,791,145	9,674	87,574
Kidon Media-Link & ABYZ News Links	1,277	60,769,702	135	1,048,314
Total	1,660	76,149,358		

Table 8 Experiments' settings and language identification results.

Experiment	One	Two	Three	Four
Training corpus	A	A and B	A and B	A and B
HTML parser	Character-sequence	Character-sequence	Byte-sequence	Byte-sequence
HTML character entities converter	Disabled	Disabled	Disabled	Enabled
Correct/Total	1,241/1,660	1,444/1,660	1,494/1,660	1,561/1,660
Accuray rate	74.76%	86.99%	90.00%	94.04%

the following format: language-index-source. Language indicates the language of the web page; index represents the accumulated number of texts in each language; and source indicates the original web site of the page.

Unlike many researches listed in Table 2, our validation corpus is totally independent from the training corpus. By selecting validation sample files from different sources, the validation corpus evenly represents the language diversity on the web, while increasing its coverage on language, script, and encoding systems on the web, as wide as possible.

### 4.3 Experiments

Four experiments were performed. Each experiment was designed to show the baseline performance and the improvement achieved by using training data, the byte-sequence based HTML parser, and the HTML character entity converter. Table 8 provides a summary of the conditions and results of each experiment.

In the first experiment, we trained the language models using training corpus A. The HTML parser adapted in this experiment was based on character sequences, which relies on the mechanism, described in Section Character and Byte-sequence based HTML Parser and integrates the Mozilla Charset Detector algorithm to determine the character encoding scheme of a web page. If no valid character encoding scheme is detected, the parser uses its predefined default encoding, i.e., UTF-8 to encode extracted text. The HTML character entity converter was not used in this experiment.

The second experiment was designed to evaluate the improvement achieved through the extension of the training corpus. Training corpus A and B were used to train the language models. The remaining settings are the same as in the first experiment.

The third experiment applied the same settings as in the second experiment; except that the character-sequence based HTML parser was replaced by a byte-sequence based HTML parser. Besides evaluating the efficacy of the

byte-sequence based HTML parser, the authors also analyzed the number of web pages with missing or invalid character encoding scheme information.

The final experiment is designed to evaluate the efficacy of the HTML character entity converter. The converter is enabled while keeping the remaining settings the same as in the previous experiment. In addition, we also examined the number of web pages that are encoded with HTML character entities in the validation corpus.

## 5. Results and Discussion

The summarized evaluation results of all experiments are presented in Table 8, where different settings are used. The first column showed the result of Experiment one, while the following right columns shows the results of Experiments two, three and four, respectively.

Experiment one was used as a base line for comparison. It adapted [Izumi Suzuki 2002] algorithm for language identification, but the correct language identification rate was only 74.76%. After manually inspecting the results and web pages, the authors found that 217 out of the 419 (i.e., 1660-1241) wrongly identified web pages were due to the unavailability of corresponding language models. This evidence that training corpus A alone was inadequate led to the decision to expand the training corpus. As a result, the authors collected new training data from the United Bible Societies web site in order to increase the number of language models.

### 5.1 Evaluation on Effectiveness of Training Corpus B

After adding the new language models of training corpus B and repeating the identification process as Experiment two, the algorithm was able to increase its accuracy of language identification from 74.76% to 86.99%, i.e., a 12.23 percent point improvement from the previous test. All of the 217 web pages wrongly identified due to unavailability of corresponding language models in Experiment one were correctly identified. However, there

Table 9 List of files in Validation Corpus that are correctly identified in Experiment one, but wrongly identified in Experiment two.

File in VC	Language Identification					
	Experiment one			Experiment two		
	Language	Script	Encoding	Language	Script	Encoding
bosnian-15-svevijesti.ba	Bosnian	Latin	Latin2	Punjabi	Gurmukhi	UTF8
indonesian-11-watchtower	Indonesian	Latin	UTF8	Aceh	Latin	Latin1
indonesian-19-pontianakpost	Indonesian	Latin	UTF8	Malay	Latin	Latin1
interlingua-01-wikipedia	Interlingua	Latin	UTF8	Spanish	Latin	UTF8
italian-13-rai.it	Italian	Latin	UTF8	Aragonese	Latin	UTF8
ndonga-01-wikipedia	Nepali	Devanagari	UTF8	Kwanyama	Latin	UTF8
persian_dari-08-afghanpaper	Persian-Dari	Arabic	UTF8	Pashto	Arabic	UTF8
portuguese-11-acorianooriental	Portuguese	Latin	Latin1	Galician	Latin	UTF8
portuguese-13-diariodoalentejo	Portuguese	Latin	Latin1	Galician	Latin	UTF8
portuguese-18-falcaodominho.pt	Portuguese	Latin	Latin1	Galician	Latin	UTF8
serbian-10-watchtower_cyrillic	Serbian	Cyrillic	UTF8	Macedonian	Cyrillic	UTF8
tibetan-03-tibettimes.net	Tibetan	Tibetan	UTF8	Dzongkha	Tibetan	UTF8
zulu-01-wikipedia	Zulu	Latin	Latin1	Ndebele	Latin	UTF8
zulu-09-zulutop.africanvoices	Zulu	Latin	Latin1	Ndebele	Latin	UTF8

were 14 web pages that had been correctly identified before, but were wrongly identified in Experiment two due to the problem of over-training. Over-training problem occurs when a language model is over-trained by a larger training data size; and/or a newly trained language model affects the accuracy of other language models. Table 9 shows the list of files that affected by this problem.

## 5.2 Evaluation on character and byte-sequence based HTML Parser

During the HTML parsing stage of Experiment two, the language identification process detected 1,466 web pages with valid charset information and 191 web pages with doubtful charset information. Of these 191 web pages, fourteen had "user-defined" charset and 177 were missing charset information.

The character-sequence based HTML parser used in Experiment one and two was defined to use UTF-8 encoding to encode web page without valid charset information. When investigated on web pages without valid charset information, it was found that the default UTF-8 character encoding scheme worked well on Latin-script based languages, but did not work well for 11 non-Latin-script based languages: Amharic, Arabic, Armenian, Belarusian, Bulgarian, Chinese, Greek, Hebrew, Macedonian, Russian and Ukrainian, respectively. Fifty wrong classifications occurred after applied UTF-8 to the text extracted from web pages belonging to those languages. Of those 50 pages, Africa(7), Asia(30),

Europe(10), International Auxiliary Language(2) and Middle East(1). As a result, the byte-sequence based HTML parser was introduced in Experiment three.

By eliminating the steps of guessing and applying charset to text using charset returned by the charset detector, the byte-sequence based parser was able to improve the accuracy of language identification in Experiment three to 90.00%. All of the previously mentioned 50 web pages were identified correctly in Experiment three.

## 5.3 Evaluation on HTML Character Entities Converter

Experiment three miss-classified 166 web pages. Among those, 76 web pages are caused by HTML character entities problem. As a result, the HTML character entities converter was introduced in Experiment four.

The accuracy of language identification in Experiment four is 94.04%. The HTML character entities converter improved the algorithm by correctly identified 67 out of the 76 (88.16%) HTML entities encoded web pages. There were 9 HTML entities encoded web pages not correctly identified, where 3 of them were due to untrained legacy font and the remaining 6 were miss identified to another closely related language, like Amharic identified as Tigrinya, Assamese identified as Bengali, Persian identified as Pashto, etc.

Table 10 shows the language identification result based on writing systems. For Non-Latin-Script based languages, the algorithm achieved perfect score on Logographic and Syllabic systems based languages; its accuracy on Abjad (93.33%), and Non-Latin Alphabet (94.17%) based languages is acceptable. The worst performance come under Abugida system based languages due to many of their web pages encoded with legacy fonts. In case of Latin-Script based languages,. The algorithm achieved 95.42% accuracy rate.

Table 10 Language identification result based on writing systems.

Writing System	Language Identification Result	Percent Correct
Abjad (e.g. Arabic)	126/135	93.33%
Abugida (e.g. Indic scripts)	211/242	87.19%
Alphabet (Latin)	938/983	95.42%
Alphabet (Non-Latin)	226/240	94.17%
Logographic (Chinese)	40/40	100.00%
Mixed logographic and syllabic (Japanese)	20/20	100.00%

## 6. Conclusion and Future Work

The primary aim of this paper was to take into account the practical issues of language identification on non-Latin-script based web pages, especially for Asia and Africa regions; and to propose corresponding methods to overcome the issues. In this paper we have shown that the adaption of UDHR and Biblical texts as training data are simple and yet effective ways of gathering data on a large variety of languages. An initial language identification accuracy rate of 86.99% was obtained based on testing 1,660 web pages in 182 different languages. We proposed and discussed the importance of a byte-sequence based HTML parser and a HTML character entity converter for non-Latin-script web pages. The evaluation results showed that our algorithm with the two new heuristics was able to improve the accuracy of language identification from 86.99% to 94.04%.

The list of future work includes finding the optimal length for training data in order to avoid the over-training problem; improvement of language identification for closely related languages; extending the algorithm to handle multi-lingual web pages; and lastly, finding a method to effectively handle the user-defined font issue.

## Acknowledgments

The authors are grateful for the sponsorship of the Japan Science Technology Agency (JST) through the Language Observatory Project (LOP) and Country Domain Governance (CDG) Project, which in turn provided the necessary resources to improve the language identification algorithm.

## References

- [1] Bruno Martins, Mário J. Silva. "Language identification in web pages." 2005 ACM symposium on Applied computing. Santa Fe: ACM New York, NY, USA, 2005. 764-768.
- [2] Clive Souter, Gavin Churcher, Judith Hayes, John Hughes, Stephen Johnson. "Natural Language Identification using Corpus-Based Models." *Hermes - Journal of Language and Communication Studies*, 1994: 183-204.
- [3] Datong Chen, Hervé Bourlard, Jean-Philippe Thiran. "Text Identification in Complex Background Using SVM." *IEEE Conference on Computer Vision and Pattern Recognition*. 2001. 621-626.
- [4] Dunning, Ted. *Statistical Identification of Language*. Technical Report, Computing Research Laboratory, New Mexico State University, 1994.
- [5] Eugene Ludovik, Ron Zacharski, Jim Cowie. "Language Recognition for Mono- and Multi-lingual Documents." In the *Proceeding of the VEXTAL Conference*. Venice, 1999.
- [6] Gold, E Mark. "Language identification in the limit." *Information and Control* 10, no. 5 (1967): 447-474.
- [7] Gordon, Raymond G. *Ethnologue: Languages of the World*. 15th. Dallas: SIL International, 2005.
- [8] Grefenstette, Greg. "Comparing Two Language Identification Schemes." *The proceedings of 3rd International Conference on Statistical Analysis of Textual Data (JADT 95)*. Rome, 1995.
- [9] Izumi Suzuki, Yoshiaki Mikami, Ario Ohsato, Yoshihide Chubachi. "A language and character set determination method based on N-gram statistics." *ACM Transactions on Asian Language Information Processing (TALIP)*, 2002: 269-278.
- [10] Lena Grothe, Ernesto William De Luca, Andreas Nürnberger. "A Comparative Study on Language Identification Methods." *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*. Marrakech, 2008. 980-985.
- [11] Lin, Dekang. "An Information-Theoretic Definition of Similarity." *Proceedings of the Fifteenth International Conference on Machine Learning*. 1998. 296-304.
- [12] Matthias Richter, Uwe Quasthoff, Erla Hallsteinsdóttir, Christian Biemann. "Exploiting the Leipzig Corpora Collection." *IS-LTC 2006*. Ljubljana, 2006.
- [13] Michael John Martino, Robert Charles Paulsen Jr. *Natural language determination using partial words*. United States Patent US 6,216,102 B1. April 10, 2001.
- [14] ÖLVECKÝ, Tomáš. "N-Gram Based Statistics Aimed at Language Identification." *IIT.SRC*. Bratislava, 2005. 1-17.
- [15] Penelope Sibun, A. Lawrence Spitz. "Language Identification: Examining the Issues." In *Proceedings of the 5th Symposium on Document Analysis and Information Retrieval*. Las Vegas, 1996. 125-135.

- [16] Peter F. Brown, Vincent J. Della Pietra, Peter V. deSouza, Jenifer C. Lai, Robert L. Mercer. "Class-Based n-gram Models of Natural Language." *Computational Linguistics* 18 (1992): 467-479.
- [17] Poutsma, Arjen. "Applying Monte Carlo Techniques to Language Identification." *Computational Linguistics in the Netherlands*. Rodopi, 2001. 179-189.
- [18] Prager, John M. "Linguini: Language Identification for Multilingual Documents." *Proceedings of the 32nd Hawaii International Conference on System Sciences*. IEEE Computer Society, 1999. 2035.
- [19] William B. Cavnar, John M. Trenkle. "N-Gram-Based Text Categorization." *SDAIR-94, 3rd Annual Symposium on Document Analysis and Information Retrieval*. 1994. 161-175.

**First Author** Yew Choong Chew received his Master's Degree of Management and Information Systems Engineering from Nagaoka University of Technology in 2007. He is currently a Ph.D candidate and researching in statistical natural language processing for written text.

**Second Author** Yoshiki Mikami received his B.Eng (Mathematical Engineering) from Tokyo University in 1975. Joined ministry of International Trade and Industry (MITI) in 1975. Worked in JETRO Singapore Office 1994-1997. Since July 1997, he is a Professor at Nagaoka University of Technology. His research interests include natural language processing and Internet governance.

**Third Author** Robin L. Nagano is currently an English language teacher at the University of Miskole, Hungary. She holds Master's degrees in Applied Linguistics (Macquarie) and Advanced Japanese Studies (Sheffield). Her areas of interest include academic writing and language use in engineering.



# Determining Covers in Combinational Circuits

Ljubomir Cvetković<sup>1</sup> and Darko Dražić<sup>2</sup>

<sup>1</sup> Teacher Training College  
Sremska Mitrovica, 22000, Serbia

<sup>2</sup> Teacher Training College  
Sremska Mitrovica, 22000, Serbia

## Abstract

In this paper we propose a procedure for determining 0– or 1– cover of an arbitrary line in a combinational circuit. When determining a cover we do not need Boolean expression for the line; only the circuit structure is used. Within the proposed procedure we use the tools of the cube theory, in particular, some operations defined on cubes. The procedure can be applied for determining 0– and 1– covers of output lines in programmable logic devices. Basically, this procedure is a method for the analysis of a combinational circuit.

**Keywords:** *Combinational Circuit, Cover, Logical Relation, Cube.*

## 1. Introduction

Traditionally, 0– or 1– cover of a line in a combinational circuit is determined using Boolean expression. There are well-known procedures for determining covers in the case when the function is given in the form of a minimal disjunctive normal form or minimal conjunctive normal form. In the case of disjunctive normal form a cube is associated to each elementary product and it represents the set of vectors on which this product has value 1. The 1– cover is determined on the basis of the correspondence between elementary products and cubes.

Getting 0– or 1– cover on the basis of truth table or Binary Decision Tree is a difficult task, especially in the case of a great number of variables.

Since we need Boolean expression for determining the cover of a line in a combinational circuit, some methods of minimization are quoted.

The Quine–McCluskey method is a program–based method that is able to carry out the exhaustive search for removing shared variables. The Quine–McCluskey method is a two step method which comprises of finding Prime Implicants and selecting a minimal set of Prime Implicants [5]. Each Boolean function can be represented by its

disjunctive normal form (DNF). A lot of Boolean function research has been devoted to minimal DNFs ([1], [8] and [9]). The generation of prime implicants (PIs) of a given function is an important first step in calculating its minimal DNF, and early interest in PIs was mainly inspired by this problem.

Generally, minimization of functions with a large number of input variables is a very time–consuming process and the results are often suboptimal. Most of the practical applications rely on heuristic minimization methods [6] with a complexity which is roughly quadratic in the number of products.

Using general DT structure, a new worst case algorithm to compute all prime implicants is presented in [4]. This algorithm has a lower time complexity than the well–known Quine–McCluskey algorithm and is the fastest corresponding worst case algorithm so far.

A SOP representation based on a “*ternary tree*” is well known. Compared to BDDs where the size can grow exponentially with the number of input variables, size of ternary tree grows only linearly with the number of inputs in the worst case. The first simple ternary tree minimization algorithms were proposed in [2], [3].

A method proposed in [7] utilizes data derived from Monte–Carlo simulations for any Boolean function with different count of variables and product term complexities. The model allows design feasibility and performance analysis prior to the circuit realization.

## 2. Preliminary considerations

Our procedure for determining covers in combinational circuits uses cube theory and therefore we provide necessary definitions.

A cube is a vector  $a_1a_2\dots a_n$ , where  $a_i \in \{0,1,X\}$  and  $X$  is a variable of the set  $\{0,1\}$  ( $i=1,2,\dots,n$ ). Hence, a cube is a set of vectors from  $\{0,1\}^n$ . Elements  $a_1, a_2, \dots, a_n$  are coordinates of the cube. A cube has rank  $r$ , if it contains  $r$  coordinates equal to  $X$ . A cube of rank  $r$  is called  $r$ -cube.

A set of cubes is called 0-cover (1-cover) of line  $i$  if it contains all input vectors generating signal of value 0 (value 1) on this line.

**Definition 1.** The intersection of cubes  $A = a_1a_2\dots a_n$  and  $B = b_1b_2\dots b_n$  is the cube  $C = c_1c_2\dots c_n$ , where  $c_i = a_i \oslash b_i$ ,  $i = 1, 2, \dots, n$ . The intersection operation  $\oslash$  is defined on the set  $\{0,1,X\}$  by Table 1. In Table 1 the symbol  $\oslash$  denotes that the operation  $\oslash$  is not defined. The intersection of cubes  $A$  and  $B$  is defined, if for any  $a_i$  and  $b_i$  the intersection operation is defined, i.e.  $a_i \oslash b_i \neq \oslash$ .

Table 1: Operation  $\oslash$

$\oslash$	0	1	X
0	0	$\oslash$	0
1	$\oslash$	1	1
X	0	1	X

**Definition 2.** The cut of cube sets  $Q_1$  and  $Q_2$  is denoted by  $Q_1 \cap Q_2$  and is the set of all cuts of a cube from  $Q_1$  with a cube from  $Q_2$ .

**Definition 3.** The union of cube sets  $Q_1$  and  $Q_2$  is denoted by  $Q_1 \cup Q_2$ . It contains all cubes from both  $Q_1$  and  $Q_2$ .

**Definition 4.** A cube  $B = b_1b_2\dots b_n$  is said to be a part of the cube  $A = a_1a_2\dots a_n$ , if all vectors of  $B$  belong also to  $A$ . Obviously,  $B$  is a part of  $A$  only if for any  $a_i \neq X$  we have  $a_i = b_i$ .

**Definition 5.** If a cube  $B$  is a part of the cube  $A$  and if both cubes belong to the same set of cubes, then  $B$  can be deleted from the considered set of cubes. This modification is called cube absorption. In particular, we say that  $A$  absorbs  $B$ . As noted, this is possible if for any  $a_i \neq X$  we have  $a_i = b_i$ .

Suppose that a cube generates a signal  $s \in \{0,1\}$  at a line  $i$  in combinational circuit which will be denoted by  $i=s$ . We shall say that the cube satisfies relation  $i=s$ , i.e. represents its solution.

Consider arbitrary lines  $i$  and  $j$  in combinational circuit.

Let  $s_i, s_j \in \{0,1\}$  be the signals at  $i$  and  $j$ , respectively. Then the following lemmas hold:

**Lemma 1.** If cubes  $A$  and  $B$  satisfy relations  $i=s_i$  and  $j=s_j$  respectively, then the cube  $C=A \oslash B$  satisfies relation  $(i=s_i) \wedge (j=s_j)$ .

**Proof.** The proof follows from the fact that the cut of cubes is equivalent to the cut of sets of vectors represented by these cubes.

**Lemma 2.** Let  $S_i, S_j$  be sets of cubes satisfying  $i=s_i, j=s_j$ , respectively, then all cubes of the set  $S_i \cup S_j$  satisfy relation  $(i=s_i) \vee (j=s_j)$ , while all cubes of the set  $S_i \oslash S_j$  satisfy relation  $(i=s_i) \wedge (j=s_j)$ .

**Proof.** The proof immediately follows from the definition of the union and the cut of cube sets.

Let  $S_{u_i}(0)$  and  $S_{u_i}(1)$  be cube sets generating signal values 0 and 1 on the input lines  $u_i, i=1,2,\dots,n$  of a logical element, considered either separately or within a combinational circuit. Based on Lemma 2 and properties of logical elements, one can formulate the following corollaries.

**Corollary 1.** In the case of elements OR and NOR the cut  $S_{u_1}(0) \oslash S_{u_2}(0) \oslash \dots \oslash S_{u_n}(0)$  represents the set of cubes generating on the output line  $v$  signal value 0 for element OR, and signal value 1 for element NOR. The union  $S_{u_1}(1) \cup S_{u_2}(1) \cup \dots \cup S_{u_n}(1)$  represents the set of cubes generating on the output line  $v$  signal value 1 for element OR, and signal value 1 for element NOR.

**Corollary 2.** In the case of elements AND or NAND the union  $S_{u_1}(1) \cup S_{u_2}(1) \cup \dots \cup S_{u_n}(1)$  represents the set of cubes generating on the output line  $v$  signal value 1 for element AND, and signal value 0 for element NAND. The cut  $S_{u_1}(0) \oslash S_{u_2}(0) \oslash \dots \oslash S_{u_n}(0)$  represents the set of cubes generating on the output line  $v$  signal value 0 for element AND, and signal value 1 for element NAND.

### 3. Determining a Cover

0- or 1-cover of input lines of the combinational circuit and of output lines of elements of the first level are determined directly (using basic rules for the logic elements).

0- or 1-cover of an arbitrary line of the combinational circuit, which is an output line of an element of the second or higher level is determined by the following two steps:

1. For an arbitrary line  $i$ , for which we want to determine a cover, we write logical relation defining conditions for generating the signal of given value. This logical relation is written on the basis of basic laws for the considered element. The left hand side of the relation determines signal values on all input lines of the element whose output line is the line  $i$ . For each line on the left hand side of the logical relation, we write a new logical relation defining conditions for generating the signal of expected value. We keep writing logical relations until we come to relations on whose left hand sides only input lines of the network or output lines of the first level appear.
2. For each line in left hand sides of the relation determined in step 1 we determine the distance in the following way. Input lines of the combinational circuit have the greatest distance  $r$ . For all lines at distance  $r-1$  we determine cube sets generating expected signal values on these lines. Next, for all lines at distance  $r-2$  we determine cube sets generating expected signal values on these lines using cube sets obtained for lines at distance  $r-1$ . We continue in this way until we get the cover for line  $i$ .

**Example 1.** Determine 1-cover of line  $i$  in the combinational circuit of Fig. 1.

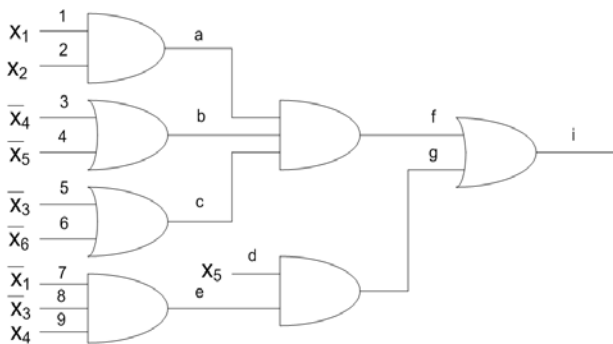


Fig. 1 Combinational circuit.

1. Logical relation defining conditions for generating the signal of value 1 reads:

$$(f = 1) \vee (g = 1) \rightarrow (i = 1) \quad (1)$$

Signals  $f=1$  and  $g=1$  are defined on the left hand side of the relation. The following logical relations define conditions for generating signals  $f=1$  and  $g=1$ :

$$(a = 1) \wedge (b = 1) \wedge (c = 1) \rightarrow (f = 1) \quad (2)$$

$$(d = 1) \wedge (e = 1) \rightarrow (g = 1) \quad (3)$$

Since output lines  $a, b, c, e$  of the first level and input line  $d$  have appeared on the left hand side of the above relations, we proceed to step 2.

2. We determine distances for all lines appearing on the left hand side of logical relations obtained in step 1. Input lines of the circuit 1-9 have the greatest distance. Lines  $a, b, c, d$  and  $e$  are at distance 2. Lines  $f$  and  $g$  are at distance 1.

We construct the table Table 2.

Table 2: Line at distance 1

a=1	11XXXX
b=1	XXX0XX XXXX0X
c=1	XX0XXX XXXXX0
d=1	XXXX1X
e=1	0X01XX

Necessary signal values at lines at distance 1 are  $f=1$  and  $g=1$ .

By the relation Eq. (2) we get 1-cover of line  $f$ :

$$\{11XXXX\} \cap \{XXX0XX\} \cap \{XX0XXX\} = \left\{ \begin{array}{l} 1100XX \\ 11X0X0 \\ 110X0X \\ 11XX00 \end{array} \right\}$$

By the relation Eq. (3) we get 1-cover of line  $g$ :

$$\{XXXX1X\} \cap \{0X01XX\} = \{0X011X\}$$

By the relation Eq. (1) and using operation  $\cup$  we get 1-cover of line  $i$ :

$$\left\{ \begin{array}{l} 1100XX \\ 11X0X0 \\ 110X0X \\ 11XX00 \end{array} \right\} \cup \{0X011X\} = \left\{ \begin{array}{l} 1100XX \\ 11X0X0 \\ 110X0X \\ 11XX00 \\ 0X011X \end{array} \right\}$$

If at least one cube can be deleted from a cover while the remaining cubes still form a cover, then the cover is redundant. In the other case the cover is irredundant.

The proposed procedure can be applied to combinational circuits with branchings of input and internal lines as well. Programmable logic devices (PLA, PAL and ROM) represent two-level combinational circuits. A PLA with  $n$  input lines,  $m$  internal lines and  $p$  output lines is represented in Fig.2.

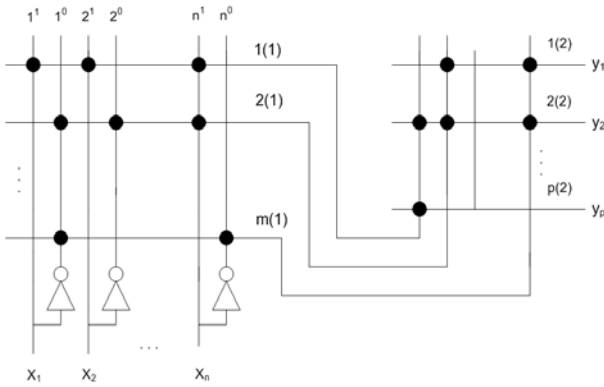


Fig. 2 Programmable logic array

Programmable elements are denoted by symbol “•”. We apply the following way of marking programmable points at PLA:

- $(i,j^1)$  - cross point of internal line  $i$  and a bit line  $j^1$  in AND array ( $i=1,2,\dots,m, j=1,2,\dots,n$ )
- $(i,j^0)$  - cross point of internal line  $i$  and a bit line  $j^0$  in AND array ( $i=1,2,\dots,m, j=1,2,\dots,n$ )
- $(i,j)$  - cross point between the lines  $i$  and  $j$  in OR array ( $i=1,2,\dots,p, j=1,2,\dots,m$ )

We propose the following procedure for determining 0- or 1-cover of a given output line  $i(2)=1,2,\dots,p$ .

1. We determine the set of test cubes  $Q_2(i)$ ,  $i=1,2,\dots,p$ , which yields 0- or 1-cover of the line  $i(2)=1,2,\dots,p$ , when applied on input lines of the OR array  $1,2,\dots,m$ . We have  $i(2)=0$  or  $i(2)=1$ , depending on whether 0- or 1-cover is determined. When determining the set  $Q_2(i)$ , we assign the coordinate  $X$  ( $X \in \{0,1\}$ ) to input lines of the OR array  $1,2,\dots,m$  which do not have cross points with the line  $i(2)$ .

2. The values from the set  $Q_2(i)$ ,  $i=1,2,\dots,p$ , obtained within step 1, are assigned using backtracking to output lines of the AND array  $i(1)$ ,  $i=1,2,\dots,m$  (when backtracking the signal complementation may occur). On the basis of signal values on the output lines of the AND array, the set of test cubes  $Q_1(i)$ ,  $i=1,2,\dots,m$ , is directly determined.

Using the cut  $\emptyset$  of cubes we have:

$$Q = Q_1(1) \emptyset Q_1(2) \emptyset \dots \emptyset Q_1(m) \quad (4)$$

Expanding the set of test cubes  $Q$  we obtain the input vectors representing 0- or 1-cover of line  $i(2)$ .

**Example 2.** Determine 1-cover of line 1(2) for the PAL of Fig. 3.

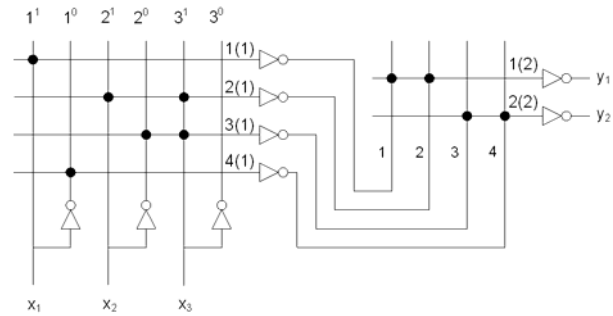


Fig. 3 Programmable array logic

We assign signal values 1 to the points (1,1) and (1,2). We get test cubes set

$$Q_2(1) = \{11XX\}$$

We go back towards output lines of the AND array and assign to these lines the values from  $Q_2(1)$ , as presented in Fig4.

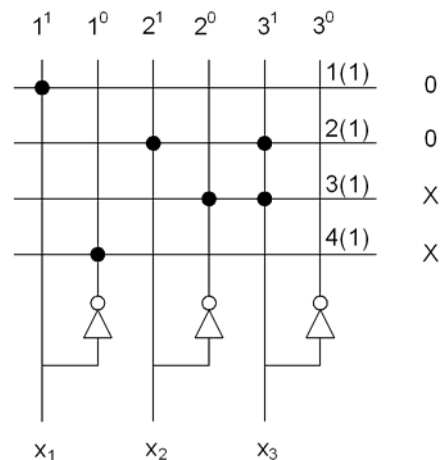


Fig. 4 Signals at output lines of the AND array

For lines 1(1) and 2(1) we determine:

$$Q_1(1) = \{0XX\} \quad Q_1(2) = \{X00, X01, X10\}$$

We apply the cut of cubes:

$$Q = Q_1(1) \emptyset Q_1(2) = \{000, 001, 010\}$$

Vectors 000, 001 and 010 represent a 1-cover of line 1(2).

## 4. Conclusion

The described procedure for determining covers is basically a method for the analysis of combinational circuits. When determining a cover we do not need an analytical expression; only the circuit structure is used. In particular, we do not need disjunctive normal forms, what is of some theoretical and practical importance. The simplification is in the fact that the cover is determined by moving from inputs towards output lines of the combinational circuit using only the operation of the cut of cubes. In addition, we present a procedure for determining covers of output lines for programmable logic devices. Within proposed procedures the cube theory plays an essential role.

## References

- [1] P. Clote, and E. Kranakis, *Boolean Functions and Computation Models*, Berlin Heidelberg: Springer Verlag, 2002.
- [2] P. Fišer, P. Rucký, and I. Váňová, "Fast Boolean Minimizer for Completely Specified Functions", Proc. 11th IEEE Design and Diagnostics of Electronic Circuits and Systems Workshop (DDECS '08), Bratislava, 2008, pp. 122-127.
- [3] Petr Fišer, and David Toman, "A Fast SOP Minimizer for Logic Functions Described by ManyProduct Terms", Proceedings of 12th Euromicro Conference on Digital System Design (DSD'09), Patras, 2009, pp. 757-764.
- [4] M. Friedel, S. Nikolajewa, and T. Wilhelm, "The Decomposition Tree for analyses of Boolean functions", *Math. Struct. in Comp. Science*, vol. 18, 2008, pp. 411-426.
- [5] E.J. McCluskey, "Minimization of Boolean functions", *The Bell System Technical Journal*, 35, No. 5, Nov. 1956, pp. 1417-1444.
- [6] A. Mishchenko, and T. Sasao, "Large-Scale SOP minimization Using Decomposition and Functional Properties", DAC 2003, pp. 149-154.
- [7] P.W. Chandana Prasad, and Azam Beg, and Ashutosh Kumar Singh, "Effect of Quine-McCluskey Simplification on Boolean Space Complexity", IEEE Conference on Innovative Technologies in Intelligent Systems & Industrial Applications, Bandar Sunway, 2009.
- [8] Y. Wang, and C. McCrosky, and X. Song, "Single-faced Boolean Functions and their Minimization", *Computer Journal* 44 (4), 2001, pp. 280-291.
- [9] I. Wegener, *Branching Programs and Binary Decision Diagrams – Theory and Application*, SIAM Monographs on Discrete Mathematics and Applications, Society for Industrial & Applied, 2000.

**Ljubomir Cvetković:** received the PhD degree from the University of Belgrade (Faculty of Electrical Engineering) in 2005. He is currently a professor in Teacher Training College in Sremska Mitrovica in Serbia. His major research interests include Digital VLSI architecture, Fault-tolerance and Fault detection. He has published about 20 publications in journals and international conferences and has written 3 books.

**Darko Dražić:** received the BSc degree from the University of Belgrade (Faculty of Organizational Sciences) in 2005. He is currently a PhD student on software engineering at Faculty of Organizational Sciences. His research interests include Computer architecture, Information system, Audio and video processing.



# Higher Order Programming to Mine Knowledge for a Modern Medical Expert System

Nittaya Kerdprasop and Kittisak Kerdprasop

Data Engineering and Knowledge Discovery (DEKD) Research Unit,  
School of Computer Engineering, Suranaree University of Technology,  
Nakhon Ratchasima 30000, Thailand

## Abstract

Knowledge mining is the process of deriving new and useful knowledge from vast volumes of data and background knowledge. Modern healthcare organizations regularly generate huge amount of electronic data stored in the databases. These data are a valuable resource for mining useful knowledge to help medical practitioners making appropriate and accurate decision on the diagnosis and treatment of diseases. In this paper, we propose the design of a novel medical expert system based on a logic-programming framework. The proposed system includes a knowledge-mining component as a repertoire of tools for discovering useful knowledge. The implementation of classification and association mining tools based on the higher order and meta-level programming schemes using Prolog has been presented to express the power of logic-based language. Such language also provides a pattern matching facility, which is an essential function for the development of knowledge-intensive tasks. Besides the major goal of medical decision support, the knowledge discovered by our logic-based knowledge-mining component can also be deployed as background knowledge to pre-treatment data from other sources as well as to guard the data repositories against constraint violation. A framework for knowledge deployment is also presented.

**Keywords:** Knowledge Mining, Association Mining, Decision-tree Induction, Higher-order Logic Programming, Medical Expert System.

## 1. Introduction

Knowledge is a valuable asset to most organizations as a substantial source to support better decisions and thus to enhance organizational competency. Researchers and practitioners in the area of knowledge management view knowledge in a broad sense as a state of mind, an object, a process, an access to information, or a capability [2, 13]. The term *knowledge asset* [24, 26] is used to refer to any organizational intangible property related to knowledge such as know-how, expertise, intellectual property. In clinical companies and computerized healthcare organizations knowledge assets include order sets, drug-

drug interaction rules, guidelines for practitioners, and clinical protocols [12].

Knowledge assets can be stored in data repositories either in implicit or explicit form. Explicit knowledge can be managed through the existing tools available in the current database technology. Implicit knowledge, on the contrary, is harder to achieve and retrieve. Specific tools and suitable environments are needed to extract such knowledge.

Implicit knowledge acquisition can be achieved through the availability of the knowledge-mining system. *Knowledge mining* is the discovery of hidden knowledge stored possibly in various forms and places in large data repositories. In health and medical domains, knowledge has been discovered in different forms such as association rules, classification trees, clustering means, trend or temporal patterns [27]. The discovered knowledge facilitates expert decision support, diagnosis and prediction. It is the current trend in the design and development of decision support systems [3, 16, 20, 31] to incorporate knowledge discovery as a tool to extract implicit information.

In this paper we present the design of a medical expert system and the implementation of knowledge mining component. Medical data mining is an emerging area of computational intelligence applied to automatically analyze electronic medical records and health databases. The non-hypothesis driven analysis approach of data mining technology can induce knowledge from clinical data repositories and health databases. Induced knowledge such as breast cancer recurrence conditions or diabetes implication is important not only to increase accurate diagnosis and successful treatment, but also to enhance safety and reduce medication-related errors.

A rapid prototyping of the proposed system is demonstrated in the paper to highlight the fact that higher order and meta-level programming are suitable schemes to



implement a complex knowledge-intensive system. For such a complicated system program coding should be done declaratively at a high abstraction level to alleviate the burden of programmers and to ease reasoning about program semantics.

The rest of this paper is organized as follows. Section 2 provides some preliminaries on two major knowledge-mining tasks, i.e. classification and association mining. Section 3 proposes the medical expert system design framework with the knowledge-mining component. Running examples on medical data set and the illustration on knowledge deployment are presented in Section 4. Section 5 discusses related work and then conclusions are drawn in Section 6. The implementation of knowledge-mining component is presented in the Appendix.

## 2. Preliminaries on Tree-based Classification and Association Mining

Decision tree induction [21] is a popular method for mining knowledge from medical data and representing the result as a classifier tree. Popularity is due to the fact that mining result in a form of decision tree is interpretability, which is more concern among medical practitioners than a sophisticated method but lack of understandability. A decision tree is a hierarchical structure with each node contains decision attribute and node branches corresponding to different attribute values of the decision node. The goal of building decision tree is to partition data with mixing classes down the tree until each leaf node contains data with pure class.

In order to build a decision tree, we need to choose the best attribute that contributes the most towards partitioning data to the purity groups. The metric to measure attribute's ability to partition data into pure class is *Info*, which is the number of bits required to encode a data mixture. The metric *Info* of positive (p) and negative (n) data mixture can be calculates as:

$$Info(P(p), P(n)) = -P(p)\log_2P(p) - P(n)\log_2P(n).$$

The symbols  $P(p)$  and  $P(n)$  are probabilities of positive and negative data instances, respectively. The symbol  $p$  represents number of positive data instances, and  $n$  is the negative cases. To choose the best attribute we have to calculate information gain, which is the yield we obtained from choosing that attribute. The information gain calculation of data with two classes (positive and negative) is given as:

$$Gain(Attribute) = Info\{p/(p+n), n/(p+n)\} - \sum_{i=1}^{10v} \{(p_i+n_i)/(p+n)\} Info\{p_i/(p_i+n_i), n_i/(p_i+n_i)\}.$$

The information gain calculates yield on *Info* of data set before splitting and *Info* after choosing attribute with  $v$  splits. The gain value of each candidate attribute is calculated, and then the maximum one has been chosen to be the decision node. The process of data partitioning continues until the data subset has the same class label.

Classification task based on decision-tree induction predicts the value of a target attribute or class, whereas association-mining task is a generalization of classification in that any attribute in the data set can be a target attribute. Association mining is the discovery of frequently occurred relationships or correlations between attributes (or items) in a database. Association mining problem can be decomposed as (1) find all sets of items that are frequent patterns, (2) use the frequent patterns to generate rules. Let  $I = \{i_1, i_2, i_3, \dots, i_m\}$  be a set of  $m$  items and  $DB = \{C_1, C_2, C_3, \dots, C_n\}$  be a database of  $n$  cases and each case contains items in  $I$ .

A *pattern* is a set of items that occur in a case. The number of items in a pattern is called the length of the pattern. To search for all valid patterns of length 1 up to  $m$  in large database is computational expensive. For a set  $I$  of  $m$  different items, the search space of all distinct patterns can be as huge as  $2^m - 1$ . To reduce the size of the search space, the *support* measurement has been introduced [1]. The function  $support(P)$  of a pattern  $P$  is defined as a number of cases in  $DB$  containing  $P$ . Thus,

$$support(P) = |\{T \mid T \in DB, P \subseteq T\}|.$$

A pattern  $P$  is called *frequent pattern* if the support value of  $P$  is not less than a predefined minimum support threshold  $minS$ . It is the  $minS$  constraint that helps reducing the computational complexity of frequent pattern generation. The  $minS$  metric has an anti-monotone property such that if the pattern contains an item that is not frequent, then none of the pattern's supersets are frequent. This property helps reducing the search space of mining frequent patterns in algorithm Apriori [1]. In this paper we adopt this algorithm as a basis for our implementation of association mining engine.

## 3. Medical Expert System Framework and the Knowledge Mining Engines

### 3.1 System Architecture

Health information is normally distributive and heterogeneous. Hence, we design the medical expert system (Figure 1) to include data integration component at the top level to collect data from distributed databases and also from documents in text format.

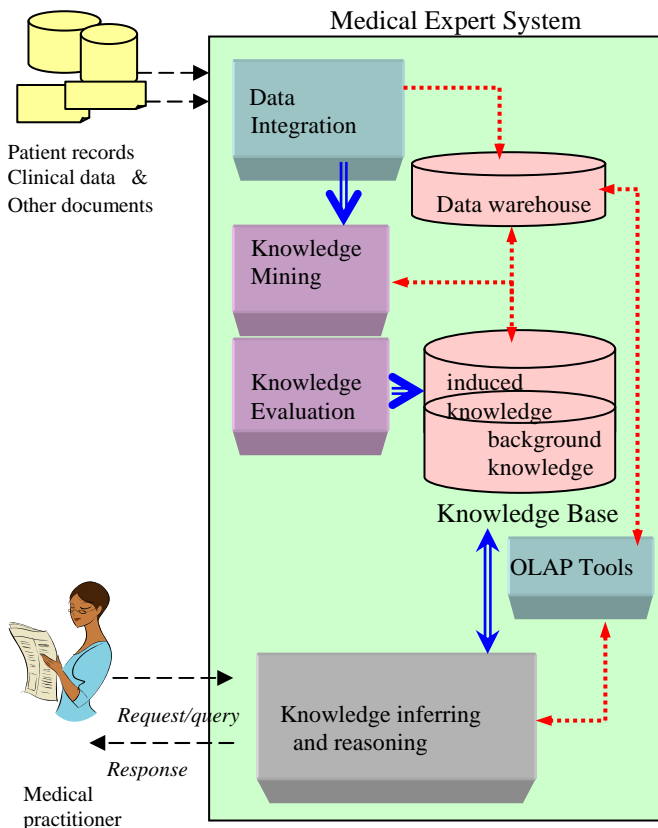


Fig. 1 Knowledge-mining component and a medical expert system framework. Double line arrows are process flow, whereas the dash line arrows are data flow.

The data integration component has been designed to input and select data with natural language processing. Data at this stage are to be stored in a warehouse to support direct querying (through OLAP tools) as well as to perform analyzing with knowledge mining engine.

Knowledge base in our design stores both induced knowledge in which its significance has to be evaluated by the domain expert, and background knowledge encoded from human experts. Knowledge inferring and reasoning is the module interfacing with medical practitioners and physicians at the front-end and accessing knowledge base at the back-end. The focus of this paper is on the implementation of knowledge-mining component, which currently contains classification and association mining engine.

### 3.2 Classification Mining Tool

Our classification mining engine is the implementation of decision-tree induction (ID3) algorithm [21]. The steps in our implementation are presented as follows:

#### Algorithm 1 Classification mining engine

**Input:** a data set formatted as Prolog clauses

**Output:** a decision tree with node and edge structures

- (1) Initialization
  - (1.1) Clear temporary knowledge base (KB) by removing all information regarding the predicates node, edge and current\_node
  - (1.2) Set node counter = 0
  - (1.3) Scan data set to get information about data attributes, positive instances, negative instances, total data instances
- (2) Building tree
  - (2.1) Increment node counter
  - (2.2) Repeat steps 2.2.1-2.2.4 until there is no more attributes left for creating decision attributes
    - (2.2.1) Compute the Info value of each candidate attribute
    - (2.2.2) Choose the attribute that yields minimum Info to be decision node
    - (2.2.3) Assert edge and node information into the knowledge base
    - (2.2.4) Split data instances along node branches
  - (2.3) Repeat steps 2.1 and 2.2 until the lists of positive and negative instances are empty
  - (2.4) Output a tree structure that contains node and edge predicates

The program source code is based on the syntax of SWI prolog ([www.swi-prolog.org](http://www.swi-prolog.org)).

```
main :-
    init(AllAttr, EdgeList), % initialize node
                                % and edge structures
    getNode(N),               % get node sequence number
    create_edge(N, AllAttr, EdgeList),
                                % recursively create tree
    print_model.              % print tree model
```

Classification mining engine is composed of two files *main* and *id3*. The main module (*main.pl*) calls initialization procedure (*init*) and starts creating edges and nodes of the decision tree. The data (*data.pl*) to be used by main module to create decision tree is also in a format of Prolog file. The mining engine induces data model of two classes: positive (class = yes) and negative (class = no). Binary classification is a typical task in medical domain. The code can be easily modified to classify data with more than two classes.

### 3.3 Association Mining Tool

The implementation of association mining engine is based primarily on the concept of higher-order Horn clauses. Such concept has been utilized through the predicates *maplist*, *include*, and *setof*.

The extensive use of these predicates contributes significantly to program conciseness and the ease of program verification. The program produces frequent patterns as a set of co-occurring items. To generate a nice representation of association rule such as  $X \Rightarrow Y$ , the list  $L$  in the predicate `association_mining` has to be further processed.

```
association_mining :-
    min_support(V), % set minimum support
    makeC1(C),      % create candidate 1-itemset
    makeL(C,L),     % compute large itemset
    apriori_loop(L,1). % recursively run apriori

makeC1(Ans):-
    input(D), % input data as a list
    allComb(1, ItemList, Ans2),
    % make combination of itemset
    maplist(countSS(D), Ans2, Ans).
    % scan database and pass countSS
    % to maplist

makeC(N, ItemSet, Ans) :-
    input(D), allComb(2, ItemSet, Ans1),
    maplist(flatten, Ans1, Ans2),
    maplist(list_to_ord_set, Ans2, Ans3),
    list_to_set(Ans3, Ans4),
    include(len(N), Ans4, Ans5), % include is
    % also a higher-order predicate
    maplist(countSS(D), Ans5, Ans).
    % scan database to find: List+N
```

#### 4. Running Examples and Knowledge Deployment

To show the running examples of our program coding, we use the following simple medical data represented as a Prolog file.

```
%% Data set: Allergy diagnosis
% Symptoms of disease and their possible values
attribute( soreThroat, [yes, no]).
attribute( fever, [yes, no]).
attribute( swollenGlands, [yes, no]).
attribute( congestion, [yes, no]).
attribute( headache, [yes, no]).
attribute( class, [yes, no]).
% Data instances
instance(1, class=no, [soreThroat=yes, fever=yes,
    swollenGlands=yes, congestion=yes,
    headache=yes]).
instance(2, class=yes, [soreThroat=no, fever=no,
    swollenGlands=no, congestion=yes,
    headache=yes]).
instance(3, class=no, [soreThroat=yes, fever=yes,
    swollenGlands=no, congestion=yes,
    headache=no]).
...
```

Data as shown are patient records suffering from allergy (class=yes). There are ten patient records in this simple data set: patient IDs 2, 6, and 8 are those who are suffering from allergy, whereas patient IDs 1, 3, 4, 5, 7, 9, 10 are suffering from other diseases but has shown some basic symptoms similar to allergy patients. To induce classification model for allergy patients from this data, we have to save this data set as a Prolog file (data.pl) and include this file name at the header declaration of the main program. By calling predicate `main`, the system should respond as `true`. At this moment we can view the tree model by calling `listing(node)`, then `listing(edge)` and get the following results.

```
1 ?- main.
true.
2 ?- listing(node).
:- dynamic user:node/2.
user:node(1, [2, 6, 8]-[1, 3, 4, 5, 7, 9, 10]).
user:node(2, []-[1, 3, 5, 9, 10]).
user:node(3, [2, 6, 8]-[4, 7]).
user:node(4, []-[4, 7]).
user:node(5, [2, 6, 8]-[]).
true.
3 ?- listing(edge).
:- dynamic user:edge/3.
user:edge(0, root-nil, 1).
user:edge(1, fever=yes, 2).
user:edge(1, fever=no, 3).
user:edge(3, swollenGlands=yes, 4).
user:edge(3, swollenGlands=no, 5).
true.
```

The node and edge structures have the following formats:  
`node(nodeID, [Positive_Cases]-[Negative_Cases])`  
`edge(ParentNode, EdgeLabel, ChildNode)`

The node structure is a tuple of `nodeID` and a mixture of positive and negative cases represented as a list pattern: `[Positive_Cases]-[Negative_Cases]`. Node 0 is a special node, representing root node of the tree. Node 1 contains a mixture of ten patients, whereas node 5 is a pure group of allergy patients. The edges leading from node 1 to node 5 capture the model of allergy patients. Therefore, the classification result represents the following data model:

```
class(allergy) :- fever=no, swollenGlands=no.
```

This model is represented as a Horn clause, thus, it provide flexibility of including this clause as a rule to select data in other group of patients who are suffering from throat infection. This kind of infection shows the same basic symptoms as allergy; therefore, screening data with the above rule can help focusing only on throat infection cases.

Applying the same data set with association mining and setting minimum support value = 50%, we got the following frequent patterns:

- {fever=yes & class=no}
- {fever=yes & congestion=yes}
- {swollenGlands=no & congestion=yes}
- {congestion=yes & headache=yes}
- {congestion=yes & class=no}
- {fever=yes & congestion=yes & class=no}

The first pattern can be interpreted as association rule as “if patient has fever, that the patient does not suffer from allergy.” This kind of rule can help accurately diagnosing patients with symptoms very close to allergy.

**Knowledge Deployment: Example 1.**

We suggest that such discovered rules, after confirming their correctness by human experts, can be added into the database system as trigger rules (Figure 2). The triggers guard database content against any updates that violates the rules. Any attempt to insert violating data will raise an error message to draw attention from the database administrator. Such trigger rules are thus deployed as a tool to enforce database integrity checking.

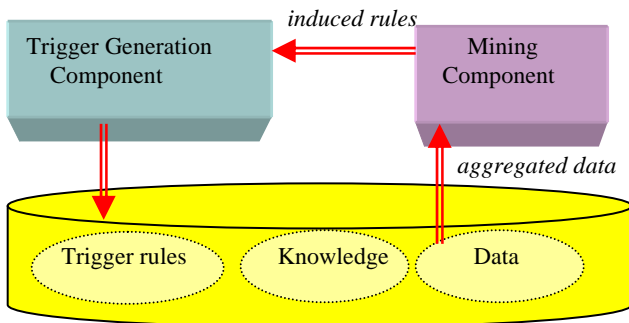


Fig. 2 The framework of knowledge deployment as triggers in a medical database.

```

1 - WordPad
File Edit View Insert Format Help
% 1.knb
% for expert shell. --- written by Postprocess
% top_goal where the inference starts.

top_goal(X,V) :- type(X,V).

type(no,0.5):-fever(yes). % generated rule
type(yes,0.3):-fever(no),swollenGlands(no). % generated rule
type(no,0.2):-fever(no),swollenGlands(yes). % generated rule

soreThroat(X):-menuask(soreThroat,X,[yes,no]). %generated menu
fever(X):-menuask(fever,X,[yes,no]). %generated menu
swollenGlands(X):-menuask(swollenGlands,X,[yes,no]). %generated menu
congestion(X):-menuask(congestion,X,[yes,no]). %generated menu
headache(X):-menuask(headache,X,[yes,no]). %generated menu
class(X):-menuask(class,X,[yes,no]). %generated menu

%end of automatic post process
    
```

Fig. 3 The content of automatically induced knowledge base.

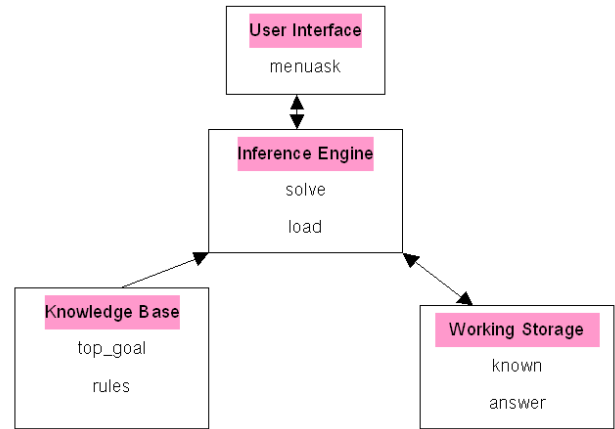


Fig. 4 Structure of a simple expert system shell with the induced knowledge base.

```

SWI-Prolog -- d:/3-2009-Nittaya/3-Research-Grants/NRCT/ชุดโครงการ/Fin
File Edit Settings Run Debug Help

For help, use ?- help(Topic). or ?- apropos(Word).

1 ?- expertshell.
This is the Easy Expert System shell.
Type help. load. solve. why. quit. or 99.
at the prompt.
expert-shell> load.
Enter file name in single quotes (ex. '1.knb'.): '1.knb'.
% 1.knb compiled 0.00 sec, 2.628 bytes
expert-shell> solve.

What is the value for fever?
[1=yes, 2=no, 99=exitShell]
Enter the choice> 2.

What is the value for swollenGlands?
[1=yes, 2=no, 99=exitShell]
Enter the choice> 2.
The answer is __yes__ with probability 0.3
expert-shell> why.

The answer is ...yes... with probability = 0.3.
The known storage are
[swollenGlands(no), fever(no)]
expert-shell>
    
```

Fig. 5 A snapshot of medical expert system inductively created from the allergy data set.

**Knowledge Deployment: Example 2.**

The induced knowledge once confirmed by the domain expert can be added to the knowledge base of the expert system shell. We illustrate the knowledge base that automatically created from the induced tree in Figure 3. This expert system shell has simple structure as diagrammatically shown in Figure 4. User can interact with the system through a line command as shown in Figure 5, in which the user can ask for further explanation by typing the ‘why’ command.

**5. Related Work**

In recent years we have witnessed increasing number of applications devising database technology and machine learning techniques to mine knowledge from biomedicine, clinical and health data. Roddick et al [22] discussed the two categories of mining techniques applied over medical

data: explanatory and exploratory. Explanatory mining refers to techniques that are used for the purpose of confirmation or making decisions. Exploratory mining is data investigation normally done at an early stage of data analysis in which an exact mining objective has not yet been set.

Explanatory mining in medical data has been extensively studied in the past decade employing various learning techniques. Bojarczuk et al [4] applied genetic programming method with constrained syntax to discover classification rules from medical data sets. Thongkam et al [28] studied breast cancer survivability using AdaBoost algorithm. Ghazavi and Liao [9] proposed the idea of fuzzy modeling on selected features of medical data. Huang et al [11] introduced a system to apply mining techniques to discover rules from health examination data. Then they employed a case-based reasoning to support the chronic disease diagnosis and treatments. The recent work of Zhuang et al [31] also combined mining with case-based reasoning, but applied a different mining method. They performed data clustering based on self-organizing maps in order to facilitate decision support on solving new cases of pathology test ordering problem. Biomedical discovery support systems are recently proposed by a number of researchers [5, 6, 10, 29, 30]. Some work [20, 25] extended medical databases to the level of data warehouses.

Exploratory, as oppose to explanatory, is rarely applied to medical domains. Among the rare cases, Nguyen et al [19] introduced knowledge visualization in the study of hepatitis patients. Palaniappan and Ling [20] applied the functionality of OLAP tools to improve visualization in data analysis.

It can be seen from the literature that most medical knowledge discovery systems have applied only some mining techniques to discover hidden knowledge with the main purpose to support medical diagnosis [4, 14, 17]. Some researchers [3, 8, 15, 16] have extended the knowledge discovery aspect to the large scale of a medical decision support system.

Our work is also in the main stream of medical decision support system development, but our methodology is different from those appeared in the literature. The system proposed in this paper is based on a logic-programming paradigm. The justification of our logic-based system is that the closed form of Horn clauses that treats program in the same way as data facilitates fusion of knowledge learned from different sources, which is a normal setting in medical domain. Knowledge reuse can easily practice in this framework.

The declarative style of our implementation also eases the future extension of the proposed medical support system to cover the concepts of higher-order mining [23], i.e. mining from the discovered knowledge, and constraint mining [7], i.e. mining with some specified constraints to obtain relevant knowledge.

## 6. Conclusions and Discussion

Modern healthcare organizations generate huge amount of electronic data stored in heterogeneous databases. Data collected by hospitals and clinics are not yet turned into useful knowledge due to the lack of efficient analysis tools. We thus propose a rapid prototyping of automatic mining tools to induce knowledge from medical data. The induced knowledge is to be evaluated and integrated into the knowledge base of a medical expert system. Discovered knowledge facilitates the reuse of knowledge base among decision-support applications within organizations that own heterogeneous clinical and health databases. Direct application of the proposed system is for medical related decision-making. Other indirect but obvious application of such knowledge is to pre-process other data sets by grouping it into focused subset containing only relevant data instances.

The main contribution of this work is our implementation of knowledge mining engines based on the concept of higher-order Horn clauses using Prolog language. Higher-order programming has been originally appeared in functional languages in which functions can be passed as arguments to other functions and can also be returned from other functions. This style of programming has soon been ubiquitous in several modern programming languages such as Perl, PHP, and JavaScript. Higher order style of programming has shown the outstanding benefits of code reuse and high level of abstraction. This paper illustrates higher order programming techniques in SWI-Prolog. The powerful feature of meta-level programming in Prolog facilitates the reuse of mining results represented as rules to be flexibly applied as conditional clauses in other applications.

The plausible extensions of our current work are to add constraints into the knowledge mining method in order to limit the search space and therefore yield the most relevant and timely knowledge, and due to the uniform representation of Prolog's statements as a clausal form, mining from the previously mined knowledge should be implemented naturally. We also plan to extend our system to work with stream data that normally occur in modern medical organizations.



## Appendix

The implementation of knowledge-mining component is based on the concept of higher-order and meta-programming styles. Higher-order programming in Prolog refers to Horn clauses that can quantify over other predicate symbols [18]. Meta-level programming is also another powerful feature of Prolog. Data and program in Prolog take the same representational format; that is clausal form. Higher-order and meta-level clauses in the following source code are typed in bold face.

```

/* Classification mining engine */
:- include('data.pl').
:- dynamic current_node/1, node/2, edge/3.

main :-
    init(AllAttr, EdgeList),
    getNode(N),          % get node sequence number
    create_edge(N, AllAttr, EdgeList),
    print_model.

init(AllAttr, [root-nil/PB-NB]) :-
    retractall(node(_, _)),
    retractall(current_node(_)),
    retractall(edge(_, _, _)),
    assert(current_node(0)),
    forall(X, attribute(X, _), AllAttr1),
    delete(AllAttr1, class, AllAttr),
    forall(X2, instance(X2, class=yes, _), PB),
    forall(X3, instance(X3, class=no, _), NB).

getNode(X) :-
    current_node(X), X1 is X+1,
    retractall(current_node(_)),
    assert(current_node(X1)).

create_edge(_, _, []) :- !.
create_edge(_, [], _) :- !.

create_edge(N, AllAttr, EdgeList) :-
    create_nodes(N, AllAttr, EdgeList).

create_nodes(_, _, []) :- !.
create_nodes(_, [], _) :- !.

create_nodes(N, AllAttr, [H1-H2/PB-NB|T]) :-
    getNode(N1), % get node sequence number N1
    assert(edge(N, H1-H2, N1)), % H1-H2 is
                                % a pattern
    assert(node(N1, PB-NB)), % PB-NB is
                              % a pattern
    append(PB, NB, AllInst),
    ((PB \== [], NB \== []) -> % if-condition
     % then clauses
     (cand_node(AllAttr, AllInst, AllSplit),
      best_attribute(AllSplit, [V, MinAttr,
                               Split]),
      delete(AllAttr, MinAttr, Attr2),
      create_edge(N1, Attr2, Split))
     ; % else clause
     true ),
    create_nodes(N, AllAttr, T).

%

```

```

% select best attribute to be a decision node
%
best_attribute([], Min, Min).

best_attribute([H|T], Min) :-
    best_attribute(T, H, Min).

best_attribute([H|T], Min0, Min) :-
    H = [V, _, _],
    Min0 = [V0, _, _],
    ( V < V0 -> Min1 = H ; Min1 = Min0),
    best_attribute(T, Min1, Min).

%
% generate candidate decision node
%
cand_node([], _, []) :- !.

cand_node(_, [], []).

cand_node([H|T], CurInstL, [[Val,H,SplitL]
                             |OtherAttr]) :-
    info(H, CurInstL, Val, SplitL),
    cand_node(T, CurInstL, OtherAttr).

%
% compute Info of each candidate node
%
info(A, CurInstL, R, Split) :-
    attribute(A,L),
    maplist(concat3(A,=), L, LI),
    suminfo(LI, CurInstL, R, Split).

concat3(A,B,C,R) :-
    atom_concat(A,B,R1),
    atom_concat(R1,C,R).

suminfo([],_,0,[]).

suminfo([H|T], CurInstL, R, [Split | ST]) :-
    AllBag = CurInstL, term_to_atom(H1, H),
    forall(X1, (instance(X1, _, LI),
               member(X1, CurInstL),
               member(H1, LI)), BagGro),
    forall(X2, (instance(X2, class=yes, L2),
               member(X2, CurInstL),
               member(H1, L2)), BagPos),
    forall(X3, (instance(X3, class=no, L3),
               member(X3, CurInstL),
               member(H1, L3)), BagNeg),
    (H11= H22) = H1,
    length(AllBag, Nall),
    length(BagGro, NGro),
    length(BagPos, NPos),
    length(BagNeg, NNeg),
    Split = H11-H22/BagPos-BagNeg,
    suminfo(T, CurInstL, R1, ST),
    ( NPos is 0 *-> L1 = 0;
      L1 is (log(NPos/NGro)/log(2)) ),
    ( 0 is NNeg *-> L2 = 0;
      L2 is (log(NNeg/NGro)/log(2)) ),
    ( NGro is 0 -> R = 999;
      R is (NGro/Nall)*
            (-NPos/NGro)*
            L1- (NNeg/NGro)*L2)+R1).

/* ===== */
/* Association mining engine */

```



```
/* ===== */
association_mining:-
  min_support(V), % set minimum support
  makeC1(C), % create candidate 1-itemset
  makeL(C,L), % compute large itemset
  apriori_loop(L,1). % recursively run apriori

apriori_loop(L, N) :- % base case of recursion
  length(L) is 1,!.

apriori_loop(L, N) :- % inductive step
  N1 is N+1,
  makeC(N1, L, C),
  makeL(C, Res),
  apriori_loop(Res, N1).

makeC1(Ans):-
  input(D), % input data as a list,
  % e.g. [[a], [a,b]]
  % then make combination of itemset
  allComb(1, ItemList, Ans2),
  % scan database and pass countSS to maplist
  maplist(countSS(D), Ans2, Ans).

makeC(N, ItemSet, Ans) :- input(D),
  allComb(2, ItemSet, Ans1),
  maplist(flatten, Ans1, Ans2),
  maplist(list_to_ord_set, Ans2, Ans3),
  list_to_set(Ans3, Ans4),
  include(len(N), Ans4, Ans5),
  % include is also a
  % higher-order predicate
  maplist(countSS(D), Ans5, Ans).
  % scan database to find: List+N

makeL(C, Res):- % for all large itemset creation
  % call higher-order predicates
  % include and maplist
  include(filter, C, Ans),
  maplist(head, Ans, Res).

%
% filter and head are for pattern matching of
% data format
%
filter(_+N):-
  input(D),
  length(D,I),
  min_support(V),
  N>=(V/100)*I.

head(H+_, H).

%
% an arbitrary subset of the set containing
% given number of elements
%
comb(0, _, []).

comb(N, [X|T], [X|Comb]) :-
  N>0, N1 is N-1,
  comb(N1, T, Comb).

comb(N, [_|T], Comb) :-
  N>0,
  comb(N, T, Comb).
allComb(N, I, Ans) :-
  setof(L, comb(N, I, L), Ans).
```

```
countSubset(A, [], 0).

countSubset(A, [B|X], N) :-
  not(subset(A, B)),

countSubset(A, X, N).

countSubset(A, [B|X], N) :-
  subset(A, B),
  countSubset(A, X, N1),
  N is N1+1.

countSS(SL, S, S+N) :-
  countSubset(S, SL, N).
```

## Acknowledgments

This work has been fully supported by research fund from Suranaree University of Technology granted to the Data Engineering and Knowledge Discovery (DEKD) research unit. This research is also supported by grants from the National Research Council of Thailand (NRCT) and the Thailand Research Fund (TRF).

## References

- [1] R. Agrawal, and R. Srikant, "Fast algorithm for mining association rules", in: *Proc. VLDB*, 1994, pp.487-499.
- [2] M. Alavi, and D.E. Leidner, "Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues", *MIS Quarterly*, Vol.25, No.1, 2001, pp.107-136.
- [3] Y. Bedard et al., "Integrating GIS components with knowledge discovery technology for environmental health decision support", *Int. J Medical Informatics*, Vol.70, 2003, pp.79-94.
- [4] C.C. Bojarczuk et al., "A constrained-syntax genetic programming system for discovering classification rules: Application to medical data sets", *Artificial Intelligence in Medicine*, Vol.30, 2004, pp.27-48.
- [5] C. Bratsas et al., "KnowBaSICS-M: An ontology-based system for semantic management of medical problems and computerised algorithmic solutions", *Computer Methods and Programs in Biomedicine*, Vol.83, 2007, pp.39-51.
- [6] R. Correia et al., "Borboleta: A mobile telehealth system for primary homecare", in: *Proc. ACM Symposium on Applied Computing*, 2008, pp.1343-1347.
- [7] L. De Raedt et al., "Constraint programming for itemset mining", in: *Proc. KDD*, 2008, pp.204-212.
- [8] E. German et al., "An architecture for linking medical decision-support applications to clinical databases and its evaluation", *J. Biomedical Informatics*, Vol.42, 2009, pp.203-218.
- [9] S. Ghazavi and T.W. Liao, "Medical data mining by fuzzy modeling with selected features", *Artificial Intelligence in Medicine*, Vol.43, No.3, 2008, pp.195-206.
- [10] D. Hristovski et al., "Using literature-based discovery to identify disease candidate genes", *Int. J Medical Informatics*, Vol.74, 2005, pp.289-298.

- [11] M.J. Huang et al., "Integrating data mining with case-based reasoning for chronic diseases prognosis and diagnosis", *Expert Systems with Applications*, Vol.32, 2007, pp.856-867.
- [12] N.C. Hulse et al., "Towards an on-demand peer feedback system for a clinical knowledge base: A case study with order sets", *J Biomedical Informatics*, Vol.41, 2008, pp.152-164.
- [13] N.K. Kakabadse et al., "From tacit knowledge to knowledge management: Leveraging invisible assets", *Knowledge and Process Management*, Vol. 8, No. 3, 2001, pp.137-154.
- [14] E. Kretschmann et al., "Automatic rule generation for protein annotation with the C4.5 data mining algorithm applied on SWISS-PROT", *Bioinformatics*, Vol.17, No.10, 2001, pp.920-926.
- [15] P.-J. Kwon et al., "A study on the web-based intelligent self-diagnosis medical system", *Advances in Engineering Software*, Vol.40, 2009, pp.402-406.
- [16] C. Lin et al., "A decision support system for improving doctors' prescribing behavior", *Expert Systems with Applications*, Vol.36, 2009, pp.7975-7984.
- [17] E. Mugambi et al., "Polynomial-fuzzy decision tree structures for classifying medical data", *Knowledge-Based System*, Vol.17, No.2-4, 2004, pp.81-87.
- [18] G. Nadathur, and D. Miller, "Higher-order Horn clauses", *J ACM*, Vol.37, 1990, pp.777-814.
- [19] D. Nguyen et al., "Knowledge visualization in hepatitis study", in: *Proc. Asia-Pacific Symposium on Information Visualization*, 2006, pp.59-62.
- [20] S. Palaniappan, and C.S. Ling, "Clinical decision support using OLAP with data mining", *Int. J Computer Science and Network Security*, Vol.8, No.9, 2008, pp.290-296.
- [21] J.R. Quinlan, "Induction of decision trees", *Machine Learning*, Vol.1, 1986, pp.81-106.
- [22] J.F. Roddick et al., "Exploratory medical knowledge discovery: experiences and issues", *ACM SIGKDD Explorations Newsletter*, Vol.5, No.1, 2003, pp.94-99.
- [23] J.F. Roddick et al., "Higher order mining", *ACM SIGKDD Explorations Newsletter*, Vol.10, No.1, 2008, pp.5-17.
- [24] C.P. Ruppel, and S.J. Harrington, "Sharing knowledge through intranets: A study of organizational culture and intranet implementation", *IEEE Transactions on Professional Communication*, Vol.44, No.1, 2001, pp.37-51.
- [25] T.R. Sahama, and P.R. Croll, "A data warehouse architecture for clinical data warehousing", in: *Proc. 12<sup>th</sup> Australasian Symposium on ACSW Frontiers*, 2007, pp.227-232.
- [26] A. Satyadas et al., "Knowledge management tutorial: An editorial overview", *IEEE Transactions on Systems, Man and Cybernetics*, Part C, Vol.31, No.4, 2001, pp.429-437.
- [27] A. Shillabeer, and J.F. Roddick, "Establishing a lineage for medical knowledge discovery", in: *Proc. 6<sup>th</sup> Australasian Conf. on Data Mining and Analytics*, 2007, pp.29-37.
- [28] J. Thongkam et al., "Breast cancer survivability via AdaBoost algorithms", in: *Proc. 2<sup>nd</sup> Australasian Workshop on Health Data and Knowledge Management*, 2008, pp.55-64.
- [29] N. Uramoto et al., "A text-mining system for knowledge discovery from biomedical documents", *IBM Systems J*, Vol.43, No.3, 2004, pp.516-533.
- [30] X. Zhou et al., "Text mining for clinical Chinese herbal medical knowledge discovery", in: *Proc. 8<sup>th</sup> Int. Conf. on Discovery Science*, 2005, pp.396-398.
- [31] Z.Y. Zhuang et al., "Combining data mining and case-based reasoning for intelligent decision support for pathology ordering by general practitioners", *European J Operational Research*, Vol.195, No.3, 2009, pp.662-675.

**Nittaya Kerdprasop** is an associate professor at the school of computer engineering, Suranaree University of Technology, Thailand. She received her B.S. in radiation techniques from Mahidol University, Thailand, in 1985, M.S. in computer science from the Prince of Songkla University, Thailand, in 1991 and Ph.D. in computer science from Nova Southeastern University, USA, in 1999. She is a member of IAENG, ACM, and IEEE Computer Society. Her research of interest includes Knowledge Discovery in Databases, Data Mining, Artificial Intelligence, Logic and Constraint Programming, Deductive and Active Databases.

**Kittisak Kerdprasop** is an associate professor and the director of DEKD (Data Engineering and Knowledge Discovery) research unit at the school of computer engineering, Suranaree University of Technology, Thailand. He received his bachelor degree in Mathematics from Srinakarinwirot University, Thailand, in 1986, master degree in computer science from the Prince of Songkla University, Thailand, in 1991 and doctoral degree in computer science from Nova Southeastern University, USA, in 1999. His current research includes Data mining, Machine Learning, Artificial Intelligence, Logic and Functional Programming, Probabilistic Databases and Knowledge Bases.

# A New Proxy Blind Signature Scheme based on ECDLP

Daniyal M. Alghazzawi<sup>1</sup>, Trigui Mohamed Salim<sup>2</sup> and Syed Hamid Hasan<sup>3</sup>

<sup>1,2,3</sup> Department of Information Systems,  
King Abdul Aziz University, Kingdom of Saudi Arabia

## Abstract

A proxy blind signature scheme is a special form of blind signature which allows a designated person called proxy signer to sign on behalf of two or more original signers without knowing the content of the message or document. It combines the advantages of proxy signature, blind signature and multi-signature scheme and satisfies the security properties of both proxy and blind signature scheme. Most of the exiting proxy blind signature schemes were developed based on the mathematical hard problems integer factorization (IFP) and simple discrete logarithm (DLP) which take sub-exponential time to solve. This paper describes an secure simple proxy blind signature scheme based on Elliptic Curve Discrete Logarithm Problem (ECDLP) takes fully-exponential time. This can be implemented in low power and small processor mobile devices such as smart card, PDA etc. Here also we describes implementation issues of various scalar multiplication for ECDLP

Keywords: ECDLP, IFP, blind signature, proxy signature.

## 1. Introduction

Blind signature scheme was first introduced by Chaum [2]. It is a protocol for obtaining a signature from a signer, but the signer can neither learn the messages nor the signatures. The recipients obtain afterwards. In 1996, mammo et al proposed the concept of proxy signature [1]. In proxy signature scheme, the original signer delegates his signing capacity to a proxy signer who can sign a message submitted on behalf of the original signer. A verifier can validate its correctness and can distinguish between a normal signature and a proxy signature. A proxy blind signature scheme is a digital signature scheme that ensures the properties of proxy signature and blind signature. In a proxy blind signature, an original signer delegates his signing capacity to proxy signer.

## 2. Preliminaries

### 2.1 Notations

Common notations used in this paper as follows:

- $p$  : The order of underlying finite field.
- $F_p$  : the underlying finite field of order  $p$
- $E$ : elliptic curve defined on finite field  $F_p$  with large order.
- $G$ : the group of elliptic curve points on  $E$ .
- $P$ : a point in  $E(F_p)$  with order  $n$ , where  $n$  is a large prime number.
- $H(\cdot)$ : a secure one-way hash function.
- $d$ : the secret key of the original signer  $S$  to be chosen randomly from  $[1, n - 1]$ .
- $Q$  is the public key of the original signer  $S$ , where  $Q = d \cdot G$ .
- $k$ : Concatenation operation between two bit strings.

## 3. Backgrounds

In this section we brief overview of prime field, Elliptic Curve over that field and Elliptic Curve Discrete Logarithm Problem.

### 3.1 The finite field $F_p$

Let  $p$  be a prime number. The finite field  $F_p$  is comprised of the set of integers  $0, 1, 2, \dots, p-1$  with the following arithmetic operations [4] [5] [6]:

- Addition: If  $a, b \in F_p$ , then  $a + b = r$ , where  $r$  is the remainder when  $a + b$  is divided by  $p$  and  $0 \leq r \leq p-1$ . This is known as addition modulo  $p$ .
- Multiplication: If  $a, b \in F_p$ , then  $a.b = s$ , where  $s$  is the remainder when  $a.b$  is divided by  $p$  and  $0 \leq s \leq p-1$ . This is known as multiplication modulo  $p$ .
- Inversion: If  $a$  is a non-zero element in  $F_p$ , the inverse of  $a$  modulo  $p$ , denoted  $a^{-1}$ , is the unique integer  $c \in F_p$  for which  $a.c = 1$ .

### 3.2 Elliptic Curve over $F_p$

Let  $p, 3$  be a prime number. Let  $a, b \in F_p$  be such that  $4a^3 + 27b^2 \neq 0$  in  $F_p$ . An elliptic curve  $E$  over  $F_p$  defined by the parameters  $a$  and  $b$  is the set of all solutions  $(x, y)$ ,  $x, y \in F_p$ , to the equation  $y^2 = x^3 + ax + b$ , together with an extra point  $O$ , the point at infinity. The set of points  $E(F_p)$  forms an abelian group with the following addition rules [8]:

1. Identity :  $P + O = O + P = P$ , for all  $P \in E(F_p)$
2. Negative: if  $P(x, y) \in E(F_p)$  then  $(x, y) + (x, -y) = O$ , The point  $(x, -y)$  is denoted as  $-P$  called negative of  $P$ .
3. Point addition: Let  $P(x_1, y_1), Q(x_2, y_2) \in E(F_p)$ , then  $P + Q = R \in E(F_p)$  and coordinate  $(x_3, y_3)$  of  $R$  is given by  $x_3 = \lambda^2 - x_1 - x_2$  and  $y_3 = \lambda(x_1 - x_3) - y_1$ .

$$\text{Where } \lambda = \frac{(y_2 - y_1)}{(x_2 - x_1)}$$

4. Point doubling: Let  $P(x_1, y_1) \in E(K)$  where  $P \neq -P$  then  $2P = (x_3, y_3)$  where  $x_3 = (3x_1^2 + a) / 2y_1 - 2x_1$  and

$$y_3 = (3x_1^2 + a) / 2y_1 (x_1 - x_3) - y_1.$$

### 3.3 Elliptic Curve Discrete Logarithm Problem (ECDLP)

Given an elliptic curve  $E$  defined over a finite field  $F_p$ , a point  $P \in E(F_p)$  of order  $n$ , and a point  $Q \in \langle P \rangle$ ,

find the integer  $l \in [0, n-1]$  such that  $Q = lP$ . The integer  $l$  is called discrete logarithm of  $Q$  to base  $P$ , denoted  $l = \log_p Q$  [8].

## 4. Proxy Signatures and Proxy Blind Signature

A proxy blind signature is a digital signature scheme that ensures the properties of proxy signature and blind signature schemes. Proxy blind signature scheme is an extension of proxy blind signature, which allows a single designated proxy signer to generate a blind signature on behalf of group of original signers. A proxy blind signature scheme consists of the following three phases[9]:

- Proxy key generation
- Proxy blind multi-signature scheme
- Signature verification

## 5. Security properties

The security properties for a secure blind multi-signature scheme are as follows [9]

- **Distinguishability:** The proxy blind multi-signature must be distinguishable from the ordinary signature.
- **Strong unforgeability:** Only the designated proxy signer can create the proxy blind signature for the original signer.
- **Non-repudiation:** The proxy signer can not claim that the proxy signer is disputed or illegally signed by the original signer.
- **Verifiability:** The proxy blind multi-signature can be verified by everyone. After verification, the verifier can be convinced of the original signer's agreement on the signed message.
- **Strong undeniability:** Due to fact that the delegation information is signed by the original signer and the proxy signature are generated by the proxy signer's secret key. Both the signer can not deny their behavior.
- **Unlinkability:** When the signer is revealed, the proxy signer can not identify the association between the message and the blind signature he generated.
- **Secret key dependencies:** Proxy key or delegation pair can be computed only by the original signer's secret key.
- **Prevention of misuse:** The proxy signer cannot use the proxy secret key for purposes other than

generating valid proxy signatures. In case of misuse, the responsibility of the proxy signer should be determined explicitly.

## 6. Proposed Protocol

The protocol involves three entities: Original signer  $S$ , Proxy signer  $P_s$  and verifier  $V$ . It is described as follows.

### 6.1 Proxy Phase

- **Proxy generation:** The original signer  $S$  selects random integer  $k$  in the interval  $[1, n-1]$ . Computes  $R = k.P$  and  $r = x_1 \bmod n$ . Where  $x_1$  is regarded as an integer between 0 and  $q-1$ . Then computes  $s = (d + k.r) \bmod n$  and computes  $Q_p = s.P$ .
- **Proxy delivery:** The original signer  $S$  sends  $(s, r)$  to the proxy signer  $P_s$  and make  $Q_p$  public.
- **Proxy Verification:** After receiving the secret key pairs  $(s, r)$ , the proxy signer  $P_s$  checks the validity of the secret key pairs  $(s, r)$  with the following equation.

$$Q_p = s.P = Q + r.R \quad (1)$$

### 6.2 Signing Phase

- The Proxy signer  $P_s$  chooses random integer  $t \in [1, n-1]$  and computes  $U = t.P$  and sends it to the verifier  $V$ .
- After receiving the verifier chooses randomly  $\alpha, \beta \in [1, n-1]$  and computes the following

$$\tilde{R} = U + \alpha.P - \beta.Q_p \quad (2)$$

$$\tilde{e} = H(\tilde{R} \| M) \quad (3)$$

$$e = (\tilde{e} + \beta) \bmod n \quad (4)$$

and verifier  $V$  sends  $e$  to the proxy signer  $P_s$ .

- After receiving  $e$ ,  $P_s$  computes the following

$$\tilde{s} = (t - s.e) \bmod n \quad (5)$$

and sends it to  $V$ .

- Now  $V$  computes

$$s_p = (\tilde{s} + \alpha) \bmod n \quad (6)$$

The tuples  $(M, s_p, \tilde{e})$  is the proxy blind signature.

### 6.3 Verification Phase

The verifier  $V$  computes the following equation.

$$\gamma = H((s_p.P + \tilde{e}.Q_p) \| M) \quad (7)$$

and verifies the validity of proxy blind signature  $(M, s_p, \tilde{e})$  with the equality  $\gamma = \tilde{e}$ .

## 7 Security Analyses

### 7.1 Security Notions

**Theorem 1** *It is infeasible for adversary  $A$  to derive signer's private key from all available public information.*

**Proof:** Assume that the adversary  $A$  wants to derive signer's private key  $d$  from his public key  $Q$ , he has to solve ECDLP problem which is computationally infeasible. Similarly, the adversary will encounter the same difficulty as she/he tries to obtain proxy signer's private key.

**Theorem 2** *Proxy signature is distinguishable from original signer's normal signature.*

**Proof:** Since proxy key is different from original signer's private key and proxy keys created by different proxy signers are different from each other, any proxy signature is distinguishable from original signer's normal signature and different proxy signer's signature are distinguishable.

**Theorem 3** *The scheme satisfies Unlinkability security requirement.*

**Proof:** In verification stage, the signer checks only whether  $\gamma = H((s_p.P + \tilde{e}.Q_p) \| M)$  holds.

He does not know the original signer's private key and proxy signer's private key. Thus the signer knows neither the message nor the signature associated with the signature scheme.

### 8. Correctness

**Theorem 4** *The proxy blind signature  $(M, s_p, \tilde{e})$  is universally verifiable by using the system Public parameters.*

**Proof:** The proof of correctness of the signature is verified as follows. We have to prove that



$H((s_p.P + \tilde{e}.Q_p) \parallel M) = H(\tilde{R} \parallel M)$  i.e. to show

$$\begin{aligned}
 s_p.P + \tilde{e}.Q_p &= \tilde{R} \\
 &= (\tilde{s} + \alpha).P + \tilde{e}.Q_p \\
 &= \tilde{s}.P + \alpha.P + \tilde{e}.Q_p \\
 &= (t - s.e).P + \alpha.P + \tilde{e}.Q_p \\
 &= t.P - (\tilde{e} + \beta).Q_p + \alpha.P + \tilde{e}.Q_p \\
 &= t.P - \beta.Q_p + \alpha.P \\
 &= U - \beta.Q_p + \alpha.P \\
 &= \tilde{R}
 \end{aligned}$$

## 9. Implémentation Issues

In this section we have discussed implementation issues, i.e. efficiency and size of the hard-ware. The basic operation for Cryptographic Protocols based on ECDLP; it is easily performed via repeated group operation. One can visualize these operations in a hierarchical structure. Point multiplication is at top level. At the next lower level is the point operations, which are closely related to coordinates used to represent the points. The lowest level consists of finite field operations such as addition, subtraction, multiplication and inversion.

### 9.1 Group Order

The order of the elliptic curve group over the underlying field is an important security parameter. There are attacks (for example Pohlig-Hellman attack) which can be launched on ECC if the group order is not divisible by a very large prime. In fact the Pohlig-Hellman attack dictates that the group order for ECC should be product of a large prime multiplied by a small positive integer less than 4. This small number is called *cofactor* of the curve. Various algorithms have been proposed in literature (for example Kedlaya's algorithm for ECC and Schoof's algorithm for ECC) for efficiently counting the group order. The group order of an elliptic curve is given by *Hasse's theorem*.

**Theorem 5.** Let  $E$  be an elliptic curve over a finite field  $F_p$  of order  $q$ . Then the order  $\#E(F_p)$  of the elliptic curve group is given by

$$\#E(F_p) = q + 1 - t, \text{ where } |t| \leq 2q^{1/2}$$

The parameter  $t$  is called trace of  $E$  over  $F_p$ . An interesting fact is that given any integer, there exists an elliptic curve  $E$  over  $F_p$  such that  $\#E(F_p) = q + 1 - t$ .

## 10. Point Representation and Cost of Group Operations

Point addition and point doubling are two important operations in ECC. Inversion in a finite field is an expensive operation. To avoid these inversions, several point representations have been proposed in literature. The cost of point addition and doubling varies depending upon the representation of the group elements. In the current section, we will briefly deal with some point representations commonly used. Let  $[i]$ ,  $[m]$ ,  $[s]$ ,  $[a]$  stand for cost of a field element inversion, a multiplication, a squaring and an addition respectively. Field element addition is considered to be a very cheap operation. In binary fields, squaring is also quite cheaper than a multiplication. If the underlying field is represented in normal basis then squaring is almost for free. Inversion is considered to be 8 to 10 times costlier than a multiplication in binary fields. In prime field the *I/M ratio* is even more. It is reported to be between 30 and 40.

### 10.1 Elliptic Curves

Point representation in ECC is a well studied area. In the following two sections we describe some of the point representation popularly used in implementations. Table 1. Cost of Group Operations in ECC for Various Point Representations for Characteristic  $> 3$

Coordinates	Cost (Addition)	Coordinates	Cost (Doubling)
$A + A \rightarrow A$	$1[i] + 2[m] + 1[s]$	$2A \rightarrow A$	$1[i] + 2[m] + 2[s]$
$P + P \rightarrow P$	$12[m] + 2[s]$	$2P \rightarrow P$	$7[m] + 3[s]$
$J + J \rightarrow J$	$12[m] + 4[s]$	$2J \rightarrow J$	$6[m] + 4[s]$
$C + C \rightarrow C$	$11[m] + 3[s]$	$2C \rightarrow C$	$5[m] + 4[s]$

Fields of Characteristic  $> 3$  Elliptic curves over fields of characteristic  $> 3$  have equations of the form  $y^2 = x^3 + ax + b$ . For such curves the following point representation methods are mostly used.

1. **In Standard Projective Coordinates** the curve has equation of the form

$$Y^2Z = X^3 + aXZ^2 + bZ^3$$

The point  $(X : Y : Z)$ , with  $Z \neq 0$  in projective coordinates is the point  $(X/Z, Y/Z)$  in affine



coordinates. The point at infinity is represented by the point  $(0: 1: 0)$  and the inverse of  $(X: Y: Z)$  is the point  $(X: -Y: Z)$ .

- In Jacobian Projective Coordinates** the curve has equation of the form  $Y^2Z = X^3 + aXZ^4 + bZ^6$ . The point,  $Z \neq 0$  in Jacobian coordinates correspond to the affine point  $(X/Z^2, Y/Z^3)$ . The point at infinity is represented by the point  $(1: 1: 0)$  and the inverse of  $(X: Y: Z)$  is the point  $(X: -Y: Z)$ . Point doubling becomes cheaper in Jacobian coordinates if the curve parameter  $a = -3$ .
- In Chudonovski Jacobian Coordinates**, the Jacobian point  $(X: Y: Z)$  is represented as  $(X: Y: Z: Z^2: Z^3)$ . Cost of point addition in Chudonovski Jacobian coordinates is the minimum among all representations.

In Table 1, we present the cost of addition and doubling in the coordinate systems described above. In the table we use  $A, P, J, C$  for affine, projective, Jacobian and Chudonovski Jacobian respectively. By  $2A \rightarrow A$ , we mean the doubling formula in which the input is in affine and so is the output. Similarly for addition and other coordinate systems.

**Fields of Characteristic 2** We will consider only non-super singular curves. Elliptic curves (non-super singular) over binary fields have equations of the form  $y^2 + xy = x^3 + ax^2 + b$ . For such curves the following point representation methods are mostly used.

- In Standard Projective Coordinates** the curve has equation of the form  $Y^2Z + XYZ = X^3 + aX^2Z + bZ^3$ . The point  $(X: Y: Z)$ , with  $Z \neq 0$  in projective coordinates is the point  $(X=Z, Y=Z)$  in affine coordinates. The point at infinity is represented by the point  $(0: 1: 0)$  and the inverse of  $(X: Y: Z)$  is the point  $(X: X + Y: Z)$ .
- In Jacobian Projective Coordinates** the curve has equation of the form  $Y^2 + XYZ = X^3 + aX^2Z^2 + bZ^6$

The point  $(X: Y: Z)$ , with  $Z \neq 0$  in Jacobian coordinates correspond to the affine point  $(X/Z^2, Y/Z^3)$ . The point at infinity is represented by the point  $(1: 1: 0)$  and the inverse of  $(X: Y: Z)$  is the point  $(X: X + Y: Z)$ .

- In Lopez-Dahab Coordinates**, the point  $(X: Y: Z)$ , with  $Z \neq 0$  represents the affine point  $(X/Z, Y/Z^2)$ . The equation of the elliptic curve in this representation is  $Y^2 + XYZ = X^3Z + aX^2Z^2 + bZ^4$ . The point at infinity is represented by the point  $(1: 0: 0)$  and the inverse of  $(X: Y: Z)$  is the point  $(X: X + Y: Z)$ .

In Table 2 we present the cost of addition and doubling in the coordinate systems over binary fields. In the table we use  $A, P, J, L$  for affine, projective, Jacobian and Lopez-Dahab respectively. The table follows the same notational convention as in last subsection. Note that in Table 2 we have neglected squaring also. That is because in binary fields squaring is a much cheaper operation than multiplication, if one point is in affine and the other is in projective or some other weighted co-ordinate, then point addition becomes relatively cheaper. This operation is called *addition in mixed coordinates or mixed addition*. In ECC, the base point is generally stored in affine coordinates to take advantage of mixed additions.

Table 2. Cost of Group Operations in ECC for Various Point Representations in Even Characteristics

Coordinates	Cost (Addition)	Coordinates	Cost (Doubling)
$A + A \rightarrow A$	$1[i] + 2[m]$	$2A \rightarrow A$	$1[i] + 2[m]$
$P + P \rightarrow P$	$13[m]$	$2P \rightarrow P$	$7[m] + 3[s]$
$J + J \rightarrow J$	$14[m]$	$2J \rightarrow J$	$5[m]$
$L + L \rightarrow L$	$14[m]$	$2L \rightarrow L$	$4[m]$

## 11. Scalar Multiplications

In ECC, computationally the most expensive operation is scalar multiplication. It is also very important from security point of view. The implementation attacks generally target the computation of this operation to break the cryptosystem. Given a point  $X$  and a positive integer  $m$ , computation of  $m \times X = X + \dots + X$  ( $m$  times) is called the operation of scalar multiplication. In this section we briefly outline various scalar multiplication algorithms proposed in literature. We do not include multi scalar

multiplication methods (i.e. methods to compute  $(lP + mQ)$ ). Also, due to the vastness of the subject and space constraints we will elaborate only those methods which are discussed in depth in this dissertation. The basic algorithms to compute the scalar multiplication are the age old binary algorithms. They are believed to have been known to the Egyptians two thousand years ago. The two versions of DBL-AND-ADD algorithm are defined above. These algorithms invoke two functions ADD and DBL. ADD takes as input two points  $X_1$  and  $X_2$  and return their sum  $X_1 + X_2$ , DBL takes as input one point  $X$  and computes its double  $2X$ .

---

Algorithm DBL-AND-ADD (Left-to-right binary method)

---

Input:  $X, m (m_{k-1} \dots m_1, m_0)$

Output:  $mX$ .

1.  $E = m_{k-1}X$
  2. for  $i = k-2$  down to 0
  3.      $E = DBL(E)$
  4.     if  $m_i = 1$
  5.          $E = ADD(E, X)$
  6. return  $E$
- 

Algorithm DBL-AND-ADD (Right-to-left binary method)

---

Input :  $X, m, (m_{k-1} \dots m_1, m_0)$

Output :  $mX$ .

1.  $E_0 = X, E_1 = 0$
2. for  $i = 0$  to  $k-1$
3.     if  $m_i = 1$
4.          $E_1 = ADD(E_0, E_1)$
5.      $E_0 = DBL(E_0)$
6. return  $(E_1)$

Both the algorithms first convert the scalar multiplier  $m$  into binary. Suppose  $m$  has a  $n$ -bit representation with hamming weight  $h$ . Then,  $mX$  can be computed by  $n-1$  invocations of DBL and  $h - 1$  invocations of ADD. Hence cost of the scalar multiplication is  $(n - 1) \times \text{cost}(DBL) + h \times \text{cost}(ADD)$ . As the average value of  $h$  is  $n=2$ , on the average these algorithms require  $(n - 1)$  doubling and  $n=2$  additions. As doublings are required more often than additions, attempts are made to reduce complexity of the doubling operation.

The scalar multiplication is the dominant operation in ECC. Extensive research has been carried out to compute it efficiently and a lot of results have been reported in literature. To compute the scalar multiplication efficiently there are three main approaches. As is seen in the basic binary algorithms the efficiency is intimately connected to the efficiency of ADD and DBL algorithms. So the first approach is to compute group operations efficiently. The second approach is to use a representation of the scalar such that the number of invocation of group operation is reduced. The third approach is to use more hardware support (like memory for pre-computation) to compute it efficiently. In some proposals these have approaches have been successfully combined to yield very efficient algorithms. As noted in the above, the cost of ADD and DBL depend to a large extent on the choice of underlying field and the point representation. Hence the cost of scalar multiplication also depends upon these choices. Based on the underlying field more efficient operations have been proposed. Over binary fields for ECC, using a point halving algorithm instead of DBL has been proved to be very efficient. Over fields of characteristic 3, point tripling has been more efficient. There are proposals for using fancier algorithms like the ones efficiently computing  $2P + Q, 3P + Q$  etc. instead of ADD and DBL.

## 12. Conclusions

The security of the scheme is hardness of solving ECDLP. The primary reason for the attractiveness of ECC over systems such as RSA and DSA is that the best algorithm known for solving the underlying mathematical problem namely, the ECDLP takes fully exponential time. In contrast, sub-exponential time algorithms are known for underlying mathematical problems on which RSA and DSA are based, namely the integer factorization (IFP) and the discrete logarithm (DLP) problems. This means that the algorithms for solving the ECDLP become infeasible much more rapidly as the problem size increases more than those algorithms for the IFP and DLP. For this reason, ECC offers security equivalent to RSA and DSA while using far smaller key sizes. The benefits of this higher-strength per-bit include higher speeds, lower power consumption, bandwidth savings, storage efficiencies, and smaller certificates. This can be implemented in low power and small processor mobile devices such as smart card, PDA etc. In this proposed scheme it is infeasible for adversary to derive signer's private key from all available public information. This protocol also achieves the security like requirements distinguishability, strong unforgeability, non-repudiation, and unlinkability.

## References

- [1]. M.Mambo, K.Usda and E.Okamoto Proxy signature: Delegation of power to sign messages "IEICE Transaction on Fundamentals", E79-A(1996), pp.1338-1353, 1996.
- [2]. D.Chaum Blind Signature for Untraceable Payments, In Crypto 82, New York, Plenum Press, pp.199-203, 1983
- [3]. S.J.Hwang and C.H.Shi A Simple multi-signature scheme, "Proceeding of 10th National conference on Information Security, Taiwan", 2000.
- [4]. N. Koblitz. A course in Number Theory and Cryptography, 2nd edition Springer-Verlag-1994
- [5]. K. H Rosen "Elementary Number Theory in Science and Communication", 2nd ed., Springer-Verlag, Berlin, 1986.
- [6]. A. Menezes, P. C Van Oorschot and S. A Vanstone Handbook of applied cryptography. CRC, Press, 1997.
- [7]. D. Hankerson, A .Menezes and S.Vanstone. Guide to Elliptic Curve Cryptography, Springer Verlag, 2004.
- [8]. "Certicom ECC Challenge and The Elliptic Curve Cryptosystem"available <http://www.certicom.com/index.php>.
- [9]. J.P.Kar Proxy Blind Multi-signature Scheme using ECC for handheld devices. Available at "International Association for Cryptology Research" <http://eprint.iacr.org/2011/043.pdf> .

**Daniyal M. Alghazzawi** has completed his Ph.D in Computer Science from University of Kansas in 2007, Master of Science in Teaching & Leadership in 2004 and Master of Science in Computer Science in 2003 from University of Kansas. He has worked as Web Programmer at ALTec (Advanced Learning Technologies) . Dr. Daniyal is currently Chairman of the Information Systems Department, Faculty of Computing and Information Technology, King Abdulaziz University. He has 04 journal papers and conferences to his credit. His research interest includes e-Security and Cryptography. Dr. Daniyal is a member of IEEE (Education Transaction) and ACM-SIGCSE (Special Interest Group in Computer Science Education) .

**Trigui Mohamed Salim** is currently working as a Lecturer at Information System Department, faculty of Computing and Information Technology, King Abdul Aziz University , KSA.. He has completed Master of Science (Information Technology) in 2009

from University, Utara Malaysia and Bachelor of Computer Science and Multimedia from University of Sfax, Tunisia in 2007. He has one conference paper to his credit. His research interest is e-Security and Cryptography.

**Syed Hamid Hasan** has completed his PhD in Computer Science from JMI, India, MSc in Statistics from AMU, India. Also he has completed Post-Graduate Diploma in Computer Science from the same university. Prof. Hamid has worked as a Head of Computer Science department at the AMU, India and was also Head of IT department at the Musana College of Technology, Sultanate of Oman. Dr Hamid is currently working as a Professor at Information Systems department, faculty of Computing and Information Technology, King Abdul Aziz University, Kingdom of Saudi Arabia. He was Reviewer for NDT 2009, Ostrava, Czech Republic, 2009. Co Sponsored by IEEE Communications Society. He is included in the Panel of referees of "The Indian journal of community health", was Chief Coordinator of the National Conference on "Vocationalization" of Computer Education" held on 28-29 September-1996 at A.M.U. Aligarh-India. He is a life Member of Indian Society for Industrial and Applicable Mathematics (ISIAM), Computer Society of India, Fellow National Association of Computer Educators & Trainers (FNACET), India. He has 20 research articles in conferences & journals to his credit. His research interest is e-Security and Cryptography.

# Web Based Application for Reading Comprehension Skills

Samir Zidat<sup>1</sup> and Mahieddine Djoudi<sup>2</sup>

<sup>1</sup> UHL Batna, Department of Computer Science, university Of Batna  
Batna, Algeria

<sup>2</sup> Laboratoire SIC et Equipe IRMA, University of Poitiers  
Poitiers, France

## Abstract

The use of the web in languages learning has been developed at very high speed these last years. Thus, we are witnessing many research and development projects set in universities and distance learning programs. However, the interest in research related to writing competence remains relatively low.

Our proposed research examines the use of the web for studying English as a second foreign language at an Algerian university. One focus is on pedagogy: therefore, a major part of our research is on developing, evaluating, and analyzing writing comprehension activities, and then composing activities into a curriculum.

The article starts with the presentation of language skills and reading comprehension. It then presents our approach of the use of the web for learning English as a second language. Finally a learner evaluation methodology is presented. The article ends with the conclusion and future trends.

**Keywords:** *Reading comprehension, E-learning, Assessment, Online Platform, Paper Submission.*

## 1. Introduction

This article describes a web based approach, where the web is used for educational activities. The main focus of this article is on reading comprehension of foreign language.

A new approach on the use of the web technology and how it was used in language learning, especially writing, is presented.

One of the main goals in our research work is to explore what are the best web learning practices and activities are in terms of assisting and supporting learning to become a more meaningful process. Another goal is to explore from a pedagogical perspective the innovative future learning practices, which are related to the new forms of studying.

## 2. Language Skills and Writing

In order to understand the problem being considered in this article, it is of primary importance to know what are the capacities concerned during a learning process of a foreign language. We point out that the capacities in learning a language represent the various mental operations that have to be done by a listener, a reader, or a writer in an unconscious way, for example: to locate, discriminate or process the data. One distinguishes in the analytical diagram, basic capacities which correspond to linguistic activities and competence in communication that involve more complex capacities.

### 2.1 Basic Language Skills

The use of a language is based on four skills. Two of these skills are from comprehension domain. These are oral and written comprehension. The last two concern the oral and written expression (see Table 1). A methodology can give the priority to one or two of these competences or it can aim at the teaching/learning of these four competences together or according to a given planned program.

On one hand, the written expression paradoxically is the component in which the learner is evaluated more often. It is concerned with the most demanding phase of the learning by requiring an in depth knowledge of different capacities (spelling, grammatical, graphic, etc.). On the other hand, listening comprehension corresponds to the most frequent used competence and can be summarized in the formula "to hear and deduce a meaning". Chronologically, it is always the one that is confronted first, except in exceptional situations (people only or initially confronted with the writing, defective in hearing, study of a dead language (a language that is not in use any more), study of a language on the basis of the autodidact writing).

	Oral	Written
Comprehension	Listening	Reading
Expression	Speaking	Writing

Table 1: Basic languages skills

## 2.2 Reading Comprehension

Reading comprehension can be defined as the level of understanding of a passage or text. It can be improved by: Training the ability to self assesses comprehension, actively test comprehension using questionnaires, and by improving met cognition. Teaching conceptual and linguistic knowledge is also advantageous. Self assessment can be conducted by summarizing, and elaborative interrogation, and those skills will gradually become more automatic through practice.

Reading comprehension skills separates the "passive" unskilled reader from the "active" readers. Skilled readers don't just read, they interact with the text. To help a beginning reader understand this concept, you might make them privy to the dialogue readers have with themselves while reading. Skilled readers, for instance:

- Predict what will happen next in a story using clues presented in text
- Create questions about the main idea, message, or plot of the text
- Monitor understanding of the sequence, context, or characters
- Clarify parts of the text which have confused them
- Connect the events in the text to prior knowledge or experience.

## 3. Related work

There has been much work on online reading focusing on new interaction techniques [1] and [2] to support practices observed in fluent readers [3] and [4], such as annotation, clipping, skimming, fast navigation, and obtaining overviews. Some work has studied the effect of presentation changes like hypertext appearance [5] on reading speed and comprehension.

The findings support what other studies have found in terms of positive influence of online environment on students' performances [6], [7], [8] and [9], but cannot be a substitution for them. The characteristics of online environment can increase students' motivation, create highly interactive learning environments, provide a variety

of learning activities, offer independence to users in the process of learning, improve learners' self confidence, and encourage learners to learn in a better way with technology-based tools.

Nowadays, most of the universities are asked to enhance of their personals' skills in order to utilize the new technologies in their teaching activities in an efficient way [10]. One of the modern technologies is online learning environment which is a software application to be used to integrate technological and pedagogical features into a well-developed virtual learning environment [11], [12] and [13]. Students have easy access to course materials, take online tests and collaborate with class mates and teacher.

## 4. Web-based application

### 4.1 Software architecture

The course management system is a web-based application with server-side processing of intensive requests. The environment provides the three principal users (teacher, learner, and administrator) a device, which has for primary functionality the availability and the remote access to pedagogical contents for language teaching, personalized learning and distance tutoring. The e-learning platform allows not only the downloading of the resources made available on line (using a standard navigator).

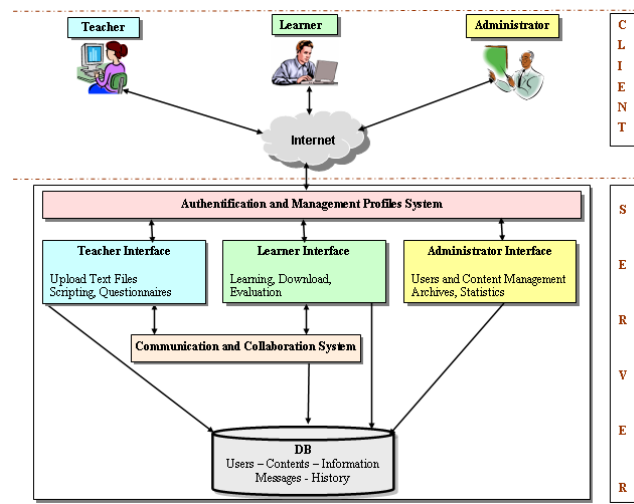


Fig. 1 Basic architecture of the environment



### 3.2 Teacher's interface

The Environment requires appropriate models for structuring and delivering content to be used. Different needs of learners require specific approaches. Then, we propose a model for structuring content that allows rendering for different users as well as presentation of the content in different levels of details according to didactic concepts like case study, definition, example, interaction, motivation, and directive. This approach allows adaptation of content (granularity of content, content selection based on didactic concepts) at run time to specific needs in a particular learning situation.

The environment allows the teacher, via a dedicated interface, to make at the learners' disposal a considerable large amount of textual documents, of excellent quality to read to. These documents are created by the teachers or recovered from Internet. The interface also makes it possible to the teacher to describe in the most complete possible way the files. Relative information to each file is: the name, the language, public concerned, expected pedagogic objectives, the period of accessibility, the source, copyright, etc. Thus documents prepared by the teacher are loaded in the database located on the platform server. If the learner can put his/her own techniques and strategies to understand the reading comprehension, then the instructor role consists in helping him/her to develop and enrich the learning strategies.



Fig. 2 Teacher interface

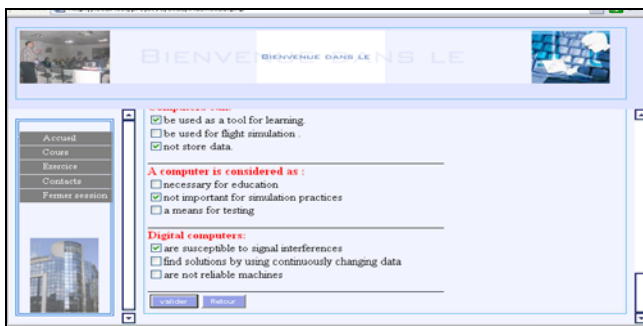


Fig. 3 Learner interface

### 3.3 Learner's interface

Learners with laptops can enter the learning space, see published courses and take part in them via their browser. Users can collaborate with other learners or teachers via discussion forums and Chat areas. Web content is dynamically adapted to the small screens.

- Collaboration Module locates people and provides for the transfer of documents and files between people logged into the environment.
- Course Access Interface provides for updating lessons, homework, and other assignments as well as the transfer of documents between learners and teachers.
- Communications tool launches a variety of communications options including text, audio and provides for 1-to-many, many-to-1, and 1-to-1 communication.
- Interactive Logbook a variety of service options including history access, and editing of user profiles

### 3.4 Interactive Logbook

The goal of the logbook is to set up an automatic book keeping information related to the learner's activity while he/she carries out a scenario on a teaching object (date and duration of each connection, exercises for self evaluation, etc). This requires an effort of information structuring and an implementation within the platform. An exploitation of this information by learners can guide them through their training plan.

The metaphor of the Interactive Logbook was conceived from the very traditional personal learning environment – a logbook, kept by learners to record lectures, laboratories, project notes and more besides. Although in many subjects, especially in the science disciplines, logbooks are still an integral part of courses, they are not well used by learners. Since many learning activities are taking place digitally: writing essays, analyzing data, browsing the web, online discussions - it doesn't make sense to print or copy out the output for a paper based logbook.

Although the interactive logbook extends to network capability and digital search, the name 'logbook' gives an impression of the flexibility and purpose of the tool. The Interactive Logbook seeks to provide a place in which personal information can be stored completely, privately and for all time.



By being integrated with the environment in which much of the learner's activity takes place, physical time and effort barriers can also be reduced, made even lower by automated logging of basic documents and events (emails, documents, diary entries, etc.) [14]. Finally, a statistical analysis of logbooks of a group of learners that have done the same activity would give a synthetic vision of the group's learning, and would be useful to all people involved in the learning.

### 3.5 Communication and collaboration using mobile devices

In the environment, learners have to find the same classical environment as they have in real life. In this environment learners can ask for all questions whenever they need and they discuss a lot together of interesting or pointless subjects.

The environment also support group communications by offering discussions, forums and shared workspaces where learners can exchange documents using podcasting tool. We distinguish between asynchronous and synchronous communication facilities. Social contacts are a crucial point in learning situations. Learners should therefore be able to present themselves in a personal homepage with a photograph, a list of hobbies and other personal aspects. Such personal presentations are not toys, but they can help the learners to get into contact even more easily than in live classroom situations. There is a great potential in using mobile terminals for communication services.

The communication and collaboration system launches a variety of communications options including text, audio, video, and whiteboard and provides for 1-to-many, many-to-1, and 1-to-1 communication. It provides a powerful architecture for the development of new educational tools to enhance different modes of teaching and learning. It is ideally suited to mobile learning and able to integrate tools developed explicitly for mobile contexts. The opportunity is to leverage the platform to develop innovative tools that are applicable to (1) synchronous formal learning (e.g., classrooms) and (2) asynchronous informal learning (e.g., discussion in the cafeteria).

There are a number of learning activities in formal educational environments (such as teacher-led classroom scenarios), which are ideally suited to mobile learning tools. Synchronous learning activities such as polling/voting and question and answer (where the system immediately collates all responses and presents an aggregate view of votes or answers to all learners) are ideal for pedagogically rich learning.

Features which are unique to the system and which would enhance the learning include:

- The ability to easily sequence activities into re-usable lesson plans (using a simple visual "drag and drop" lesson planner).
- Recording of learner responses for later review by learners/teachers and the option for teachers to create "question & answer" activities with either anonymous or identified answers from learners (which provides a basis for more honest answers due to the lack of peer pressure).

Informal learning scenarios (such as student discussion in a cafeteria) provide environments where mobile devices can support flexible, "on the fly" learning opportunities. Valuable learning activities in these contexts could be supported by a content sharing tool, and discussion forums and live chat/instant messaging for questions and responses to other learners or the teacher.

Again, the environment provides unique features to support these activities by providing an environment to manage and deliver these tools in the context of asynchronous (and synchronous) informal learning, including recording of activities for later learner/teacher review, and creation of re-usable lesson plans (based around informal student learning using flexible toolsets).

## 4. Experimentation

In Algeria, we evaluate the reading ability of student's university by giving them reading comprehension tests. These tests typically consist of a short text followed by questions. Presumably, the tests are designed so that the reader must understand important aspects of the text to answer the questions correctly. For this reason, we believe that reading comprehension tests can be a valuable tool to assess the state of the art in natural language understanding.

The main hypothesis of the present research study is as follow: the ongoing integration and utilization of the computer within the English language reading comprehension will firstly enhance the learners' affect exemplified by high motivations, and self-confidence. Consequently, when learners are motivated, they learn better and acquire more knowledge.

Secondly, empower the English teachers' roles and responsibilities as active agents of pedagogical and technological change.

The main objectives of the current work are to investigate, firstly, the validity of computer-assisted comprehension

reading and secondly, to attract both teachers and learners' attention as to the crucial relevance of the urgent integration the computer in English language learning in Algerian university. This study was conducted in intranet-based English language classroom with student of fifth year preparing the engineering degree of Computer Science Department in the Faculty of Engineering of the University of Batna, Algeria. Therefore, any obtained conclusions or results will apply of them.

There is a myriad of appropriate methodologies for the study of different learning problems. The selection of one and the avoidance of other is not a simple task at all. The nature and purpose of the investigation and the population involved will help the research to which method to be dealt with. In our present research work which investigates the possibility to adopt and adapt the computer in English language as instructional means and the way it can affect positively the learners, we found it more convenient to opt for the experimental research methods.

The Reading comprehension has come to be recognized as an active rather than a passive skill and its importance acknowledged in the acquisition of language. With the emergence of multimedia as teaching tools, it is being given renewed attention.

To verify if comprehension is reached, the learners are invited to answer to short instructions written in English language, without required that they write them in the sentences forms. The tasks of comprehension credited on the marks-scale, which appears on the specific grid, provide for each support and are distributed to the learners [15].

#### 4.1 Material

Text has been used in an exploratory study with similar students, the findings of which showed the texts as suitable in terms of content and level.

The text is general enough to be understood by students and do not require a deep specialist's knowledge of the topic discussed.

A set of multiple-choice comprehension questions was prepared for the text. All the questions were conceptual, focusing on main ideas and purpose of the writer, organization of the text. The multiple-choice format was chosen, despite much of the criticism on this method of testing, after a preliminary study with the open-ended format of questions which yielded too many divergent responses.

#### 4.2 Procedure

Every student participated in two sessions separated by 2 weeks. In the first session, the "Computer" text was used; in the second session, the "Network" text was used. In each session, the students of one group received the "computer" condition and other the "paper" condition. Thus, every student was exposed to the two contents (Computer and network), each content in one of two processing conditions ("paper" and "laptop"). In the first session, one group of the students used the paper sheets like support of work (reading and answering) for the "Computer text"; the other one used the laptops as work support for the "Network" text. In the second session, those students who had received the "laptop" condition in the first session received the "paper" condition for the "Computer" text, and those that had received the "paper" condition in the first session received the "laptop" condition for the "Network" text. The information concerning every session are summarized in the table 2.

The test condition involved the following instructions:

Read the following text. You have fifty minutes for this task. The conditions were explained to students who asked for clarification. The set of multiple-choice questions was distributed to the subjects with the text on their desks.

After 50 min, all the materials were collected. The students has a good or a very good knowledge of computing and didn't know at all or a few about the principle application (two persons out of five knew a little its principle of working).

The choice deliberated of this kind of people was conclusive because, contrary to beginners, they proved to be cooperative, and looked for testing the system, what helped us to identify the limits and weakness of this first version of the application.

In our project, we proceeded to the experimentation of the understanding of English language by using our developed system. In other words, we submit a text in English language to read, followed by exercises of Multiple Choice Question (MCQ) and True/False type on sheet of paper (classical method) for a group of users, and on a laptop for another group of users. The text to read and the exercises are elaborated by a specialist teacher at the department of English language of Batna University.

The set of the proposed exercises are marked on 20 points.

Our population is constituted of 20 students' 4th year computing engineer distributed in two groups:

- 10 students participate in this experimentation on sheets of paper.
- 10 students participate in this experimentation on laptop.

The interest of this experimentation is to answer the following question: Does the use of sheet of paper in written comprehension is more efficient than the use of the laptop (H0 hypothesis)? To answer this question (H0 hypothesis), Fisher statistical method is adopted.

Among the 10 students, 5 students work in an individual way, two groups (formed of 2 and 3 students) work together, i.e. they collaborate to read and to understand the text and solve the proposed exercises together. The same thing is made for the experimentation on sheets of paper, but in that case, the students find the text and the exercises on a laptop and are marked in an automatic way.

Every student participated in two sessions separated by 2 weeks. In the first session, the "Computer" text was used; in the second session, the "Network" text was used. In each session, the students of one group received the "computer" condition and other the "paper" condition. Thus, every student was exposed to the two contents (Computer and network), each content in one of two processing conditions ("paper" and "laptop"). In the first session, one group of the students used the paper sheets like support of work (reading and answering) for the "Computer text"; the other one used the laptops as work support for the "Network" text.

In the second session, those students who had received the "laptop" condition in the first session received the "paper" condition for the "Computer" text, and those that had received the "paper" condition in the first session received the "laptop" condition for the "Network" text. The information concerning every session are summarized in the table 2.

The test condition involved the following instructions: Read the following text. You have fifty minutes for this task. The conditions were explained to students who asked for clarification. The set of multiple-choice questions was distributed to the subjects with the text on their desks. After 50 min, all the materials were collected. The students has a good or a very good knowledge of computing and didn't know at all or a few about the principle application (two persons out of five knew a little its principle of working).

The choice deliberated of this kind of people was conclusive because, contrary to beginners, they proved to be cooperative, and looked for testing the system, what

helped us to identify the limits and weakness of this first version of the application.

In our project, we proceeded to the experimentation of the understanding of English language by using our developed system. In other words, we submit a text in English language to read, followed by exercises of MCQ and True/False type on sheet of paper (classical method) for a group of users, and on a laptop for another group of users. The text to read and the exercises are elaborated by a specialist teacher at the department of English language of Batna University. The set of the proposed exercises are marked on 20 points.

Our population is constituted of 20 students' 4th year computing engineer distributed in two groups:

10 students participate in this experimentation on sheets of paper.

10 students participate in this experimentation on laptop.

The interest of this experimentation is to answer the following question: Does the use of sheet of paper in written comprehension is more efficient than the use of the laptop (H0 hypothesis)? To answer this question (H0 hypothesis), Fisher method is adopted [16].

Among the 10 students, 5 students work in an individual way, two groups (formed of 2 and 3 students) work together, i.e. they collaborate to read and to understand the text and solve the proposed exercises together. The same thing is made for the experimentation on sheets of paper, but in that case, the students find the text and the exercises on a laptop and are marked in an automatic way.

Table 2: The groups of work

Group 1	5 students working separately on laptop
Group 2	5 students working separately on sheet of paper
Group 1.s	2 groups of students (2 or 3) working in collaboration on laptop
Group 2.s	2 groups of students working in collaboration on sheet of paper

### 4.3 Statistic study

Our main objective is to try to answer the following question: "Is the traditional use of paper sheets as work support in the reading comprehension more effective than the use of the laptop concerning this population (assumption H0) ? "

By applying the Fisher method, one calculates the sum of square method, SS meth, and the sum of the residual,

SSres to reach the factor Fisher F. The results are available in figure 3 with: Degree of freedom = 3 and the critical point of Fisher  $F_{3,16}(0.05) = 3.23$ :

From the obtained results we have  $F > F_{3,16}(0,05)$ , therefore, one rejects  $H_0$  i.e. the use of laptop is more effective than the use of traditional paper sheets.

One can note starting from the Fisher's result that the use of microcomputer by our learners helped us obtain a higher performance than working on the traditional paper sheet and we noted the collaborative learning with a help of a micro portable provided us with the better performances.

Variance Analysis: (Fisher) factor						
Groups	Number of samples	Sum	Average values	Variance		
Paper=individually	5	58	11.6	4.3		
Paper+collaboratively	5	58	11.6	0.3		
Laptop+individually	5	72	14.4	2.3		
Laptop+collaboratively	5	77	15.4	0.3		
Variance Analysis						
Variation sources	Sum of squares	Degree of freedom	Mean of squares	F	Probability	Critical value of F
Between Groups (SSmeth)	56.95	3	18.98333	10.5462	0.00045258	3.2388
Among groups (SSres)	28.8	16	1.8			
Total	85.75	19				

Fig. 4 Fisher results

By applying the Fisher method, one calculates the sum of square method, SS meth, and the sum of the residual, SSres to reach the factor Fisher F. The results are available in table 3 with: Degree of freedom = 3 and the critical point of Fisher  $F_{3,16}(0.05) = 3.23$ :

From the obtained results we have  $F > F_{3,16}(0,05)$ , therefore, one rejects  $H_0$  i.e. the use of laptop is more effective than the use of traditional paper sheets.

One can note starting from the Fisher's result that the use of microcomputer by our learners helped us obtain a higher performance than working on the traditional paper sheet and we noted the collaborative learning with a help of a micro portable provided us with the better performances.

#### 4.3 Limitations

The present study deals with four year learners' poor reading performances at the department of Computer Science at Batna University. Any conclusion drawn from

the experiment will be limited to the targeted population only.

### 5. Future Trends

We have started experimenting with the use of the environment in real teaching/learning situation. This experimentation allows us to collect information on the effective activities of the users. We can thus validate or question certain technical choices and determine with more precision the adaptations that have to be made to the integrated tools Feedback from a panel was very positive and the mobile aspect of environment was seen as a novel and interesting approach as a research tool. A detailed evaluation of the effectiveness of the learning environment has yet to be completed. In prospect, the approach aims at developing in the learners other language skills, so that they can express themselves in foreign language.

### 6. Conclusion

We presented in this paper an original approach for reading comprehension of English second foreign language by using web-based application. According to the study of the experimentation result, we can conclude that learning by computer doesn't stop evolving, and the learner finds a simple method of education.

The obtained results supported our hypothesis that claims that the use of web based application can contribute in improving the students' reading comprehension. Henceforth, we recommend the generalization of this new technology in our schools and universities to allow students take a maximum advantage of it.

### References

- [1] J., Graham, "The reader's helper: a personalized document reading environment", Proceedings of CHI '99, 1999, pp. 481-488.
- [2] B. N. Schilit, G. Golovchinsky and M. N. Price, "Beyond paper: supporting active reading with free form digital ink annotations", Proceedings of CHI '98, 1998, pp. 249-256.
- [3] G. B.Duggan, S. J., "Payne: How much do we understand when skim reading?," Proceedings of CHI '06, 2006, pp. 730-735.
- [4] K. O'Hara, A. Sellen, "A comparison of reading paper and on-line documents". Proceedings of CHI '97, 1997, pp. 335-342.
- [5] D. Cook, "A new kind of reading and writing space the online course site", The Reading Matrix, Vol.2, No.3, September, 2002.
- [6] V. Fernandez, P. Simoa, J. Sallana : "Podcasting: A new technological tool to facilitate good practice in higher education", Computers & Education, Volume 53, Issue 2, September,2009, pp. 385-392.

- [7] W. Tsou, W. Wang and H. Li, "How computers facilitate English foreign language learners acquire English abstract words", *Computers & Education*, 2002, pp. 415–428.
- [8] Y.L. Chen, "A mixed-method study of EFL teachers' Internet use in language instruction", *Teaching and Teacher Education*, 2008, pp. 1015–1028.
- [9] M. Rahimi, S. Yadollahia, "Foreign language learning attitude as a predictor of attitudes towards computer-assisted language learning", *Procedia Computer Science Volume 3, World Conference on Information Technology*, 2011 Pages 167-174.
- [10] Turan, "Student Readiness for Technology Enhanced History Education in Turkish High Schools", *Cypriot Journal Of Educational Sciences*, 5(2). Retrieved, from <http://www.worldeducationcenter.org/index.php/cjes/article/view/75>, 2010.
- [11] S. Zidat, S. Tahy and M. Djoudi, "Système de compréhension à distance du français écrit pour un public arabophone", *Colloque Euro Méditerranéen et Africain d'Approfondissement sur la FORMation A Distance, CEMAFORAD 4*, 9, 10 et 11 avril, Strasbourg, France, 2008.
- [12] S. Zidat, M. Djoudi, "Online evaluation of Ibn Sina elearning environment", *Information Technology Journal (ITJ)*, ISSN: 1812-5638, Vol. 5, No. 3, 2006, pp. 409-415.
- [13] S. Zidat, M. Djoudi, "Task collaborative resolution tool for elearning environment", *Journal of Computer Science*, ISSN: 1549-3636, Vol. 2, No. 7, pp. 558-564.
- [14] O. Kiddie, T. Marianczak, N. Sandle, L. Bridgefoot, C. Mistry, D. Williams, D. Corlett, M. Sharples. and S. Bull, "Logbook: The Development of an Application to Enhance and Facilitate Collaborative Working within Groups in Higher Education". *Proceedings of MLEARN 2004: Learning Anytime, Everywhere*, Rome, 5-6 July.
- [15] J.-F. Rouet, A. Goumi, A. Maniez and A. Raud. "Liralec : A Web-based resource for the assessment and training of reading-comprehension skills", In C.P. Constantinou, D. Demetriou, A. Evagorou, M. Evagourou, A. Kofteros, M. Michael, Chr. Nicolaou, D. Papademetriou & N. papadouris (Eds.), *Multiple Perspectives on Effective Learning Environments*, 2005, (pp. 113).

**Samir Zidat** got a Ph. D. in Computer Science from Université Hadj Lakhdar of Batna (Algeria) in 2006. He works as associate professor in Batna University.

His main scientific interests are e-learning and mobile learning, with a special emphasis on language learning. His PhD thesis research was in Assessment of e-learning platforms. His teaching interests include Software Engineering, Information & Communication Technology and Artificial Intelligence.

Dr. Samir Zidat  
Université de Batna  
05 avenue Chahid Boukhrouf, 05000 Batna, Algérie.  
Phone: 0213 5 57 83 25 19  
Fax: 0213 33 86 940

**Mahieddine Djoudi** received a PhD in Computer Science from the University of Nancy, France, in 1991. He is currently an Associate

Professor at the University of Poitiers, France. He is a member of SIC (Signal, Images and Communications) Research laboratory. He is also a member of IRMA E-learning research group. His PhD thesis research was in Continuous Speech Recognition. His current research interest is in E-Learning, Mobile Learning, Computer Supported Cooperative Work and Information Literacy. His teaching interests include Programming, Data Bases, Artificial Intelligence and Information & Communication Technology. He started and is involved in many research projects which include many researchers from different Algerian universities.

Dr. Mahieddine Djoudi  
XLIM-SIC Lab. & IRMA Research Group, University of Poitiers  
Bât. Sp2mi, Boulevard Marie et Pierre Curie, BP 30179,  
86962 Futuroscope Cedex France  
Phone: +33 5 49 45 39 89 Fax: +33 5 49 45 38 16

URL: <http://mahieddine.djoudi.online.fr>



# Active Fault Tolerant Control (FTC) Design for Takagi-Sugeno Fuzzy Systems with Weighting Functions Depending on the FTC

Atef Khedher, Kamel Ben Othman, Mohamed Benrejeb

Ecole Nationale d'Ingénieurs de Tunis,  
UR. LARA Automatique, BP 37, le Belvédère, 1002 Tunis

## Abstract

In this paper the problem of active fault tolerant control design for noisy systems described by Takagi-Sugeno fuzzy models is studied. The proposed control strategy is based on the known of the fault estimated and the error between the faulty system state and a reference system state. The considered systems are affected by actuator and sensor faults and have the weighting functions depending on the fault tolerant control. A mathematical transformation is used to conceive an augmented system in which all the faults affecting the initial system appear as actuator faults. Then, an adaptive proportional integral observer is used in order to estimate the state and the faults. The problem of conception of the proportional integral observer and of the fault tolerant control strategy is formulated in linear matrices inequalities which can be solved easily. To illustrate the proposed method, it is applied to the three tanks systems.

**Keywords:** *fault estimation, active fault tolerant control, proportional integral observer, nonlinear Takagi-Sugeno fuzzy models, actuator faults, sensor faults.*

## 1. Introduction

State observers are always used to estimate system outputs by the known of the system model and some measures of the system control and output [27]. This estimation is compared to the measured value of the output to generate residuals. The residuals are used as reliable indicators of the process behavior. They are equal to zero if the system is not affected by faults. The residuals depend of faults if they are present. There are three categories of faults detection methods: sensors faults detection, actuators faults detection and system faults detection.

In most cases, processes are subjected to disturbances which have as origin the noises due to its environment and the model uncertainties. Moreover, sensors and/or actuators can be corrupted by different faults or failures. Many works are dealing with state estimation for systems with unknown inputs or parameter uncertainties. In [37], Wang *et al.* propose an observer able to entirely reconstruct the state of a linear system in the presence of unknown inputs and in [28], to estimate the state, a model inversion method is used. Using the Walcott and Zak

structure observer [36], Edwards *et al.* [7] and [8] have also designed a convergent observer using the Lyapunov approach.

In the context of nonlinear systems described by Takagi-Sugeno fuzzy models, some works tried to reconstruct the system state in spite of the unknown input existence. This reconstruction is assured via the elimination of unknown inputs [10]. Other works choose to estimate the unknown inputs and system state simultaneously [1], [12], [18], [23] and [30]. Unknown input observers can be used to estimate actuator faults provided they are assumed to be considered as unknown inputs. This estimation can be obtained by using a proportional integral observer [15], [17], [21-23]. That kind of observers gives some robustness property of the state estimation with respect to the system uncertainties and perturbations [4], [31].

Faults affecting systems have harmful effects on the normal behavior of the process and their estimation can be used to conceive a control strategy able to minimize their effects (named fault tolerant control (FTC)). A control loop can be considered fault tolerant if there exist adaptation strategies of the control law included in the closed-loop that introduce redundancy in actuators [38]. Fault Tolerant Control (FTC) is, relatively, a new idea in the research literature [5] which allows having a control loop that fulfils its objectives when faults appear [11], [16], [19] and [20]

There are two main groups of control strategies: the active and the passive techniques. The passive techniques are control laws that take into account the faults appearance as system perturbations [38]. Thus, within certain margins, the control law has inherent fault tolerant capabilities, allowing the system to cope with the fault presence [38]. This kind of control is described in [5], [6] [25], [26], [32] and [33]. The active fault tolerant control techniques consist on adapting the control law using the information given by the FDI block [5], [16], [19], [20] and [40]. With this information, some automatic adjustments are done trying to reach the control objectives [38].



In this paper, an active FTC strategy inspired from that given in [38] is proposed. In [38] Witczak *et al.* designed a FTC strategy for the class of discrete systems. This FTC is conceived using the error between the faulty and the reference system states. However, in real cases the faulty system state is unknown. The main contribution in this work is to conceive the FTC for the case of non linear systems described by Takagi-Sugeno fuzzy models with weighting functions depending on the fault tolerant control. This case is not treated enough in the literature [11]. It is important to consider this system class because if the weighting functions are depending on the system input and if the system input changes because of the action of the fault affecting the system, the weighting functions must depend on the new system input. State and faults estimation is made using an adaptive proportional integral observer. A mathematical transformation is used to conceive an augmented system in which the sensor fault affecting the initial system appears as an actuator fault. The actuator fault is considered as an unknown input. Once the fault is estimated, the FTC controller is implemented as a state feedback controller. In this work the observer design and the control implementation can be made simultaneously.

The paper is organized as follows. Section 2 recalls an elementary background about the Takagi-Sugeno fuzzy models (named also multiple models). In section 3 the proposed method of fault tolerant control design is presented. The application of the proposed control to the three tanks system is the subject of section 4.

## 2. On the Takagi-Sugeno fuzzy systems

Takagi-Sugeno fuzzy models are non linear systems described by a set of if-then rules which gives local linear representations of an underlying system [1], [12], [14] and [39] Such models can approximate a wide class of non linear systems [39]. They can even describe exactly some non linear systems [38] and [39].

Each non linear dynamic system can be simply, described by a Takagi-Sugeno fuzzy model [35] and [34]. A Takagi-Sugeno fuzzy model is the fuzzy fusion of many linear models [1-3], [12] and [30] each of them represents the local system behavior around an operating point. A Takagi-Sugeno model is described by fuzzy IF-THEN rules which represent local linear input/output relations of the non-linear system [38]. It has a rule base of  $M$  rules, each having  $p$  antecedents, where the  $i^{th}$  rule is expressed as:

$$R^i : \text{IF } \xi_1 \text{ is } F_1^i \text{ and ... and } \xi_p \text{ is } F_p^i$$

$$\text{THEN} : \begin{cases} \dot{x}(t) = A_i x(t) + B_i u(t) \\ y(t) = C_i x(t) \end{cases} \quad (1)$$

in which  $i=1 \dots M$ ,  $F_j^i (j=1 \dots p)$  are fuzzy sets and  $\xi = [\xi_1 \ \xi_2 \ \dots \ \xi_p]$  is a known vector of premise variables [23] which may depend on the state, the input or the output.

The final output of the normalized Takagi-Sugeno fuzzy model can be inferred as:

$$\begin{cases} \dot{x}(t) = \sum_{i=1}^M \mu_i(\xi(t))(A_i x(t) + B_i u(t)) \\ y(t) = \sum_{i=1}^M \mu_i(\xi(t))C_i x(t) \end{cases} \quad (2)$$

The weighting functions  $\mu_i(\xi(t))$  are non linear and depend on the decision variable  $\xi(t)$ .

The weighting functions are normalized rules defined as:

$$\mu_i(\xi(t)) = \frac{T_{j=1}^p \omega_i(\xi(t))}{\sum_{j=1}^M T_{j=1}^p \omega_j(\xi(t))} \quad (3)$$

where  $\omega_i(\xi(t))$  is the grade of membership of the premise variable  $\xi(t)$  and  $T$  denotes a t-norm. The weighting functions satisfy the sum convex property expressed in the following equations:

$$0 \leq \mu_i(\xi(t)) \leq 1 \quad \text{and} \quad \sum_{i=1}^M \mu_i(\xi(t)) = 1 \quad (4)$$

If, in the equation which defines the output, we impose that  $C_1 = C_2 = \dots = C_M = C$ , the output of the model (2) is reduced to:  $y(t) = Cx(t)$  and the Takagi-Sugeno fuzzy model becomes:

$$\begin{cases} \dot{x}(t) = \sum_{i=1}^M \mu_i(\xi(t))(A_i x(t) + B_i u(t)) \\ y(t) = Cx(t) \end{cases} \quad (5)$$

This model, known also as Takagi-Sugeno multiple model, has been initially proposed, in a fuzzy modeling framework, by Takagi and Sugeno [34] and in a multiple model modeling framework in [13] and [29]. This model has been largely considered for analysis [29], [34] and [9], modeling [13] and [41], control [21] and [9] and state estimation [1-3], [12], [22], [23] and [30] of non linear systems.

### 3. Active fault tolerant control design

A non linear system described by multiple model can be expressed as follow:

$$\begin{cases} \dot{x}(t) = \sum_{i=1}^M \mu_i(u(t))A_i x(t) + Bu(t) \\ y(t) = Cx(t) \end{cases} \quad (6)$$

where  $x(t) \in R^n$  is the state vector,  $u(t) \in R^r$  is the input vector,  $y(t) \in R^m$  the output vector and  $A_i, B$  and  $C$  are known constant matrices with appropriate dimensions. The scalar  $M$  represents the number of local models. Consider the following nonlinear Takagi-Sugeno model affected by actuator and sensor faults and measurement noise:

$$\begin{cases} \dot{x}_f(t) = \sum_{i=1}^M \mu_i(u_f(t))A_i x_f(t) + Bu_f(t) + Ef_a(t) \\ y_f(t) = Cx_f(t) + Ff_s(t) + Dw(t) \end{cases} \quad (7)$$

where  $x_f(t) \in R^n$  is the state vector,  $u_f(t) \in R^r$  is the fault tolerant control which will be conceived,  $y_f(t) \in R^m$  is the output vector.  $f_a(t)$  and  $f_s(t)$  are respectively the actuator and sensor faults which are assumed to be bounded and  $w(t)$  represents the measurement noise.  $E, F$  and  $D$  are respectively the faults and the noise distribution matrices which are assumed to be known. Let us define the following states [15]:

$$\dot{z}(t) = \sum_{i=1}^M \mu_i(u(t))(-\bar{A}z(t) + \bar{A}Cx(t)) \quad (8)$$

$$\dot{z}_f(t) = \sum_{i=1}^M \mu_i(u_f(t))(-\bar{A}z(t) + \bar{A}Cx(t) + \bar{A}_i Ff_s(t) + \bar{A}Dw(t))$$

where  $-\bar{A}$  is a stable matrix with appropriate dimension. Defining the two augmented states  $X(t)$  and  $X_f(t)$  as:

$$X(t) = \begin{bmatrix} x(t)^T & z(t)^T \end{bmatrix}^T \quad \text{and} \quad X_f(t) = \begin{bmatrix} x_f(t)^T & z_f(t)^T \end{bmatrix}^T$$

these two augmented state vectors can be written:

$$\begin{cases} \dot{X}(t) = \sum_{i=1}^M \mu_i(u(t))A_{ai} X(t) + B_a u(t) \\ Y(t) = C_a X(t) \end{cases} \quad (9)$$

and

$$\begin{cases} \dot{X}_f(t) = \sum_{i=1}^M \mu_i(u_f(t))A_{ai} X_f(t) + B_a u_f(t) + E_a f(t) + D_a w(t) \\ Y_f(t) = C_a X_f(t) \end{cases} \quad (10)$$

with:

$$A_{ai} = \begin{bmatrix} A_i & 0 \\ -\bar{A}C & -\bar{A} \end{bmatrix}, \quad E_a = \begin{bmatrix} E & 0 \\ 0 & \bar{A}F \end{bmatrix}, \quad f = \begin{bmatrix} f_a \\ f_s \end{bmatrix}, \quad B_a = \begin{bmatrix} B \\ 0 \end{bmatrix},$$

$$D_a = \begin{bmatrix} 0 \\ \bar{A}D \end{bmatrix} \quad \text{and} \quad C_a = [0 \quad I], \quad (11)$$

A proportional integral observer is used to estimate the augmented state  $X_f(t)$  and the generalized fault  $f(t)$ . It is given by the following equations:

$$\begin{cases} \dot{\hat{X}}_f(t) = \sum_{i=1}^M \mu_i(u_f(t))(A_{ai} \hat{X}_f(t) + K_i \tilde{Y}_f(t)) + B_a u_f(t) + E_a \hat{f}(t) \\ \hat{f}(t) = \sum_{i=1}^M \mu_i(u_f(t))L_i \tilde{Y}_f(t) \\ \hat{Y}(t) = C_a \hat{X}(t) \end{cases} \quad (12)$$

where  $\hat{X}_f(t)$  is the estimated system state,  $\hat{f}(t)$  represents the estimated fault,  $\hat{Y}_f(t)$  is the estimated output,  $K_i$  are the proportional gains of the local observers and  $L_i$  are their integral gains to be computed and  $\tilde{Y}_f(t) = Y_f(t) - \hat{Y}_f(t)$ .

The fault tolerant control  $u_f(t)$  is conceived on the base of the strategy described by the following expression [38].

$$u_f(t) = -S\hat{f}(t) + G(X(t) - \hat{X}_f(t)) + u(t) \quad (13)$$

where  $S$  and  $G$  are two constant matrices with appropriate dimensions.

Let us define  $\tilde{X}(t)$  the error between the states  $X(t)$  and  $X_f(t)$ ,  $\tilde{X}_f(t)$  the estimation error of the state  $X_f(t)$  and  $\tilde{f}(t)$  the fault estimation error :

$$\tilde{X}(t) = X(t) - X_f(t)$$

$$\tilde{X}_f(t) = X_f(t) - \hat{X}_f(t) \quad (14)$$

$$\tilde{f}(t) = f(t) - \hat{f}(t)$$

Choosing the matrix  $S$  verifying  $E_a = B_a S$ , the dynamics of  $\tilde{X}(t)$  is given by:

$$\begin{aligned} \dot{\tilde{X}}(t) &= \dot{X}(t) - \dot{X}_f(t) \\ &= \sum_{i=1}^M \mu_i(u(t))(A_{ai} - B_a G)\tilde{X}(t) - E_a \tilde{f}(t) - B_a G\tilde{X}_f(t) + \Delta_1(t) \end{aligned} \quad (15)$$

with :

$$\Delta_1(t) = \sum_{i=1}^M \mu_i(u_f(t) - \mu_i u(t))A_{ai} \tilde{X}_f(t) - D_a w(t) \quad (16)$$

The dynamic of  $\tilde{X}_f(t)$  can be written:

$$\begin{aligned} \dot{\tilde{X}}_f(t) &= \dot{X}_f(t) - \dot{X}_f(t) \\ &= \sum_{i=1}^M \mu_i(u(t))(A_{ai} - K_i C_a) \tilde{X}_f(t) + E_a \tilde{f}(t) + \Delta_2(t) \end{aligned} \quad (17)$$

with :

$$\Delta_2(t) = \sum_{i=1}^M \mu_i(u_f(t) - \mu_i u(t))(A_{ai} - K_i C_a) \tilde{X}_f(t) + D_a w(t) \quad (18)$$

The dynamic of the fault error estimation is:

$$\begin{aligned} \dot{\tilde{f}}(t) &= \dot{f}(t) - \dot{\hat{f}}(t) \\ &= -\sum_{i=1}^M \mu_i(u(t)) L_i C_a \tilde{X}_f(t) + \Delta_3(t) \end{aligned} \quad (19)$$

with :

$$\Delta_3(t) = \sum_{i=1}^M \mu_i(u_f(t) - \mu_i u(t)) L_i C_a \tilde{X}_f(t) + D_a w(t) + \dot{f}(t) \quad (20)$$

The equations (15), (17) and (19) can be rewritten:

$$\dot{\varphi}(t) = A_m \varphi(t) + \varepsilon(t) \quad (21)$$

where :

$$\varphi(t) = \begin{bmatrix} \tilde{X}(t) \\ \tilde{X}_f(t) \\ \tilde{f}(t) \end{bmatrix}, \quad \varepsilon(t) = \begin{bmatrix} \Delta_1(t) \\ \Delta_2(t) \\ \Delta_3(t) \end{bmatrix} \text{ and } A_m = -\sum_{i=1}^M \mu_i(u(t)) A_{mi} \quad (22)$$

where

$$A_{mi} = \begin{bmatrix} A_{ai} - B_a G & -B_a G & B_a \\ 0 & A_{ai} - K_i C_a & B_a \\ 0 & L_i C_a & 0 \end{bmatrix} \quad (23)$$

Considering the Lyapunov function  $V(t) = \varphi(t)^T P \varphi(t)$ , the generalized error vector  $\varphi(t)$  converges to zero if  $\dot{V}(t) < 0$ ,

$$\dot{V}(t) < 0 \text{ if } A_{mi}^T P + P A_{mi} < 0 \quad \forall i \in \{1 \dots M\} .$$

The problem of robust state and faults estimation and of the fault tolerant control design is reduced to find the gains  $K$  and  $L$  of the observer and the matrix  $G$  to ensure an asymptotic convergence of the generalized error vector  $\varphi(t)$  toward zero if  $\varepsilon(t) = 0$  and to ensure a bounded error in the case where  $\varepsilon(t) \neq 0$ , i.e.:

$$\begin{aligned} \lim_{t \rightarrow \infty} \varphi(t) &= 0 & \text{for } \varepsilon(t) &= 0 \\ \|\varphi(t)\|_{Q_\varphi} &\leq \lambda \|\varepsilon(t)\|_{Q_\varepsilon} & \text{for } \varepsilon(t) &\neq 0 \end{aligned} \quad (24)$$

where  $\lambda > 0$  is the attenuation level. To satisfy the constraints (13), it is sufficient to find a Lyapunov function  $V(t)$  such that:

$$\dot{V}(t) + \varphi(t)^T Q_\varphi \varphi(t) - \lambda^2 \varepsilon(t)^T Q_\varepsilon \varepsilon(t) < 0 \quad (25)$$

where  $Q_\varphi$  and  $Q_\varepsilon$  are two positive definite matrices.

The inequality (25) can be written:

$$\begin{bmatrix} \varphi(t) \\ \varepsilon(t) \end{bmatrix}^T \Phi \begin{bmatrix} \varphi(t) \\ \varepsilon(t) \end{bmatrix} < 0 \quad (26)$$

where:

$$\Phi = \begin{bmatrix} A_m^T P + P A_m + Q_\varphi & P \\ P & -\lambda^2 Q_\varepsilon \end{bmatrix} \quad (27)$$

Choosing  $Q_\varphi = Q_\varepsilon = I$  and assume that the Lyapunov matrix  $P$  has the form:  $diag(I, P_2, P_3)$ , the matrix  $\Phi$  is written :

$$\Phi = \sum_{i=1}^M \mu_i(u(t)) \Phi_i \quad (28)$$

where:

$$\Phi_i = \begin{bmatrix} \Phi_{11i} & -B_a G & B_a & I & 0 & 0 \\ -G^T B_a^T & \Phi_{22i} & \Phi_{23i} & 0 & P_2 & 0 \\ B_a^T & \Phi_{32i} & I_3 & 0 & 0 & P_3 \\ I & 0 & 0 & \lambda_1 I_{01} & 0 & 0 \\ 0 & P_2 & 0 & 0 & \lambda_2 I_{02} & 0 \\ 0 & 0 & P_3 & 0 & 0 & \lambda_3 I_{03} \end{bmatrix} \quad (29)$$

with:

$$\begin{aligned} \Phi_{11i} &= A_{ai} - B_a G + A_{ai}^T - G^T B_a^T + I_1 \\ \Phi_{22i} &= P_2 A_{ai} - P_2 K_i C_a + A_{ai}^T P_2 - C_a^T K_i^T P_2 + I_2 \\ \Phi_{23i} &= P_2 B_a + C_a^T L_i^T P_3 \end{aligned} \quad (30)$$

$$\Phi_{32i} = \Phi_{23}^T$$

$\Phi < 0$  if  $\Phi_i < 0 \quad \forall i \in \{1 \dots M\}$ , the inequalities  $\Phi_i < 0$  are bilinear, they can be linearised using the changes of variables :  $U_{2i} = P_2 K_i$  and  $U_{3i} = P_3 L_i$ . The observer gains are then computed using the equations:

$$K_i = P_2^{-1} U_{2i} \quad (31)$$

$$L_i = P_3^{-1} U_{3i}$$

Summarizing the following theorem can be proposed:

**Theorem:**

The system (21) describing the evolution of the errors  $\tilde{X}(t)$ ,  $\tilde{X}_f(t)$  and  $\tilde{f}(t)$  is stable if there exist symmetric definite positive matrices  $P_2$  and  $P_3$  and matrices  $U_{3i}$ ,  $U_{2i}$  and  $G$ ,  $i \in \{1 \dots M\}$  so that the LMI  $\Phi_i < 0$  are verified  $\forall i \in \{1 \dots M\}$  where :

$$\Phi_i = \begin{bmatrix} \Phi_{11i} & -B_a G & B_a & I & 0 & 0 \\ -G^T B_a^T & \Phi_{22i} & \Phi_{23i} & 0 & P_2 & 0 \\ B_a^T & \Phi_{32i} & I_3 & 0 & 0 & P_3 \\ I & 0 & 0 & \lambda_1 I_{01} & 0 & 0 \\ 0 & P_2 & 0 & 0 & \lambda_2 I_{02} & 0 \\ 0 & 0 & P_3 & 0 & 0 & \lambda_3 I_{03} \end{bmatrix} \quad (32)$$

and:

$$\begin{aligned} \Phi_{11i} &= A_{ai} - B_a G + A_{ai}^T - G^T B_a^T + I_1 \\ \Phi_{22i} &= P_2 A_{ai} - P_2 U_{2i} + A_{ai}^T P_2 - C_a^T U_{2i}^T + I_2 \\ \Phi_{23i} &= P_2 B_a + C_a^T U_{3i}^T \\ \Phi_{32i} &= \Phi_{23}^T \end{aligned} \quad (33)$$

The observer gains are obtained by:

$$L_i = P_3^{-1} U_{3i} \text{ and } K_i = P_2^{-1} U_{2i}$$

#### 4. Application to the three tanks system

The main objective of this part is to show the robustness of the proposed method by its application to a hydraulic process made up of three tanks [3] and [34].

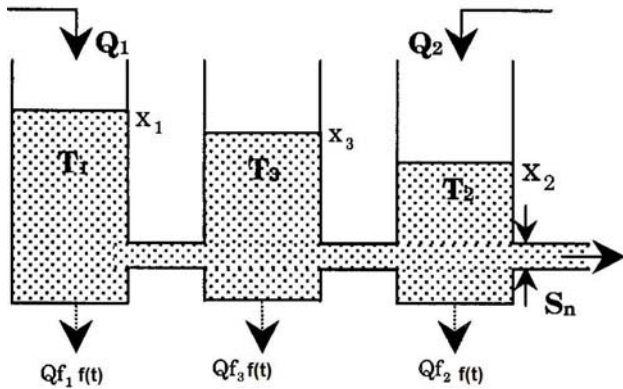


Fig. 1 Three tanks system

The considered system is affected simultaneously by sensor and actuator faults. The three tanks  $T_1, T_2$ , and  $T_3$  with identical sections  $\sigma$ , are connected to each others by cylindrical pipes of identical sections  $S_n$ . The output valve is located at the output of tank  $T_2$ ; it ensures to empty the tank filled by the flow of pumps 1 and 2 with respectively flow rates  $Q_1$  and  $Q_2$ . Combinations of the three water levels are measured. The pipes of communication between the tanks are equipped with manually adjustable ball valves, which allow the corresponding pump to be closed or open. The three levels  $x_1, x_2$  and  $x_3$  are governed by the constraint  $x_1 > x_3 > x_2$ ; the process model is given by the equation (33). Indeed, taking into account the fundamental laws of conservation of the fluid, one can describe the operating mode of each tank; one then obtains a non linear model expressed by the following state equations [3] and [41]

$$\begin{cases} \sigma \frac{dx_1}{dt} = -\alpha_1 S_n (2g(x_1(t) - x_3(t)))^{1/2} + Q_1(t) + Qf_1 \cdot f_a(t) \\ \sigma \frac{dx_2}{dt} = -\alpha_3 S_n (2g(x_3(t) - x_2(t)))^{1/2} \\ \quad - \alpha_2 S_n (2g(x_2(t)))^{1/2} + Q_2(t) + Qf_2 \cdot f_a(t) \\ \sigma \frac{dx_3}{dt} = -\alpha_1 S_n (2g(x_1(t) - x_3(t)))^{1/2} + Qf_3 \cdot f_a(t) \\ \quad - \alpha_3 S_n (2g(x_3(t) - x_2(t)))^{1/2} \end{cases} \quad (34)$$

where  $\alpha_1, \alpha_2$  and  $\alpha_3$  are constants,  $f_a(t)$  is the actuator fault regarded as an unknown input.  $Qf / f_i, i \in \{1...3\}$  denote the additional mass flows into the tanks caused by leaks and  $g$  is the gravity constant. The multiple model, with  $\xi(t) = u(t)$ , which approximates the non linear system (34), is:

$$\begin{cases} \dot{x}(t) = \sum_{i=1}^M \mu_i(\xi(t)) (A_i x(t) + B u(t) + E f_a(t) + d_i) \\ y(t) = C x(t) + F f_s(t) + D w(t) \end{cases} \quad (35)$$

The matrices  $A_i, B_i$ , and  $d_i$  are calculated by linearizing the initial system (34) around four points chosen in the operation range of the system. Four local models have been selected in a heuristic way. That number guarantees a good approximation of the state of the real system by the multiple models [3] and [41]. The following numerical values were obtained:

$$\begin{aligned} A_1 &= \begin{bmatrix} -0.0109 & 0 & 0.0109 \\ 0 & -0.0206 & 0.0106 \\ 0.0109 & 0.0106 & -0.0215 \end{bmatrix}, d_1 = 10^{-3} \begin{bmatrix} -2.86 \\ -0.38 \\ 0.11 \end{bmatrix} \\ A_2 &= \begin{bmatrix} -0.0110 & 0 & 0.0110 \\ 0 & -0.0205 & 0.0104 \\ 0.0110 & 0.0104 & -0.0215 \end{bmatrix}, d_2 = 10^{-3} \begin{bmatrix} -2.86 \\ -0.34 \\ 0.038 \end{bmatrix} \\ A_3 &= \begin{bmatrix} -0.0084 & 0 & 0.0084 \\ 0 & -0.0206 & 0.0095 \\ 0.0084 & 0.0095 & -0.0180 \end{bmatrix}, d_3 = 10^{-3} \begin{bmatrix} -3.7 \\ -0.14 \\ 0.69 \end{bmatrix} \\ A_4 &= \begin{bmatrix} -0.0085 & 0 & 0.0085 \\ 0 & -0.0205 & 0.0095 \\ 0.0085 & 0.0095 & -0.0180 \end{bmatrix}, d_4 = 10^{-3} \begin{bmatrix} -3.67 \\ -0.18 \\ 0.62 \end{bmatrix} \\ B_i &= \frac{1}{\sigma} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, C = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix} \end{aligned}$$

In the following, the functions  $Qf_1, Qf_2$  and  $Qf_3$  are constant, the numerical application are performed with:

$Qf_i = 10^{-4} \forall i \in \{1...4\}$  and  $t \in [0, x[$ ,  $g = 9.8$ ,  $\alpha_1 = 0.78$ ,  
 $\alpha_2 = 0.78$  and  $\alpha_3 = 0.75$ ,  $S_n = 5 \cdot 10^{-5}$  and  $\sigma = 0.0154$ .

The two actuator faults signals  $f_a(t) = [f_{a1}(t) \ f_{a2}(t)]$  are defined as:

$$f_{a1}(t) = \begin{cases} \sin(0.4\pi t), & \text{for } 15s \leq t \leq 75s \\ 0, & \text{elsewhere} \end{cases} \text{ and}$$

$$f_{a2}(t) = \begin{cases} 0.3, & \text{for } 20s \leq t \leq 70s \\ 0.5, & \text{for } t > 70s \\ 0, & \text{elsewhere} \end{cases}$$

It is supposed that a sensor fault  $f_s(t)$  is affecting the system. This fault is defined as follows:

$f_s(t) = [f_{s1}(t) \ f_{s2}(t)]$  with:

$$f_{s1}(t) = \begin{cases} 0, & \text{for } t < 35s \\ 0.6, & \text{for } t \geq 35s \end{cases} \text{ and}$$

$$f_{s2}(t) = \begin{cases} 0, & \text{for } t < 25s \\ \sin(0.6\pi t), & \text{for } t \geq 25s \end{cases}$$

The chosen weighting functions depends on the system input  $u(t)$ . They have been created on the basis of Gaussian membership functions. Figure (2) shows their time-evolution showing that the system is clearly nonlinear since  $\mu_i, i \in \{1, \dots, 4\}$  are not constant functions.

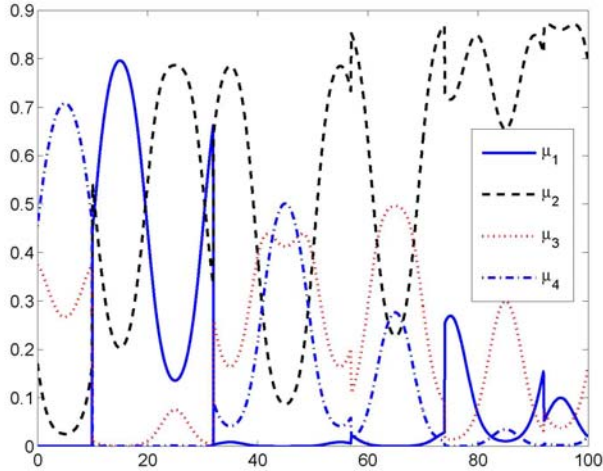


Fig. 2. Weighting functions

Choosing,  $\bar{A} = 10 \times I$  the  $\lambda, K_1, K_2, K_3, K_4, L_1, L_2, L_3, L_4$  and  $G$  computation gives:  $\lambda = 1.2936$ ,

$$L_1 = \begin{bmatrix} 37.57 & 11.81 & -18.60 \\ 26.93 & 8.46 & 23.66 \\ -12.26 & 27.19 & -31.1 \\ -3.78 & 39.31 & 20.545 \end{bmatrix} \quad L_2 = \begin{bmatrix} 37.16 & 11.1 & -20.35 \\ 27.73 & 8.56 & 23.34 \\ -21.75 & 53.53 & -49.31 \\ -11.23 & 74.74 & 33.12 \end{bmatrix}$$

$$L_3 = \begin{bmatrix} 37.98 & 10.65 & -22.71 \\ 30.44 & 8.71 & -24.84 \\ -32.98 & 81.68 & -71.08 \\ -18.38 & 113.29 & 47.59 \end{bmatrix} \quad L_4 = \begin{bmatrix} 36.87 & 10.1 & -22.74 \\ 30.96 & 8.56 & 24.05 \\ -43.14 & 106.16 & -92.72 \\ -24.36 & 148.41 & 62.18 \end{bmatrix}$$

$$K_1 = \begin{bmatrix} -5.24 & -1.17 & -14 \\ 15.32 & 17.34 & 39.18 \\ -7.40 & -1.30 & -9.80 \\ 0.47 & 5.87 & 3.80 \\ -0.08 & 8.87 & 2.12 \\ 4.93 & 3.29 & 14.11 \end{bmatrix} \quad K_2 = \begin{bmatrix} -4.67 & 1.20 & -12.76 \\ 13.85 & 20.79 & 40.87 \\ -9.07 & -1.29 & -8.01 \\ -3.88 & 6.37 & 5.61 \\ 2.36 & 4.68 & 4.49 \\ 5.99 & 3.977 & 11.88 \end{bmatrix}$$

$$K_3 = \begin{bmatrix} -3.91 & 2.67 & -13.4 \\ 10.22 & 25.26 & 45.13 \\ -10.14 & -1.40 & -7.17 \\ -9.05 & -5.74 & -6 \\ 4.56 & -0.03 & 4.28 \\ 7 & 5.03 & 9.81 \end{bmatrix} \quad K_4 = \begin{bmatrix} -3.79 & 6.85 & -12.68 \\ 9.01 & 26.6 & 36.03 \\ -11.86 & -1.24 & -5.94 \\ -13.37 & 6.66 & 7.90 \\ 6.86 & -3.94 & 5.54 \\ 8.47 & 3.65 & 5.40 \end{bmatrix}$$

$$G = \begin{bmatrix} -2.44 & -2.99 & 1.96 & -4.35 & 3.17 & 7.58 \\ -0.53 & 4.53 & 7.53 & -2.52 & 5.43 & -1.02 \end{bmatrix}$$

The obtained results are shown in figures (3) to (7).

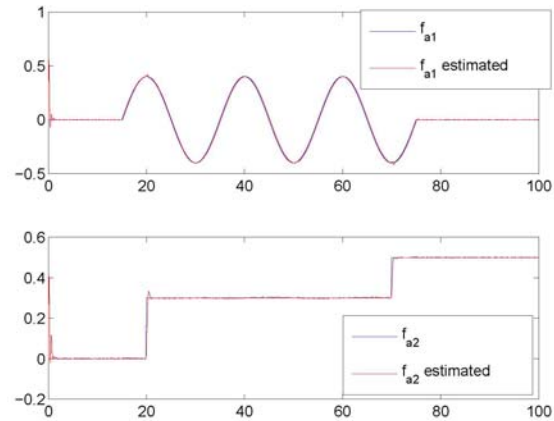


Fig. 3. Actuator faults and their estimation

Figure (3) visualizes the two actuator faults ( $f_{a1}(t)$  and  $f_{a2}(t)$ ) and their estimations, the two sensor faults ( $f_{s1}(t)$  and  $f_{s2}(t)$ ) and their estimations are represented in figure (4). In figure (5), the state error estimation is visualized.

These three figures show that the proposed observer permits to estimate simultaneously the sensor and actuator faults and the system state. The application of the proposed method to the three tanks system shows its robustness. Simulation results show that the fault is estimated well and the effect of the measurement noise is minimized. This method allows estimating well the sensor and actuator faults even in the case of time-varying faults.



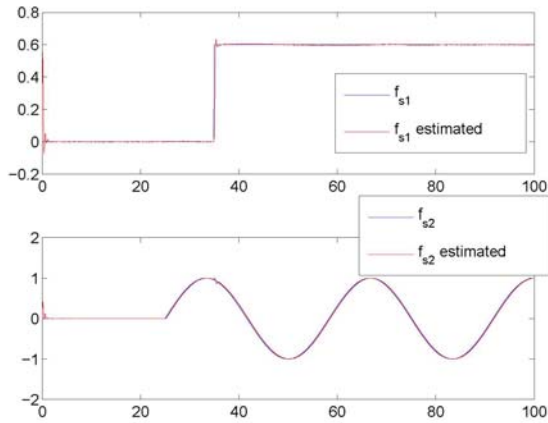


Fig. 4. Sensor faults and their estimation

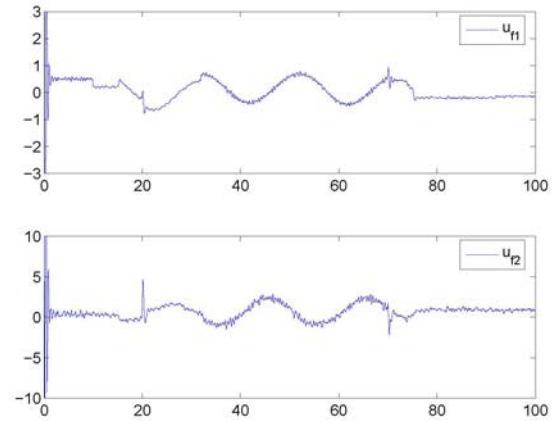


Fig. 7. Fault tolerant control  $u_f$

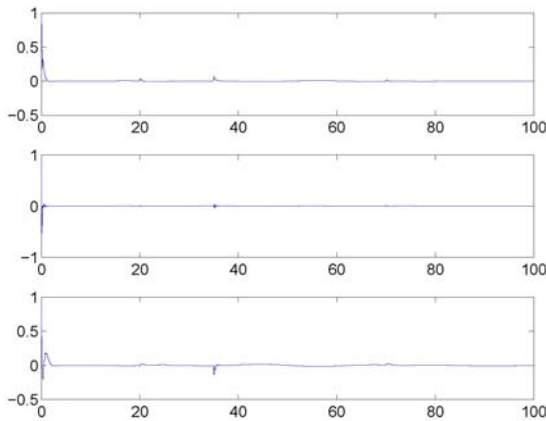


Fig. 5. state error estimation

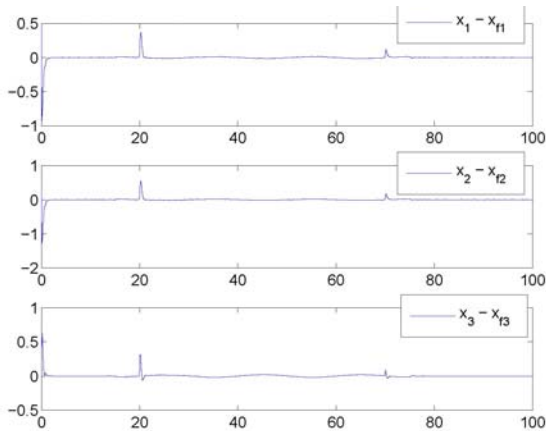


Fig.6. error between  $x$  and  $x_f$

Figure (6) shows the time-evolution of the error  $\tilde{X}(t)$  between the reference state  $x(t)$  and the faulty state  $x_f(t)$ . This error converges toward zero. So the application of the conceived fault tolerant control law  $u_f(t)$  to the faulty system let the behavior of the system affected by the sensor and the actuator fault similar to the reference system behavior. The action of the proposed fault tolerant control is quick.

Fault and state estimation is very important because the fault and state estimated are used to conceive the fault tolerant control strategy. This control is shown in the figure (7)

## 5. Conclusion

This work proposes a direct application of the use of proportional integral observer to the fault tolerant control design. This control was conceived for systems described by Takagi-Sugeno fuzzy models with weighting function depending on the FTC. The proposed method is based on the estimation of the state and faults affecting the system. To make faults estimation, a mathematical transformation was used to conceive an augmented system in which the sensor fault affecting the initial system appears as an actuator fault. Then an adaptive proportional integral observer is used to estimate simultaneously actuator and sensor faults and the system state. The main contribution in this work is that the considered systems have the weighting functions depending on the fault tolerant control which is a very important case and is the subject of few works and in the use of the mathematical transformation and the proportion integral observer to estimate time-varying sensor and actuator faults. The FTC controller is implemented as a state feedback controller. This controller



is designed such that it can stabilize the faulty plant using Lyapunov theory and LMIs.

## References

- [1] A. Akhenak, M. Chadli, J. Ragot and D. Maquin, "Design of observers for Takagi-Sugeno fuzzy models for Fault Detection and Isolation", 7<sup>th</sup> IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes SAFEPROCESS'09, Barcelona, Spain, June 30th - July 3rd, 2009.
- [2] A. Akhenak M. Chadli J. Ragot and D. Maquin "Design of sliding mode unknown input observer for uncertain Takagi-Sugeno model". 15<sup>th</sup> Mediterranean Conference on Control and Automation, MED'07, Athens, Greece, June 27-29, 2007.
- [3] A. Akhenak, M. Chadli, D. Maquin, and J. Ragot, "State estimation via multiple observers. The three tank system". 5<sup>th</sup> IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes, Safeprocess'03, Washington, D.C., USA, June 9-11, 2003.
- [4] S. Beale and B. Shafai "Robust control system design with a proportional integral observer", International Journal of Control, Vol. 50 No. 1, 1989, pp. 97-111.
- [5] M. Blanke, M. Kinnaert, J. Lunze, and M. Staroswiecki "Diagnosis and Fault-Tolerant Control". Springer-Verlag Berlin Heidelberg. ISBN 3-540-01056-4, 2003
- [6] J. Chen and R.J. Patton "Fault-tolerant control systems design using the linear matrix inequality approach". 6<sup>th</sup> European Control Conference, Porto, Portugal, 4-7 September, 2001
- [7] C. Edwards, "A comparison of sliding mode and unknown input observers for fault reconstruction". IEEE Conference on Decision and Control, Vol. 5, 2004, pp. 5279-5284.
- [8] C. Edwards, and S.K Spurgeon, "On the development of discontinuous observers". International Journal of Control, Vol. 59 No. 5, 1994, pp. 1211-1229.
- [9] D. Filev, Fuzzy "modeling of complex systems. International Journal of Approximate Reasoning", Vol. 5, N 3, 1991, pp. 281-290.
- [10] Y. Guan, and M. Saif, "A novel approach to the design of unknown input observers", IEEE Trans on Automatic Control, AC-Vol. 36, No. 5, 1991, pp. 632-635.
- [11] D. Ichalal, B. Marx, J. Ragot and D. Maquin, "New fault tolerant control strategy for nonlinear systems with multiple model approach". Conference on Control and Fault-Tolerant Systems, SysTol'10, October 6-10, 2010
- [12] D. Ichalal, B. Marx, J. Ragot and D. Maquin, "Simultaneous state and unknown inputs estimation with PI and PMI observers for Takagi-Sugeno model with unmeasurable premise variables". 17<sup>th</sup> Mediterranean Conference on Control and Automation, MED'09, Thessaloniki, Greece, June 24-26, 2009.
- [13] T. A. Johansen, and A.B. Foss, "Non linear local model representation for adaptive systems". Singapore International Conference on Intelligent Control and Instrumentation, Singapore, February 17-21, 1992.
- [14] S. Kawamoto, K. Tada, N. Onoe, A. Ishigame and T. Taniguchi, "Construction of exact fuzzy system for non linear system and its stability analysis", 8<sup>th</sup> Fuzzy System Symposium, Hiroshima, Japan, 1992, pp. 517-520.
- [15] A. Khedher, K. Ben Othman, D. Maquin and M. Benrejeb "Sensor fault estimation for nonlinear systems described by Takagi-Sugeno models International Journal" Transaction on system, signal & devices, Issues on Systems, Analysis & Automatic Control, Vol. 6, No. 1, 2011, pp.1-18.
- [16] A. Khedher, K. Ben Othman, D. Maquin and M. Benrejeb "Design of an adaptive faults tolerant control: case of sensor faults", Wseas Transactions on Systems. Vol. 9, No. 7, 2010, pp 794-803.
- [17] A. Khedher, K. Ben Othman, M. Benrejeb and D. Maquin "Adaptive observer for fault estimation in nonlinear systems described by a Takagi-Sugeno model". 18<sup>th</sup> Mediterranean Conference on Control and Automation, MED'10, June 24-26, Marrakech, Morocco, 2010.
- [18] A. Khedher, K. Ben Othman, D. Maquin and M. Benrejeb "An approach of faults estimation in Takagi-Sugeno fuzzy systems" 8<sup>th</sup> ACS/IEEE International Conference on Computer Systems and Applications Hammamet Tunisia May 16-19, 2010.
- [19] A. Khedher, K. Ben Othman, D. Maquin and M. Benrejeb "Fault tolerant control for nonlinear system described by Takagi-Sugeno models" 8<sup>th</sup> International Conference of Modeling and Simulation - MOSIM'10 - Hammamet - Tunisia - May 10-12, 2010.
- [20] A. Khedher, K. Ben Othman, D. Maquin and M. Benrejeb "Active sensor faults tolerant control for Takagi-Sugeno multiple models". 6<sup>th</sup> WSEAS International Conference on dynamical systems & control (CONTROL'10) Sousse, Tunisia, May 3-6, 2010.
- [21] A. Khedher and K. Ben Othman, "Proportional Integral Observer Design for State and Faults Estimation: Application to the Three Tanks System". International Review of Automatic Control Vol. 3, No 2, 2010, pp. 115-124,
- [22] A. Khedher, K. Ben Othman, D. Maquin and M. Benrejeb "State and sensor faults estimation via a proportional integral observer". 6<sup>th</sup> international multi-conference on Systems signals & devices SSD'09 March 23-26, Djerba, Tunisia, 2009.
- [23] A. Khedher, K. Ben Othman, M. Benrejeb and D. Maquin "State and unknown input estimation via a proportional integral observer with unknown inputs". 9<sup>th</sup>

international conference on Sciences and Techniques of Automatic control and computer engineering STA'2008 December 20-23, Sousse, Tunisia, 2008.

[24] J. Korbicz, J. Kościelny, Z. Kowalczyk, and W. Cholewa, "Fault diagnosis. Models, Artificial Intelligence, Applications". Springer-Verlag, Berlin, 2004.

[25] Y. Liang, D. Liaw, and T. Lee, "Reliable control of nonlinear systems". Vol. 45, 2000, pp. 706-710.

[26] F. Liao, J. Wang, and G. Yang, "Reliable robust flight tracking control: an LMI approach". IEEE Transaction Control Systems Technic, Vol. 10, 2002, pp. 76-89.

[27] D.G. Luenberger, "An introduction to observers. IEEE Transactions on Automatic Control", vol. 16 No. 6, 1971, pp. 596-602.

[28] L. M. Lyubchik and Y. T. Kostenko, "The output control of multivariable systems with immeasurable arbitrary disturbances - The inverse model approach". ECC'93, pp. 1160-1165, Groningen, Netherlands, June 28-July 1, 1993.

[29] R. Murray-Smith and T. Johansen "Multiple model approaches to modeling and control". Taylor and Francis, London, 1997.

[30] R. Orjuela, B. Marx, J. Ragot, and D. Maquin, "On the simultaneous state and unknown inputs estimation of complex systems via a multiple model strategy", IET Control Theory & Applications, Vol. 3 (7):877-890, 2009.

[31] R. Orjuela, B. Marx, J. Ragot, and D. Maquin, "Proportional-Integral observer design for nonlinear uncertain systems modeled by a multiple model approach" 47<sup>th</sup> IEEE Conference on Decision and Control, Cancun, Mexico, December 9-11, 2008.

[32] Z. Qu, C. M. Ihlefeld, J. Yufang, and A. Saengdeejing, "Robust fault-tolerant self-recovering control of nonlinear uncertain systems". Automatica, Vol. 39, 2003, pp. 1763-1771.

[33] Z. Qu, C.M. Ihlefeld, J. Yufang, and A. Saengdeejing, "Robust control of a class of nonlinear uncertain systems. fault tolerance against sensor failures and subsequent self recovery". In Proceedings of the IEEE Conference on Decision and Control, Vol. 2, 2001, pp: 1472-1478.

[34] M., Takagi, and M. Sugeno, "Fuzzy identification of systems and its application to modeling and control", IEEE Transactions on Systems Man and Cybernetics, Vol. 15 No. 1, 1985, pp. 116-132.

[35] K. Tanaka, T. Ikeda, and Y. Y. He, "Fuzzy regulators and fuzzy observers: relaxed stability conditions and LMI-based design", IEEE Transaction on Fuzzy Systems, Vol. 6, No.1, 1998, pp. 250-256.

[36] B. L. Walcott and S. H. Zak, "Observation of dynamical systems in the presence of bounded

nonlinearities/uncertainties". 25<sup>th</sup> IEEE Conference on Decision and Control, pp. 961-966, 1988.

[37] S.H. Wang, E. J. Davison and P. Dorato, "Observing the states of systems with immeasurable disturbances". IEEE Transactions on Automatic Control, AC-20, pp. 716-717, 1975.

[38] M. Witczak, L. Dziekan, V. Puig and J. Korbicz. "A fault-tolerant control strategy for Takagi-Sugeno fuzzy systems". 17<sup>th</sup> IFAC World Congress, Seoul, Korea, July 6-11, 2008.

[39] J. Yoneyama, "Robust  $H_\infty$  control of uncertain fuzzy system under time-varying sampling", Fuzzy Sets and Systems Vol. 160, 2009, 1738-1748.

[40] Y. Zhang and J. Jiang, "Bibliographical review on reconfigurable fault-tolerant control systems". Proceedings of IFAC SAFEPROCESS, pp.265-276, 2003.

[41] A. Zolghadri, D. Henry, and M. Morsion, "Design of non linear observers for fault diagnosis: a case study", Control Engineering Practice, Vol. 4 No.11, 1996, pp. 1535-1544.

**Atef Khedher** was born in Tunisia in 1980. He obtained the Engineer degree in electro-Mechanical engineering from the "Ecole Nationale d'Ingénieurs de Sfax (ENIS)" in 2003 and obtained the master degree in automatic and industrial Maintenance from the "Ecole Nationale d'Ingénieur de Monastir" in 2005. He obtained the PhD degree in the electrical engineering from the "Ecole Nationale d'Ingénieur de Tunis (ENIT)" in 2011. His research is related to state and faults estimation and the faults tolerant control for takagi-sugeno fuzzy systems.

**Kamel Ben Othman** was born in Tunisia in 1958. He obtained the Engineer degree in Mechanical and Energetic engineering from the "Université de Valenciennes" in 1981 and obtain the PhD degree in automatic and signal processing from the "Université de Valenciennes" in 1984 and the HDR from the "Ecole Nationale d'Ingénieur de Tunis" in 2008. He is currently professor at "ISSTE Gafsa". His research is related to Reliability, fuzzy systems and Diagnosis of complex systems.

**Mohamed Benrejeb** was born in Tunisia in 1950. He obtained the Diploma of "Ingénieur IDN" (French "Grande Ecole") in 1973, The Master degree of Automatic Control in 1974, the PhD in Automatic Control of the University of Lille in 1976 and the DSc of the same University in 1980. Full Professor at "Ecole Nationale d'Ingénieurs de Tunis" since 1985 and at "Ecole Centrale de Lille" since 2003, his research interests are in the area of analysis and synthesis of complex systems based on classical and non conventional approaches.

## Efficient Spatial Data mining using Integrated Genetic Algorithm and ACO

Mr.K.Sankar<sup>1</sup> and Dr. V.Vankatachalam<sup>2</sup>

<sup>1</sup>Assistant Professor(Senior), Department of Master of Computer Applications, KSR  
College of Engineering, Tiruchengode

<sup>2</sup> Principal, The KAVERY Engineering College, Mecheri, Salem

### Abstract

Spatial data plays a key role in numerous applications such as network traffic, distributed security applications such as banking, retailing, etc., The spatial data is essential mine, useful for decision making and the knowledge discovery of interesting facts from large amounts of data. Many private institutions, organizations collect the number of congestion on the network while packets of data are sent, the flow of data and the mobility of the same. In addition other databases provide the additional information about the client who has sent the data, the server who has to receive the data, total number of clients on the network, etc. These data contain a mine of useful information for the network traffic risk analysis. Initially study was conducted to identify and predict the number of nodes in the system; the nodes can either be a client or a server. It used a decision tree that studies from the traffic risk in a network. However, this method is only based on tabular data and does not exploit geo routing location. Using the data, combined to trend data relating to the network, the traffic flow, demand, load, etc., this work aims at deducing relevant risk models to help in network traffic safety task.

The existing work provided a pragmatic approach to multi-layer geo-data mining. The process behind was to prepare input data by joining each layer

table using a given spatial criterion, then applying a standard method to build a decision tree. The existing work did not consider multi-relational data mining domain. The quality of a decision tree depends, on the quality of the initial data which are incomplete, incorrect or non relevant data inevitably leads to erroneous results. The proposed model develops an ant colony algorithm integrated with GA for the discovery of spatial trend patterns found in a network traffic risk analysis database. The proposed ant colony based spatial data mining algorithm applies the emergent intelligent behavior of ant colonies. The experimental results on a network traffic (trend layer) spatial database show that our method has higher efficiency in performance of the discovery process compared to other existing approaches using non-intelligent decision tree heuristics.

Keywords: Spatial data mining, Network Traffic, ACO, GA

### 1. Introduction

Data mining is the process of extracting patterns from large data sets by combining methods from statistics and artificial intelligence with database management. Given an informational system, data mining is seen to be used as an important tool to transform data into business intelligence process. It is currently used in wide range of areas such as marketing, surveillance, fraud

detection, and scientific discovery. Automatic data processing is the result of the increase in size and complexity of the data set. This has been used in other areas of computer science as neural networks, support vector machines, genetic algorithms and decision trees. A primary reason for using data mining is to assist in the analysis of collections of observations of network user behavior.

Spatial data mining try to find patterns in geographic data. Most commonly used in retail, it has grown out of the field of data mining, which initially focused on finding patterns in network traffic analysis, security threats over a period of time, textual and numerical electronic information. It is considered to be more complicated challenge than traditional mining because of the difficulties associated with analyzing objects with concrete existences in space and time. Spatial patterns may be discovered using techniques like classification, association, and clustering and outlier detection. New techniques are needed for SDM due to spatial auto-correlation, importance of non-point data types, continuity of space, regional knowledge and separation between spatial and non-spatial subspace. The explosive growth of spatial data and widespread use of spatial databases emphasize the need for the automated discovery of spatial knowledge. Our focus of this work is on the methods of spatial data mining, i.e., discovery of interesting knowledge from spatial data of network traffic patterns. Spatial data are related to traffic data objects that occupy space.

The institutions concern the routing network studies the application of data mining techniques for network

traffic risk analysis. The proposed work aims at spatial feature of the traffic load and demand requirements and their interaction with the geo routing environment. In previous work, the system has implemented some spatial data mining methods such as generalization and characterization. The proposal of this work uses intelligent ant agent to evaluate the search space of the network traffic risk analysis along with usage of genetic algorithm for risk pattern.

## 2. Literature Review

Spatial data mining fulfills real needs of many geomantic applications. It allows taking advantage of the growing availability of geographically referenced data and their potential richness. This includes the spatial analysis of risk such as epidemic risk or network traffic accident risk in the router. This work deals with the method of decision tree for spatial data classification. This method differs from conventional decision trees by taking account implicit spatial relationships in addition to other object attributes. Ref [2, 3] aims at taking account of the spatial feature of the packets transmissions and their interaction with the geographical environment.

How are spatial data handled in usual data mining systems? Although many data-mining applications deal at least implicitly with spatial data they essentially ignore the spatial dimension of the data, treating them as non-spatial. This has ramifications both for the analysis of data and for their visualization. First, one of the basic tasks of exploratory data analysis is to present the salient features of a data set in a

format understandable to humans. It is well known that visualization in geographical space is much easier to understand than visualization in abstract space. Secondly, results of a data mining analysis may be suboptimal or even be distorted if unique features of spatial data, such as spatial autocorrelation ([7]), are ignored. In sum, convergence of GIS and data mining in an Internet enabled spatial data mining system is a logical progression for spatial data analysis technology. Related work in this direction has been done by Koperski and Han, Ester et al. [4, 9].

Rather than aggregate data, Gridfit [1] avoids overlap in the 2D display by repositioning pixels locally. In areas with high overlap, however, the repositioning depends on the ordering of the points in the database, which might be arbitrary. Gridfit places the first data item found in the database at its correct position, and moves subsequent overlapping data points to nearby free positions, making their placement quasirandom. Cartograms [5] are another common technique dealing with advanced map distortion. Cartogram techniques let data analysts trade shape against area and preserve the map's topology to improve map visualization by scaling polygonal elements according to an external parameter. Thus, in cartogram techniques, the rescaling of map regions is independent of a local distribution of the data points. A cartogram-based map distortion provides much better results, but solves neither the overlap nor the pixel coherence problems. Even if the cartogram provides a perfect map distortion (in many cases, achieving a perfect distortion is impossible), many data points might be at the same location, and

there might be little pixel coherence. Therefore, cartogram-based distortion is primarily a preprocessing step.

In [8] the author proposes an Improved Ant Colony Optimization (IACO) and Hybrid Particle Swarm Optimization (HPSO) method for SCOC. In the process of doing so, the system first use IACO to obtain the shortest obstructed distance, which is an effective method for arbitrary shape obstacles, and then the system develop a novel HPKSCOC based on HPSO and K-Medoids to cluster spatial data with obstacles, which can not only give attention to higher local constringency speed and stronger global optimum search, but also get down to the obstacles constraints. Spatial clustering is an important research topic in Spatial Data Mining (SDM). Many methods have been proposed in the literature, but few of them have taken into account constraints that may be present in the data or constraints on the clustering. These constraints have significant influence on the results of the clustering process of large spatial data. In this project, the system discuss the problem of spatial clustering with obstacles constraints and propose a novel spatial clustering method based on Genetic Algorithms (GAs) and KMedoids, called GKSCOC, which aims to cluster spatial data with obstacles constraints.[9]

### **3. Genetic and ACO Based Spatial Data Mining Model**

Before data mining algorithms can be used, a target data set must be collected. As data mining only uncover patterns already present in the data, the target dataset must be large enough to contain these patterns. A common source

for data is a data mart or data warehouse. Pre-process is essential to analyze the multivariate datasets before clustering or data mining. The target set is then cleaned. Cleaning removes the observations with noise and missing data. The clean data are reduced into feature vectors, one vector per observation. A feature vector is a summarized version of the raw data observation. This might be turned into a feature vector by locating the eyes and mouth in the image. The feature vectors are divided into two sets, the "training set" and the "test set". The training set is used to "train" the data mining algorithm(s), while the test set is used to verify the accuracy of any patterns found

The proposed spatial data mining model uses ACO integrated with GA for network risk pattern storage. The proposed ant colony based spatial data mining algorithm applies the emergent intelligent behavior of ant colonies. The proposed system handle the huge search space encountered in the discovery of spatial data knowledge. It applies an effective greedy heuristic combined with the trail intensity being laid by ants using a spatial path. GA uses searching population to produce a new generation population. The proposed system develops an ant colony algorithm for the discovery of spatial trends in a GIS network traffic risk analysis database. Intelligent ant agents are used to evaluate valuable and comprehensive spatial patterns.

### 3.1. Geo-Spatial Data Mining

Data volume was a primary factor in the transition at many federal agencies from delivering public domain data via physical mechanisms. Algorithmic requirements differ

substantially for relational (attribute) data management and for topological (feature) data management. Geographic data repositories increasingly include ill structured data such as imagery and geo referenced multimedia. The strength of network GIS is in providing a rich data infrastructure for combining disparate data in meaningful ways by using spatial proximity.

The next logical step to take Network GIS analysis beyond demographic reporting to true market intelligence is to incorporate the ability to analyze and condense a large number of variables into a single forecast or score. This is the strength of predictive data mining technology and the reason why there is such a true relationship between Network GIS & data mining. Depending upon the specific application, Network GIS can combine historical customer or retail store sales data with syndicated demographic, business, network traffic, and market research data. This dataset is then ideal for building predictive models to score new locations or customers for sales potential, cross-selling, targeted marketing, customer churn, and other similar applications. Geospatial data repositories tend to be very large. Moreover, existing GIS datasets are often splintered into feature and attribute components that are conventionally archived in hybrid data management systems. Algorithmic requirement differ substantially for relational (attribute) data management and for topological (feature) data management.

### 3.2 Ant Colony Optimization

Ant colony Optimization algorithm (ACO), a probabilistic



technique is deployed for evaluating spatial data inference from network traffic patterns which find load and demand at various instances. In the natural world, ants (initially) wander randomly, and upon finding food return to their colony while laying down pheromone trails. If other ants find such a path, they are likely not to keep traveling at random, but to instead follow the trail, returning and reinforcing it if they eventually find food.

Ant Colony Optimization (ACO) is a paradigm for designing meta-heuristic algorithms for combinatorial optimization problems. Meta-heuristic algorithms are algorithms which, in order to escape from local optima, drive some basic heuristic, either a constructive heuristic starting from a null solution and adding elements to build a good complete one, or a local search heuristic starting from a complete solution and iteratively modifying some of its elements in order to achieve a better one. The metaheuristic part permits the low level heuristic to obtain solutions better than those it could have achieved alone, even if iterated. The characteristic of ACO algorithms is their explicit use of elements of previous solutions

Over time, however, the pheromone trail starts to evaporate, thus reducing its attractive strength. The more time it takes for an ant to travel down the path and back again, the more time the pheromones have to evaporate. A short path, by comparison, gets marched over faster, and thus the pheromone density remains high as it is laid on the path as fast as it can evaporate. Pheromone evaporation has also the advantage of avoiding the convergence to a locally

optimal solution. If there were no evaporation at all, the paths chosen by the first ants would tend to be excessively attractive to the following ones. In that case, the exploration of the solution space would be constrained.

Thus, when one ant finds a good (i.e., short) path from the colony to a food source, other ants are more likely to follow that path, and positive feedback eventually leads all the ants following a single path. The idea of the ant colony algorithm is to mimic this behavior with "simulated ants" walking around the graph representing the problem to solve.

### 3.3 Genetic Algorithm

The proposed algorithm of spatial clustering based on GAs is described in the following procedure. Divide an individual risk pattern of the network traffic generating objects (chromosome) into  $n$  part and each part is corresponding to the classification of a datum element. The optimization criterion is defined by a Euclidean distance among the data frequently, and the initial number of packets that has to be sent is produced at random. Its genetic operators are similar to standard GA's. This method can find the global optimum solution and not influenced by an outlier, but it only fits for the situation of small network traffic risk pattern data sets and classification number.

## 4. Experimental Evaluation

ACO with GA integration SPDM model is proposed to be tested in the framework of network traffic risk analysis. The analysis is done on a spatial database provided in the

framework of an industrial collaboration. It contains data on the number of packets to be sent and others on the number of nodes that is ready to be served in the network. The objective is to construct a predictive model. The system model looks correspondences between the packet and the other trend layers as the number of nodes, time taken for the packet to reach at the other end etc. It applies classification by decision tree while integrating the number of packets to be transmitted via spatial character and their interaction with the geographical environment. The experimental evaluation is made on a geographical network traffic (trend layer) spatial database to depict higher efficiency in performance of the discovery process. It proves that better quality of trend patterns discovered compared to other existing approaches using non-intelligent decision tree heuristics. Reliable data constitute the key to success of a decision tree. An efficient parallel and near global optimum search for network traffic risk patterns are evaluated using genetic algorithm. It combines the concept of survival of the fittest with a structured interchange. GAs imitates natural selection of the biological evolution. Improvements in the identification of high or low risk areas can assist the emergency preparedness planning and resource evaluation.

#### **4.1 Spatial Data Mining on Network Traffic Risk Patterns**

Spatial data mining on network traffic risk pattern focuses on the human vulnerability in built environments. It considers issues like differences between common and rare collision, commuting of people, and relations between

accidents and networks. Visualization and interaction helps to understand the dependencies within and between data sets. Visualization supports formulating hypotheses and answering questions about correlations between certain variables and collision. Explorative visualization may reveal new variables relevant to the model and relevance of already used variables. It is highly required to analyze the correlations combine spatial data analysis methods with visualization. Risk model development is an interactive and explorative process.

#### **4.2 ACO on SPDM**

ACO has been recently used in some data mining tasks, e.g. classification rule discovery. Considering the challenges faced in the problem of spatial trend detection, ACO suggest efficient properties in these aspects. Ant agents search for the trend starting from their own start point in a completely distributed manner. This guides the search process to infer to a better subspace potentially containing more and better trend patterns. Finally some measures of attractiveness can be defined for selecting a feasible spatial object from the neighborhood graph. ACO on SPDM Effectively guide the trend detection process of an ant ACO has been recently used in some data mining tasks, e.g., classification rule discovery. Considering the challenges faced in the problem of spatial trend detection, ACO suggest efficient properties in these aspects. Ant agents search for the trend starting from their own start point in a completely distributed manner. Finally some measures of attractiveness can be

defined for selecting a feasible spatial object from the neighborhood graph.

### 4.3 Spatial Clustering GA

Genetic algorithms are an efficient parallel and near global optimum search method based on nature genetic and selection. GA combines the concept of survival of the fittest with a structured interchange. Gas imitates natural selection of the biological evolution. It uses searching population (set) to produce a new generation population. GAs automatically achieve and accumulate the knowledge about the search space. GA adaptively controls the search process to approach a global optimal solution. GA performs well in highly constrained problems, where the number of “good” solutions is very small relative to the size of the search space. GAs provides better solution in a shorter time, including complex problems to solve by traditional methods.

## 5. Result and Discussions

The proposed results provide spatial decision trees for network traffic risk patterns with optimized route structure with the ant agents. The proposed model classifies objects according to spatial information (using the ant agent and the distance pheromone). Spatial classification provided by the proposed scheme is simple and efficient. It allows adapting to different decision tree algorithm for the spatial modeling of network traffic risk patterns. It uses the structure of geo-data in multiple trend layers which is characteristic of geographical databases. Finally, the quality of this analysis is improved by enriching the spatial database by multiple geographical trends, and by a close collaboration with

a domain specialist in traffic risk analysis. The advantage of proposed technique allows the end-user to evaluate the results without any assistance by an analyst or statistician. Gas automatically achieve and accumulate the knowledge about the search space of the ACO. GA adaptively controls the traffic risk pattern search process to approach a global optimal solution. Perform well in highly constrained traffic risk pattern, where the number of “good” solutions is very small relative to the size of the search space.

The current application results show a use case of spatial decision trees. The contribution of this approach to spatial classification lies in its simplicity and its efficiency. It makes it possible to classify objects according to spatial information (using the distance). It allows adapting any decision tree algorithm or tool for a spatial modeling problem. Furthermore, this method considers the structure of geo-data in multiple trends (patterns) which is characteristic of geographical databases. The graph below indicates the number of trends found and paths examined using SPDM Decision Tree and SPDM-ACO-GA models for traffic risk pattern analysis.

## 6. Conclusion

The Spatial data mining system of ACO with GA have shown that network traffic risk patterns are discovered efficiently and recorded in the genetic property for avoiding the collision risk in highly dense spatial regions. The proposal of our system analyzes existing methods for spatial data mining and mentioned their strengths and weaknesses. The variety of yet unexplored topics and problems

makes knowledge discovery in spatial databases an attractive and challenging research field. This work gives an efficient approach to multi-layer geo-data mining. The main idea is to prepare input data by joining each layer table using a given spatial criterion, then applying a standard method to build a decision tree. The most advantage is to demonstrate the feasibility and the interest of integrating neighborhood properties when analyzing spatial objects. Our future work will focus on adapting recent work in multi-relational data mining domain, in particular on the extension of the spatial decision trees based on neural network. Another extension will concern automatic filtering of spatial relationships. The system will study of its functional behavior and its performances for concrete cases, which has never been done before. Finally, the quality of this analysis could be improved by enriching the spatial database by other geographical trends, and by a close collaboration with a domain specialist in traffic risk analysis. Indeed, the quality of a decision tree depends, on the whole, of the quality of the initial data.

## REFERENCES

[1] . D.A. Keim and A. Herrmann, "The Gridfit Algorithm: An Efficient and Effective Approach to Visualizing Large Amounts of Spatial Data," *Proc. IEEE Visualization Conf.*, IEEE CS Press, 1998, pp. 181-188.  
[2] Anselin, L. 1988. *Spatial Econometrics: Methods and Models*. Dordrecht, Netherlands: Kluwer.  
Anselin, L. 1994. *Exploratory Spatial Data Analysis and Geographic Information Systems*. In Painho, M., ed., *New Tools for Spatial Analysis*, 45-54.

[3] Anselin, L. 1995. *Local Indicators of Spatial Association: LISA*. *Geographical Analysis* 27(2):93-115.

[4] Ester, M., Frommelt, A., Kriegel, H.P, Sander, J., "Spatial Data Mining: Database Primitives, Algorithms and Efficient DBMS Support", in *Data Mining and Knowledge Discovery, an International Journal*, 1999

[5] D.A. Keim, S.C. North, and C. Panse, "Cartodraw: A Fast Algorithm for Generating Contiguous Cartograms," *IEEE Trans. Visualization and Computer Graphics* (TVCG), vol. 10, no. 1, 2004, pp. 95-110.

[6] Haining, R. *Spatial data analysis in the social and environmental sciences*, Cambridge Univ. Press, 1991

[7] Hawkins, D. 1980. *Identification of Outliers*. Chapman and Hall. [Jain & Dubes1988] Jain, A., and Dubes, R. 1988. *Algorithms for Clustering Data*. Prentice Hall.

[8] Jhung, Y., and Swain, P. H. 1996. *Bayesian Contextual Classification Based on Modified M-Estimates and Markov Random Fields*. *IEEE Transaction on Pattern Analysis and Machine Intelligence* 34(1):67-75.

[9] Koperski, K., Han, J. "GeoMiner: A System Prototype for Spatial Mining", *Proceedings ACM SIGMOD*, Arizona, 1997.

**K.Sankar** is a Research Scholar at the Anna University Coimbatore. He is now working as a Assistant Professor(Sr) at KSR College of Engineering, Tiruchengode. His Research interests are in the field of Data Mining and Optimization Techniques.

**Dr.V.Vankatachalam** is a principal of The Kavary Engineering College. He received his B.E in Electronics and

Communication at Coimbatore Institute of Technology Coimbatore. He obtained his M.S degree in Software systems from Birla Institute of Technology Pilani. He did his M.Tech in Computer Science at Regional Engineering College (REC) Trichy. He obtained his Ph.D degree in Computer Science and Engineering from Anna University Chennai. He has published 3 papers in International Journal and 20 papers in International & National conferences.

# Electronic Seal Stamping Based on Group Signature

Girija Srikanth<sup>1</sup>

<sup>1</sup> Department of CSE, Birla Institute of Technology,  
Al Dhait south, Ras Al Kaimah, UAE

## Abstract

This paper describes a new electronic official seal stamping based on Group Signature, USB Key. Bill/Contract in E-commerce must be seal stamped to gain tamper proof and non-repudiation. The seal stamping control is designed based on the certificate-based public key. This technique is more efficient for generating and verifying individual/group signatures in terms of computational efforts and communication costs. Web page electronic seal-stamping system is implemented which has been adopted by CNBAB platform since Mar., 2008.

**Keywords:** *Digital Signature, Self certified public key, Seal Stamp, USB key*

## 1. Introduction

CNBAB [1][2][3] is an e-commerce platform which constructs a credit worthy trade environment and provides financing channels for Chinese small and medium-sized enterprise (SME) and even small enterprises. CNBAB has carried an operation in Shandong province and achieved a great success. CNBAB adopts a brand new business pattern called BAB (Business agent business) [4]. Business Agent is an agent who handles business affairs for another, especially one who deals with employers. An agent is a representative who acts on behalf of other persons/organizations. BAB pattern can provide credit guarantee and solve quickly transactional fund storage for SMEs. CNBAB constructs an aggregate called agent to guarantee reliable trading environment by combining banks, the government, the digital authentication centre and third party quality supervision institutions, in which every party undertakes different responsibility throughout the entire trading process. There are three kinds of users and eighteen kinds of agent staffs in CNBAB.

Enterprise user or individual user can become register by registering in CNBAB. Register user can apply for becoming contracted user. Contracted user can trade in CNBAB platform. Contracted user can apply for becoming core user by submitting appointed materials to CNBAB and banks. If these materials are materials are audited to pass, contracted user can become core user.

CNBAB launches trade-currency service similar to short term loans for core user to resolve financing problem. Trade currency guarantees trade between users, banks guarantees the value of trade-currency to assure smooth trade steps.

After achieving a transaction between users, both sides need to sign a contract. And in sequence every stage of trade process, users need to fill in some bills and agent staffs need to audit these bills; some agent staffs need to fill in some bills and other agent staffs need to audit these bills. There are twenty-six kinds of contract templates and sixteen kinds of bills in CNBAB. Bill/contract must be seal-stamped to gain tamperproof and non-repudiation.

Seal-stamping on web page is a method that allows a person to 'seal' documents in a manner parallel to the traditional seal. Seal-stamping on web page can be regarded as electrification of the traditional seal and the handwritten signature. Combining the digital signature with the seal image prevent bill/contract from altering and denying.

## 2. Literature Survey

### 2.1 Web page seal-stamping and verify

CNBAB is developed in java language based on IBM Rational Application Developer IDE using JSF web framework and hibernate Middleware, adopts Oracle 10g release 2 as DBMS and IBM Web sphere as application server. CNBAB has 600,000 lines code approximately.

Seal-stamping control based on proposed digital signature scheme using self-certified public key is an ActiveX control on client which is available in IE browser. It is developed in C++ language based on Microsoft Visual Studio 2008 IDE. It provides JavaScript interface functions, the most important two functions are sign and verify. Internal specific cryptography operations of the two functions are described in "Signature generation and verification" section. Sign function executes seal-stamping operation. Verify function verifies the validity of public





key and signature, but the verification of the public key is accomplished within the signature verification Procedure.

As compared with seal-stamping control designed based on the certificate-based public key [5][6], this control is more efficient for generating and verifying signatures in terms of computational efforts and communication costs.

After Users/agent staff logins CNBAB, there appears a web page include many menu items according to their individual rights. And there is a session bean storing user information, including key-information. There is a table recording username and corresponding public key in database.

When user views a contract, user/agent staff views a bill by clicking a menu item on web page, corresponding functional page is opened. According to the status of bill/contract and the privileges of user/agent staff, web page backend business logic judges whether there is a seal-stamping button on the page. If web page contains seal-stampings, control will verify the validity of every signature, then valid seal image is showed if passing verify, otherwise invalid seal image.

## 2.2 Web page Seal-Stamping

There is a processing step on server side before corresponding functional page is opened. Entire bill/contract page's html data is converted into XML data, stored as a property of page bean.

If bill/contract need to be seal-stamped by user/agent staff, a seal-stamping control and a seal-stamping button are inserted in the right position of the page. User/agent staff can trigger the seal-stamping button. After this button being triggered, control executes following steps accomplished on client.

(1) Examines whether there is a valid USB key on computer USB interface. If yes, require user /agent staff input USB key PIN; if no, prompt user to insert USB key.

(2) Examines whether this USB key is owned by login person according to public key information.

(3) Reads seal image in USB key, then sign organized XML data (mentioned at the beginning of this section) using private key in USB key, then seal image is inserted into the web page and floats above the web page automatically. Signature data include the signature value of organized XML data, seal image and public

key. At the same time, signature data is assigned to a hidden html element in bill/contract web page whose value is corresponding to a property of page bean. The maximum size of signature data is 15K, commonly 4K.

Thus, seal-stamping finished. After saving bill/contract, organized XML data and signature data is saved into database and the status of bill/contract is updated. When agent staff needs to audit bill, only if all seal image is valid, there is a seal-stamping button in the right position of page.

## 2.3 Verify

When user/agent staff views seal-stamped web page, all seal-stamping controls execute verify operation. There is a processing step on server side before corresponding functional page is opened. Data before signature and after signature for every seal-stamping must be retrieved from database to verify the validity of signature, stored as two property of page bean. According to CNBAB SRS [7], at most there are three seal stamps in a web page, commonly two.

If signature passes verify, controls in web page show valid Seal image, otherwise invalid seal image.

## 2.4 Digital Signature

A digital signature or digital signature scheme is a mathematical scheme for demonstrating the authenticity of a digital message or document. A valid digital signature gives a recipient reason to believe that the message was created by a known sender, and that it was not altered in transit. Digital signatures are commonly used for software distribution, financial transactions, and in other cases where it is important to detect forgery or tampering.

Digital signatures are often used to implement electronic signatures, a broader term that refers to any electronic data that carries the intent of a signature, but not all electronic signatures use digital signatures. In some countries, including the United States, India, and members of the European Union, electronic signatures have legal significance. However, laws concerning electronic signatures do not always make clear whether they are digital cryptographic signatures in the sense used here, leaving the legal definition, and so their importance, somewhat confused.

Digital signatures employ a type of asymmetric cryptography. For messages sent through a nonsecure channel, a properly implemented digital signature gives the receiver reason to believe the message was sent by the claimed sender. Digital signatures are equivalent to traditional handwritten signatures in many respects; properly implemented digital signatures are more difficult to forge than the handwritten type. Digital signature schemes in the sense used here are cryptographically based, and must be implemented properly to be effective. Digital signatures can also provide non-repudiation, meaning that the signer cannot successfully claim they did not sign a message, while also claiming their private key remains secret; further, some non-repudiation schemes offer a time stamp for the digital signature, so that even if the private key is exposed, the signature is valid nonetheless. Digitally signed messages may be anything representable as a bit string: examples include electronic mail, contracts, or a message sent via some other cryptographic protocol.

As organizations move away from paper documents with ink signatures or authenticity stamps, digital signatures can provide added assurances of the evidence to provenance, identity, and status of an electronic document as well as acknowledging informed consent and approval by a signatory. The United States Government Printing Office (GPO) publishes electronic versions of the budget, public and private laws, and congressional bills with digital signatures. Universities including Penn State, University of Chicago, and Stanford are publishing electronic student transcripts with digital signatures.

Below are some common reasons for applying a digital signature to communications:

**Authentication:** Although messages may often include information about the entity sending a message, that information may not be accurate. Digital signatures can be used to authenticate the source of messages. When ownership of a digital signature secret key is bound to a specific user, a valid signature shows that the message was sent by that user. The importance of high confidence in sender authenticity is especially obvious in a financial context. For example, suppose a bank's branch office sends instructions to the central office requesting a change in the balance of an account. If the central office is not convinced that such a message is truly sent from an authorized source, acting on such a request could be a grave mistake.

**Integrity:** In many scenarios, the sender and receiver of a message may have a need for confidence that the message has not been altered during transmission. Although encryption hides the contents of a message, it may be possible to change an encrypted message without understanding it. (Some encryption algorithms, known as nonmalleable ones, prevent this, but others do not.) However, if a message is digitally signed, any change in the message after signature will invalidate the signature. Furthermore, there is no efficient way to modify a message and its signature to produce a new message with a valid signature, because this is still considered to be computationally infeasible by most cryptographic hash functions

**Non-repudiation:** Non-repudiation, or more specifically non-repudiation of origin, is an important aspect of digital signatures. By this property an entity that has signed some information cannot at a later time deny having signed it. Similarly, access to the public key only does not enable a fraudulent party to fake a valid signature.

## 2.5 Group Signature

Based on digital signature scheme, we develop an ActiveX control on client to accomplish seal-stamping and verify. As compared with seal-stamping control designed based on the certificate-based public key [8][9], this control is more efficient for generating and verifying signatures in terms of computational efforts and communication costs. Further, we propose an electronic seal stamping based on Group signature which overcomes the disadvantages and retains all merits of the original scheme.

Group signatures allow individual members to make signatures on behalf of the group while providing, all previously proposed schemes are not very efficient and are also not to secure.

Group signatures allow individual members to make signatures on behalf of the group. Group oriented signature is a method to distribute the ability to sign among a set of users in such a way that only certain subsets of a group of users can collaborate to produce a valid signature on any given message. A group signature scheme has the following three properties

- (1) Only legal member of the group can sign messages.
- (2) The receiver can verify that it is indeed a valid group signature, but cannot discover which group member made it.

(3) In the case of a later dispute, the signer can be identified by either the group members together or a group authority.

Group signature scheme with signature claiming and variable linkability is a digital signature scheme with three types of participants: A group manager, an open authority, and group members. It consists of the following procedures:

- Setup: For a given security parameters, the group manager produce system-wide public parameters and a group manager master key for group membership certificate generation.
- Join: An interactive protocol between a user and the group manager. The user obtains a group membership certificate to become a group member. The public certificate and the user's identity information are stored by the group manager in a database for future use.
- Sign: Using his group membership certificate and his private key, a group member creates an anonymous group signature for a message.
- Verify: A signature is verified to make sure it originates from a legitimate group member without the knowledge of which particular one.
- Open: Given a valid signature, an open authority discloses the underlying group membership certificate.
- Claim (Self-trace): A group member creates a proof that he created a particular signature.
- Claim Verify: A party verifies the correctness of the claiming transcript. Similar to a group signature, our signature scheme should satisfy the following properties:
  - Correctness: Any valid signature can be correctly verified by the Verify protocol and a valid claiming proof can be correctly verified.
  - Forgery-Resistance: A valid group membership certificate can only be created by a user and the group manager through Join protocol.
  - Anonymity: It is infeasible to identify the real signer of a signature except by the open authority or if the signature has been claimed.

- Unlinkability: It is infeasible to link two different signatures of the same group member.
- Non-framing: No one (including the group manager) can sign a message in such a way that it appears to come from another user if it is opened.
- Non-appropriation: No one (including the group manager) can make a valid claim for signature which they did not create.

### 3. Proposed Signature Scheme using Self-Certified Public keys

#### 3.1 System Model

In the system environments, there exists a DUC (Digital Authentication Centre). The responsibilities of digital authentication centre are to generate the system parameters and to issue users' public keys. Stages of the proposed signature scheme include the system setup, the registration, the signature generation and verification.

In the system setup stage, digital authentication centre generates system parameters, including digital authentication centre's private key and public key pair. In the registration stage, digital authentication centre deals with the registration requests submitted by a registering user for issuing self certified public keys. After that, digital authentication centre publishes all self-certified public keys and sends each user a witness. Note that digital authentication centre does not need to generate any certificates for these public keys. With the received witness and the secret shadow, each user can solely compute his private key.

Moreover, each user could directly verify the validity of his self-certified public key with his private key, which demands on any additional public key Certificate. It should be assured that digital authentication centre does not have any useful knowledge of any user's private key. Note that the validity of signature and the authenticity of the signer have self-certified public key can be simultaneously verified in the signature verification.

#### 3.2 Realization of the Proposed Scheme

Following the system model as mentioned in the previous section, we propose a signature scheme using self-certified public keys in this section. The system setup, the registration, the signature generation and verification are described below in detail.

### 3.2.1 System Setup

Initially, digital authentication centre chooses a one-way hash function  $h$ , a large primes  $p$  such that  $p-1$  has also a large prime factor (e.g.  $(p-1)/2$ ) and a generator  $g$  of  $Z_p^*$ .

Then digital authentication centre randomly selects an integer  $a$  ( $a \in [1, p-2]$ ) and computes

$$b = g^a \text{ mod } p \quad (1)$$

The parameters  $b, g, p$  are published by digital authentication centre while  $a$  is kept secret.

### 3.2.2 Registration

When a user  $U_i$  with identity  $ID_i$  wants to join the system, the procedure for generating self-certified private-key/public-key pair is described below.

Step 1:  $U_i$  chooses a random integer  $j$  in  $Z_p^*$  ( $j \in [1, p-2]$ ),  $j$  is co-prime with  $p-1$ , computes

$$u = g^j \text{ mod } p \quad (2)$$

and  $U_i$  sends  $\{ID_i, u\}$  to digital authentication center for registration. Then he proves to digital authentication center that she knows  $j$  without revealing it by using an interactive zero knowledge proof.

Step 2: Upon receiving  $\{ID, u_i\}$  digital authentication center selects a random integer  $k$ , computes the public key for  $U_i$  as

$$P_i = u_i^k \text{ mod } p \quad (3)$$

and solves  $x$  in the equation using extended Euclidean algorithm

$$aP_i + kx = ID_i \text{ mod } (p-1) \quad (4)$$

Step 3: Digital authentication centre returns  $(P_i, ID_i, x)$  to  $U_i$ , who calculates:

$$s_i = xj^{-1} \text{ mod } (p-1) \quad (5)$$

So that,

$$b^{P_i} P_i^{s_i} = g^{ID_i} \text{ mod } p \quad (6)$$

$U_i$ 's secret key is  $s_i$  and self-certified public key is  $P_i$ .  $U_i$  computes solely his private key, so level 3 [10] is reached.  $U_i$  can check the validity of  $P_i$  by verifying (6). The correctness of the verification for the self-certified public key is shown through the following theorems.

Theorem 1: The self-certified public key  $P_i$  is valid provided that (6) holds.

Proof: Substituting  $ID_i$  with (4), we can rewrite (6) as

$$b^{P_i} P_i^{s_i} = g^{(aP_i + kx)} \text{ mod } p \quad (7)$$

combining (5), (1), (2), (3), we can infer (6).

If  $U_i$  wants to prove his identity to some verifier, he can perform the following procedure:

Step 1:  $U_i$  sends  $\{ID_i, P_i\}$  to the verifier, who computes

$$v_i b^{-P_i} g^{ID_i} \text{ mod } p \quad (8)$$

Step 2:  $U_i$  selects a random integer  $r_i$  in  $Z_p^*$  computes

$$t_i = P_i^{r_i} \text{ mod } p \quad (9)$$

and sends  $t_i$  to the verifier.

Step 3: The verifier randomly selects an integer  $k$  in  $Z_p^*$  and sends it to  $U_i$ .

Step 4:  $U_i$  computes

$$x_i = r_i + s_i k \quad (10)$$

Step 5: The verifier checks the following verification equation:

$$P_i^{x_i} = t_i v_i^k \text{ (mod } p) \quad (11)$$

If it holds, then the verifier accepts the validity of the identity of  $U_i$ , otherwise rejects the identity claimed by  $U_i$ .

Note that no additional certificate is required when verifying the validity of the identity of  $U_i$ , since  $P_i$  is self-certified. Except for  $U_i$ , another user cannot infer  $s_i$  from  $P_i$  and all available public information, under the cryptographic assumptions that the discrete logarithm problems are hard [5].

Also note that digital authentication centre might impersonate  $U_i$  by randomly choosing a random integer  $j'$ , computing public key  $P_i'$  and private key  $s_i'$  by (2), (3), (4), and (5). The forged public key  $P_i'$  will pass the verification check in (6).

However, the existence of two valid public keys linked to  $U_i$  gives the proof that digital authentication centre is dishonest.

## 4. Signature Generation & Verification

### 4.1 Signature Generation

Let  $M$  be the signing message. To generate the signature for  $M$ , each user  $U_i$  performs the following procedure:

$U_i$  first chooses an random integer  $w_i$  in  $Z_p^*$  and then computes the signature for  $M$ , i.e.,  $(r_i, x_i)$  where

$$r_i = P_i^{w_i} \bmod p \quad (12)$$

$$x_i = w_i + s_i h(M, r_i) \quad (13)$$

### 4.2 Signature Verification

Upon receiving  $M$  and its signature  $(r_i, x_i)$ , the verifier checks the following signature verification equation:

$$P_i^{x_i} = r_i (b^{-P_i} g^{ID_i})^{h(M, r_i)} \pmod{p} \quad (14)$$

If it holds, then the verifier accepts the validity of the signature, otherwise rejects the signature

Theorem 2: If (13) holds, then the signature of  $M$  is verified, and meanwhile, the public key of  $U_i$  is authenticated.

Proof: Raising both sides of (12) to exponents with the base  $P_i$  yields

$$P_i^{x_i} = P_i^{w_i} \cdot P_i^{s_i h(M, r_i)} \pmod{p} \quad (15)$$

Thus,  $(r_i, x_i)$  are verified if  $P_i$  is authenticated

### 4.3 Group Signature Generation & verification

If all individual signatures are verified, then CLK computes

$$R = \prod_{i=1}^t r_i \bmod p \quad (16)$$

$$S = \sum_{i=1}^t s_i \bmod q \quad (17)$$

Thus  $(R, S)$  is the group signature of  $M$  with respect to  $G$ . To verify the group signature, any verifier checks the following equality:

$$g^S = R^{h(m|R)} ((Y_G + h(GID)) \beta^{h(Y_G | GID)})^R \bmod p \quad (18)$$

If it holds, then  $(R, S)$  is a valid group signature of  $M$  signed by  $G$  with the self certified public key  $Y_G$  [11], [12], [13].

## 5. USB Key

USB Key is a smart hardware of USB interface within CPU, memory and chip operating systems (COS) inside. It is used to store user's self certified private key/ public key pair and watermarked seal image. The procedure for generating self certified private-key/public-key pair is described in "REGISTRATION". User/Agent staff seal is scanned into computer to seal image. After Hollow processing, semitransparent processing, Gray Processing, Seal image is returned into USB key at the same time seal image is watermarked using User's private key. Inside USB key, there are algorithms to verify private-key/public-key pair and watermark seal image.

Each USB key has PIN protection [14]. Since PIN is input on the computer, then the attacker may get PIN by program. If the user does not take USB key in time, the attacker may pass the fake authentication through having gotten PIN. So there is dynamic password algorithm inside USB key to work out frequently changed, unpredictable and one time valid password, so that PIN may be produced dynamically. Even if the attacker can get the last PIN, it has been already disposable. Time stamp can be implemented with the USB key. This can be considered as future work.

## 6. Conclusions

In this paper, we present a group signature scheme using self certified key. Electronic commerce, commonly known as e-commerce or ecommerce, consists of the buying and selling of products or services over electronic systems such as the Internet and other computer networks. Bill/Contract in E-commerce must be seal stamped to gain tamper proof and non-repudiation. Non-repudiation refers to a state of affairs where the purported maker of a statement will not be able to successfully challenge the validity of the statement or contract. The term is often seen in a legal setting wherein the authenticity of a signature is being challenged. In such an instance the authenticity is being "repudiated". The seal stamping control is designed based on the certificate-based public key. This technique is more efficient for generating and verifying individual/group signatures in terms of computational efforts and communication costs. The security of the proposed scheme is based on the hash function.



## References

- [1] Shijin Yaun, Bin Mu and Xianing Zhang, "Implementation for electronic seal-stamping using self certified public key in e-commerce", Internet technology and application, wuhan, 20-22, Aug, 2010.
- [2] <http://www.cnbab.com/>, last visited on 28th, April. 2011.
- [3] Bin Mu, Shijin Yaun, "software analyze and design for resources operations and its supporting technologies project", unpublished
- [4] Girault M, " self -certified public keys", In Advances in cryptology- EUROCRPYT'91, springer-verlag, Berlin, pp 491-497,1991.
- [5] Shahrokh Saeednia, "A short note on Girault's self certified model", <http://eprint.iacr.org/2001/100.ps.gz>, 2001.
- [6] Tzong-Sun Wu, Chien-Lung Hsu, "Threshold signature scheme using sel certified public keys", The journal of systems, Vol.67, pp.89-97, 2003
- [7] Li Guo, zang Jinmei, "Realization of electronic official seal system based on WORD", Proceedings-International conference on Networks Security, Wireless Communications and Trusted Computing, NSWCTC 2009,v 1, p 501-504, 2009
- [8] <http://www.bjca.org.cn/> last visited on 28th, Mar., 2011
- [9] Cheng Zhen-bo, Xiao Gang Zhang Fei," Design and Implementaion on digital stamp system for public document", Journal of Zhejiang University of Technology, Volume 36 issue 5, 2008
- [10] Shi-yuan Zheng; Jun Liu "An USB-Key\_based approach for software tamper resistance", Advanced Computer Theory and Engineering (ICACTE), 2010 3rd International Conference on 20-22 Aug. 2010.
- [11] Ueda, K. Mutoh, T. Matsuo, K. Dept. of Inf. & Computer. Eng., Nara Nat. Coll. of Technol. "Automatic verification system for seal imprints on Japanese bankchecks ", Pattern Recognition, 1998. Proceedings. Fourteenth International Conference on IssueDate: 16-20, Aug, 1998; Volume: 1, On page(s): 629 - 632 vol.1
- [12] Jianhong Zhang, Qin Geng;North China University of Technology;Beijing." On the Security of a Group Signature Scheme", networking, Sensing & Control, 2008. ICNSC 2008. IEEE International Conference,6-8 April 2008, Pages:1310-1314
- [13] Popescu, C.; Noje, D.; Bede, B.; Mang, I.; Dept. of Math., Univ. of Oradea, Romania ,"A group signature scheme with revocation" Video/Image Processing and Multimedia Communications, 2003.
- 4th EURASIP Conference, Issue Date: 2-5 July 2003  
On page(s): 245 - 250 Vol.1
- [14] Park, , Haeryong; Kim, Hyun; Chun, Kilsoo; Lee, Jaeil; Lim, Seongan; Yie, Ikkwon; Cryptography Technol. Team, Korea Inf. Security Agency, Seoul , " Untraceability of Group Signature Schemes based on Bilinear Mapping and Their Improvement" Information Technology, 2007. ITNG '07. Fourth International Conference on Issue Date: 2-4April2007, on page(s): 747 - 753

**Girija Srikanth** has completed her BE Degree in Electronics & Communication Engineering [2005], M.Tech degree in Computer Science & Engineering with the specialization of Information Security [2008] from Pondicherry Engineering College. Currently she is working as Lecturer, Department of Computer Science & Engineering, Birla Institute of Technology, Ras-Al-Kaimah, Dubai. She presented two papers in National Conference and participated in an International Conference [IACITS]. Her research interests include Cryptography, Image Processing, Steganography, Network Security and Web Security.



# Arithmetic and Frequency Filtering Methods of Pixel-Based Image Fusion Techniques

Mrs. Firouz Abdullah Al-Wassai<sup>1</sup>, Dr. N.V. Kalyankar<sup>2</sup>, Dr. Ali A. Al-Zuky<sup>3</sup>

<sup>1</sup> Research Student, Computer Science Dept., Yeshwant College, (SRTMU), Nanded, India

<sup>2</sup> Principal, Yeshwant Mahavidyala Colleg, Nanded, India

<sup>3</sup> Assistant Professor, Dept. of Physics, College of Science, Mustansiriyah Un. Baghdad – Iraq.

## Abstract

In remote sensing, image fusion technique is a useful tool used to fuse high spatial resolution panchromatic images (PAN) with lower spatial resolution multispectral images (MS) to create a high spatial resolution multispectral of image fusion (F) while preserving the spectral information in the multispectral image (MS). There are many PAN sharpening techniques or Pixel-Based image fusion techniques that have been developed to try to enhance the spatial resolution and the spectral property preservation of the MS. This paper attempts to undertake the study of image fusion, by using two types of pixel –based image fusion techniques i.e. Arithmetic Combination and Frequency Filtering Methods of Pixel-Based Image Fusion Techniques. The first type includes Brovey Transform (BT), Color Normalized Transformation (CN) and Multiplicative Method (MLT). The second type include High-Pass Filter Additive Method (HPFA), High – Frequency- Addition Method (HFA) High Frequency Modulation Method (HFM) and The Wavelet transform-based fusion method (WT). This paper also devotes to concentrate on the analytical techniques for evaluating the quality of image fusion (F) by using various methods including Standard Deviation (SD), Entropy(En), Correlation Coefficient (CC), Signal-to Noise Ratio (SNR), Normalization Root Mean Square Error (NRMSE) and Deviation Index (DI) to estimate the quality and degree of information improvement of a fused image quantitatively.

**Keywords:** *Image Fusion; Pixel-Based Fusion; Brovey Transform; Color Normalized; High-Pass Filter ; Modulation, Wavelet transform.*

## 1. INTRODUCTION

Although Satellites remote sensing image fusion has been a hot research topic of remote sensing image processing [1]. This is obvious from the amount of conferences and workshops focusing on data fusion, as well as the special issues of scientific journals dedicated to the topic. Previously, data fusion, and in

particular image fusion belonged to the world of research and development. In the meantime, it has become a valuable technique for data enhancement in many applications. More and more data providers envisage the marketing of fused products. Software vendors started to offer pre-defined fusion methods within their generic image processing packages [2].

Remote sensing offers a wide variety of image data with different characteristics in terms of temporal, spatial, radiometric and Spectral resolutions. Although the information content of these images might be partially overlapping [3], imaging systems somehow offer a tradeoff between high spatial and high spectral resolution, whereas no single system offers both. Hence, in the remote sensing community, an image with ‘greater quality’ often means higher spatial or higher spectral resolution, which can only be obtained by more advanced sensors [4]. However, many applications of satellite images require both spectral and spatial resolution to be high. In order to automate the processing of these satellite images new concepts for sensor fusion are needed. It is, therefore, necessary and very useful to be able to merge images with higher spectral information and higher spatial information [5].

The term “fusion” gets several words to appear, such as merging, combination, synergy, integration ... and several others that express more or less the same concept have since appeared in literature [6]. Different definitions of data fusion can be found in literature, each author interprets this term differently depending his research interests, such as [7-8] . A general definition of data fusion can be adopted as following “Data fusion is a formal framework which expresses means and tools for the alliance of data originating from different sources. It aims at obtaining information of greater quality; the exact definition of ‘greater quality’ will depend upon the

application” [11-13]. Image fusion forms a subgroup within this definition and aims at the generation of a single image from multiple image data for the extraction of information of higher quality. Having that in mind, the achievement of high spatial resolution, while maintaining the provided spectral resolution, falls exactly into this framework [14].

## 2. Pixel-Based Image Fusion Techniques

Image fusion is a sub area of the more general topic of data fusion [15]. Generally, Image fusion techniques can be classified into three categories depending on the stage at which fusion takes place; it is often divided into three levels, namely: pixel level, feature level and decision level of representation [16, 17]. This paper will focus on pixel level image fusion. The pixel image fusion techniques can be grouped into several techniques depending on the tools or the processing methods for image fusion procedure. It is grouped into four classes: 1) Arithmetic Combination techniques (AC) 2) Component Substitution fusion techniques (CS) 3) Frequency Filtering Methods (FFM) 4) Statistical Methods (SM). This paper focuses on using two types of pixel-based image fusion techniques Arithmetic Combination and Frequency Filtering Methods of Pixel-Based Image Fusion Techniques. The first type is included BT; CN; MLT and the last type includes HPFA; HFA HFM and WT. In this work to achieve the fusion algorithm and estimate the quality and degree of information improvement of a fused image quantitatively used programming in VB.

To explain the algorithms through this report, Pixels should have the same spatial resolution from two different sources that are manipulated to obtain the resultant image. So, before fusing two sources at a pixel level, it is necessary to perform a geometric registration and a radiometric adjustment of the images to one another. When images are obtained from sensors of different satellites as in the case of fusion of SPOT or IRS with Landsat, the registration accuracy is very important. But registration is not much of a problem with simultaneously acquired images as in the case of Ikonos/Quickbird PAN and MS images. The PAN images have a different spatial resolution from that of MS images. Therefore, resampling of MS images to the spatial resolution of PAN is an essential step in some fusion methods to bring the MS images to the same size of PAN, , thus the resampled MS images will be noted by  $M_k$  that represents the set of DN of band k in the resampled MS image. Also the following notations will be used:

P as DN for PAN image,  $F_k$  the DN in final fusion result for band k.  $\bar{M}_k$ ,  $\bar{P}$ ,  $\sigma_P$ ,  $\sigma_{M_k}$  Denotes the local means and standard deviation calculated inside the window of size (3, 3) for  $M_k$  and P respectively.

## 3. The AC Methods

This category includes simple arithmetic techniques. Different arithmetic combinations have been employed for fusing MS and PAN images. They directly perform some type of arithmetic operation on the MS and PAN bands such as addition, multiplication, normalized division, ratios and subtraction which have been combined in different ways to achieve a better fusion effect. These models assume that there is high correlation between the PAN and each of the MS bands [24]. Some of the popular AC methods for pan sharpening are the BT, CN and MLM. The algorithms are described in the following sections.

### 3.1 Brovey Transform (BT)

The BT, named after its author, uses ratios to sharpen the MS image in this method [18]. It was created to produce RGB images, and therefore only three bands at a time can be merged [19]. Many researchers used the BT to fuse a RGB image with a high resolution image [20-25]. The basic procedure of the BT first multiplies each MS band by the high resolution PAN band, and then divides each product by the sum of the MS bands. The following equation, given by [18], gives the mathematical formula for the BT:

$$F_{k(i,j)} = \frac{M_{k(i,j)} \times P(i,j)}{\sum_k M_{k(i,j)}} \quad (1)$$

The BT may cause color distortion if the spectral range of the intensity image is different from the spectral range covered by the MS bands.

### 3.2 Color Normalized Transformation (CN)

CN is an extension of the BT [17]. CN transform also referred to as an energy subdivision transform [26]. The CN transform separates the spectral space into hue and brightness components. The transform multiplies each of the MS bands by the p imagery, and these resulting values are each normalized by being divided by the sum of the MS bands. The CN transform is defined by the following equation [26, 27]:

$$F_{k(i,j)} = \frac{(M_{k(i,j)}+1.0)(P(i,j)+1.0) \times 3.0}{\sum_k M_{k(i,j)}+3.0} - 1.0 \quad (2)$$

(Note: The small additive constants in the equation are included to avoid division by zero.)

### 3.3 Multiplicative Method (MLT)

The Multiplicative model or the product fusion method combines two data sets by multiplying each pixel in each band  $k$  of MS data by the corresponding pixel of the PAN data. To compensate for the increased brightness, the square root of the mixed data set is taken. The square root of the Multiplicative data set, reduce the data to combination reflecting the mixed spectral properties of both sets. The fusion algorithm formula is as follows [1; 19; 20]:

$$F_{k(i,j)} = \sqrt{M_{k(i,j)} \times P_{(i,j)}} \quad (3)$$

## 4. Frequency Filtering Methods (FFM)

Many authors have found fusion methods in the spatial domain (high frequency inserting procedures) superior over the other approaches, which are known to deliver fusion results that are spectrally distorted to some degree [28] Examples of those authors are [29-31].

Fusion techniques in this group use high pass filters, Fourier transform or wavelet transform to model the frequency components between the PAN and MS images by injecting spatial details in the PAN and introducing them into the MS image. Therefore, the original spectral information of the MS channels is not or only minimally affected [32]. Such algorithms make use of classical filter techniques in the spatial domain. Some of the popular FFM for pan sharpening are the HPF, HFA, HFM and the WT based methods.

### 4.1 High-Pass Filter Additive Method (HPFA)

The High-Pass Filter Additive (HPFA) technique [28] was first introduced by Schowengerdt (1980) as a method to reduce data quantity and increase spatial resolution for Landsat MSS data [33]. HPF basically consists of an addition of spatial details, taken from a high-resolution Pan observation, into the low resolution MS image [34]. The high frequencies information is computed by filtering the PAN with a high-pass filter through a simple local pixel averaging, i.e. box filters. It is performed by emphasize the detailed high frequency components of an image and deemphasize the more general low frequency information [35]. The HPF method uses standard

square box HP filters. For example, a 3\*3 pixel kernel given by [36], which is used in this study:

$$P_{HPF} = \frac{1}{9} \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix} \quad (4)$$

In its simplest form, The HP filter matrix is occupied by “-1” at all but at the center location. The center value is derived by  $c = n * n - 1$ , where  $c$  is the center value and  $n * n$  is the size of the filter box [28]. The HP are filters that compute a local average around each pixel in the PAN image.

The extracted high frequency components of  $P_{HPF}$  superimposed on the MS image [1] by simple addition and the result divided by two to offset the increase in brightness values [33]. This technique can improve spatial resolution for either colour composites or an individual band [16]. This is given by [33]:

$$F_k = \frac{(M_k + P_{HPF})}{2} \quad (5)$$

The high frequency is introduced equally without taking into account the relationship between the MS and PAN images. So the HPF alone will accentuate edges in the result but loses a large portion of the information by filtering out the low spatial frequency components [37].

### 4.2 High –Frequency- Addition Method (HFA)

High-frequency-addition method [32] is a technique of filter techniques in spatial domain similar the previous technique, but the difference between them is the way how to extract the high frequencies. In this method, to extract the PAN channel high frequencies; a degraded or low-pass-filtered version of the panchromatic channel has to be created by applying the following set of filter weights (in a 3 x 3 convolution filter example) [38]:

$$P_{LFF} = \frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad (6)$$

A low pass or smoothing filter, which corresponds to computing a local average around each pixel in the image, is achieved. Since the goal of contrast enhancement is to increase the visibility of small detail in an image, subsequently, the high frequency addition method (HFA) extracts the high frequencies using a subtraction procedure. This approach is known as Unsharp masking USM [39]:

$$P_{USM} = P - P_{LFF} \quad (7)$$

Some authors, for example [40]; defined USM as HPF; while [36, 41] multiply the original image by an implication factor, denoted by  $a$ , and hence define it as a High Boost Filter (HBF) or high-frequency-emphasis filter: in the original, that is:

$$HBF = a * P - P_{LPF} \quad (8)$$

The general process by using equation (8) called unsharp masking [36] and adds them to the MS channels via addition as shown by equation [32]:

$$F_k = M_k + P_{USM} \quad (9)$$

When this technique is applied, it really leads to the enhancement of all high spatial frequency detail in an image including edges, line and points of high gradient [42]

### 4.3 High Frequency Modulation Method (HFM)

The problem of the addition operation is that the introduced texture will be of different size relative to each multispectral channel, so a channel wise scaling factor for the high frequencies is needed. The alternative high frequency modulation method HFM extracts the high frequencies via division for the P on the PAN channel low frequency  $P_{LPF}$  which is obtained by using equation (9) to extract the PAN channel low-frequency  $P_{LPF}$  and then adds them to each multispectral channel via multiplication [32]:

$$F_k = M_k \times \frac{P}{P_{LPF}} \quad (10)$$

Because of the multiplication operation, every multispectral channel is modulated by the same high frequencies [32].

### 4.4 Wavelet Transformation (WT) Based Image Fusion

Wavelet-based methods Multi-resolution or multi-scale methods [24] is a mathematical tool developed in the field of signal processing [9] have been adopted for data fusion since the early 1980s (MALAT, 1989). Recently, the wavelet transform approach has been used for fusing data and becomes hot topic in research [43]. The wavelet transform provides a framework to decompose (also called analysis) images into a number of new images, each one of them with a different degree of resolution as well as a perfect reconstruction of the signal (also called synthesis). Wavelet-based approaches show some favorable properties compared to the Fourier transform [44]. While the Fourier transform gives an idea of the frequency content in the image, the wavelet representation is an intermediate representation between the Fourier and the spatial representation, and it can provide good localization in both frequency and space domains [45]. Furthermore, the multi-resolution nature of the wavelet transforms allows for control of fusion quality by controlling the number of resolutions [46] as will as the wavelet transform does not operate on color images directly so we have transformed the color image from RGB domain to another domain [47].

For more information about image fusion based on wavelet transform have been published in recent years [48 -50].

The block diagram of a generic wavelet-based image fusion scheme is shown in Fig. 3. Wavelet transform based image fusion involves three steps; forward transform coefficient combination and backward transform. In the forward transform, two or more registered input images are wavelet transformed to get their wavelet coefficients [51]. The wavelet coefficients for each level contain the spatial (detail) differences between two successive resolution levels [9].

The basic operation for calculating the DWT is convolving the samples of the input with the low-pass and high-pass filters of the wavelet and down sampling the output [52]. Wavelet transform based image fusion involves various steps:

Step (1): the PAN image  $P$  is first reference stretched three times, each time to match one of multispectral  $M_k$  histograms to produce three new PAN images.

Step (2): the wavelet basis for the transform is chosen. In this study the upper procedure is for one level wavelet decomposition, and we used to implement the image fusion using wavelet basis of Haar because it is found that the choice of the wavelet basis does affect the fused images [53]. The Haar basis vectors are simple [37]:

$$L = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \quad (10)$$

Then performing the wavelet decomposition analysis to extract The structures or "details" present between the images of two different resolution. These structures are isolated into three wavelet coefficients, which correspond to the detailed images according to the three directions. The decomposition at first level we will have one approximation coefficients, ( $A^N_{R,G,B}$ ) and 3N wavelets Planes for each band by the following equation [54]:

$$\begin{aligned} R &\xrightarrow{WT} A^N_R + \sum_l^N (H^l_R + V^l_R + D^l_R) \\ G &\xrightarrow{WT} A^N_G + \sum_l^N (H^l_G + V^l_G + D^l_G) \\ B &\xrightarrow{WT} A^N_B + \sum_l^N (H^l_B + V^l_B + D^l_B) \end{aligned} \quad (11)$$

$A^N$ : is Approximation coefficient at level N or approximation plane

-  $H^l$  : is Horizontal coefficient at level l or horizontal wavelet plane

-  $V^l$  : is Vertical Coefficient at level l or vertical wavelet plane

-  $D^l$ : is Diagonal coefficient at level l or diagonal wavelet plane

Step (3): Similarly by decomposing the panchromatic high-resolution image we will have one approximation coefficients, ( $A_P^N$ ) and  $3N$  wavelets Planes for Panchromatic image, where  $PAN$  means, panchromatic image.

Step (4): the wavelet coefficients sets from two images are combined via substitutive or additive rules. In the case of substitutive method, the wavelet coefficient planes (or details) of the  $R$ ,  $G$ , and  $B$  decompositions are replaced by the similar detail planes of the panchromatic decomposition, which that used in this study.

Step (5): Then, for obtaining the fused images, the inverse wavelet transform is implemented on resultant sets. By reversing the process in step (2) the synthesis is equation [54]:

$$\begin{aligned} A_R^N + \sum_l^N (H_P^l + V_P^l + D_P^l) &\xrightarrow{IWT} R_{new} \\ A_G^N + \sum_l^N (H_P^l + V_P^l + D_P^l) &\xrightarrow{IWT} G_{new} \\ A_B^N + \sum_l^N (H_P^l + V_P^l + D_P^l) &\xrightarrow{IWT} B_{new} \end{aligned} \quad (12)$$

Wavelet transform fusion is obtained. This reverse process is referred to as reconstruction of the image in which the finer representation is calculated from coarser levels by adding the details according to the synthesis equation [44]. Thus at high resolution, simulated are produced.

## 5. Experiments

In order to validate the theoretical analysis, the performance of the methods discussed above was further evaluated by experimentation. Data sets used for this study were collected by the Indian IRS-1C PAN (0.50 - 0.75  $\mu\text{m}$ ) of the 5.8- m resolution panchromatic band. Where the American Landsat (TM) the red (0.63 - 0.69  $\mu\text{m}$ ), green (0.52 - 0.60  $\mu\text{m}$ ) and blue (0.45 - 0.52  $\mu\text{m}$ ) bands of the 30 m resolution multispectral image were used in this experiment. Fig. 3 shows the IRS-1C PAN and multispectral TM images. The scenes covered the same area of the Mausoleums of the Chinese Tang – Dynasty in the PR China [55] was selected as test sit in this study. Since this study is involved in evaluation of the effect of the various spatial, radiometric and spectral resolution for image fusion, an area contains both manmade and natural features is essential to study these effects. Hence, this work is an attempt to study the quality of the images fused from different sensors with various characteristics. The size of the PAN is 600 \* 525 pixels at 6 bits per pixel and the size of the original multispectral is 120

\* 105 pixels at 8 bits per pixel, but this is upsampled to by Nearest neighbor was used to avoid spectral contamination caused by interpolation.

To evaluate the ability of enhancing spatial details and preserving spectral information, some Indices including Standard Deviation (SD), Entropy(En), Correlation Coefficient (CC), Signal-to Noise Ratio (SNR), Normalization Root Mean Square Error (NRMSE) and Deviation Index (DI) of the image were used (Table 1), and the results are shown in Table 2. In the following sections,  $F_k$ ,  $M_k$  are the measurements of each the brightness values of homogenous pixels of the result image and the original multispectral image of band k,  $\bar{M}_k$  and  $\bar{F}_k$  are the mean brightness values of both images and are of size  $n * m$ .  $BV$  is the brightness value of image data  $\bar{M}_k$  and  $\bar{F}_k$ . To simplify the comparison of the different fusion methods, the values of the En, CC, SNR, NRMSE and DI index of the fused images are provided as chart in Fig. 1

Table 1: Indices Used to Assess Fusion Images.

Equation
$\sigma_k = \sqrt{\frac{\sum_{i=1}^m \sum_{j=1}^n (BV_k(n, m) - \mu_k)^2}{m \times n}}$
$CC_k = \frac{\sum_i^n \sum_j^m (F_k(i, j) - \bar{F}_k)(M_k(i, j) - \bar{M}_k)}{\sqrt{\sum_i^n \sum_j^m (F_k(i, j) - \bar{F}_k)^2} \sqrt{\sum_i^n \sum_j^m (M_k(i, j) - \bar{M}_k)^2}}$
$En = - \sum_0^{I-1} P(i) \log_2 P(i)$
$DI_k = \frac{1}{nm} \sum_i^n \sum_j^m \frac{ F_k(i, j) - M_k(i, j) }{M_k(i, j)}$
$SNR_k = \sqrt{\frac{\sum_i^n \sum_j^m (F_k(i, j))^2}{\sum_i^n \sum_j^m (F_k(i, j) - M_k(i, j))^2}}$
$NRMSE_k = \sqrt{\frac{1}{nm * 255^2} \sum_i^n \sum_j^m (F_k(i, j) - M_k(i, j))^2}$

## 6. Discussion Of Results

The Fig. 1 shows those parameters for the fused images using various methods. It can be seen that from fig.1a. The SD of the fused images remains constant for HFA and HFM. According to the computation results En, the increased En indicates the change in quantity of information content for radiometric resolution through the merging. From fig.1b, it is obvious that En of the fused images have been changed when compared to the original multispectral but some methods such as (BT and HPFA) decrease the En values to below the original. In Fig.1c. Correlation values also remain practically constant, very near the maximum possible value except BT and CN. The results of SNR, NRMSE and DI appear changing significantly. It can be observed, from the diagram of Fig. 1., that the results of NRMSE & DI, of the fused image, show that the HFM and HFA methods give the best results with respect to the other methods indicating that these methods maintain most of information spectral content of the original multispectral data set which get the same values presented the lowest value of the NRMSE & DI as well as the higher of the SNR. Hence, the spectral qualities of fused images by HFM and HFA methods are much better than the others. In contrast, it can also be noted that the BT, HPFA images produce highly NRMSE & DI values indicate that these methods deteriorate spectral information content for the reference image. In a comparison of spatial effects, it can be seen that the results of the HFM; HFA; WT and CN are better than other methods. Fig.3. shows the original images and the fused image results.

By combining the visual inspection results, it can be seen that the experimental results overall method are The HFM and HFA results which are the best result. The next higher the visual inspection results are obtained with WT, CN and MUL.

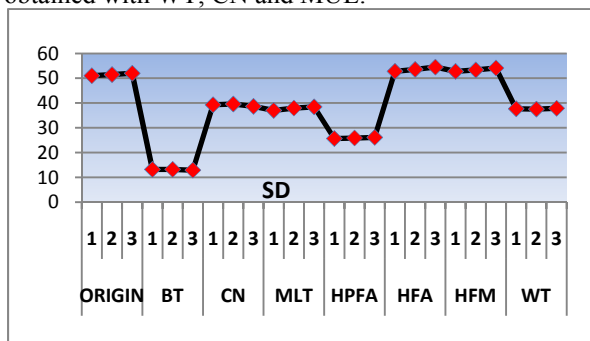


Fig. 1a: Chart Representation of SD of Fused Images

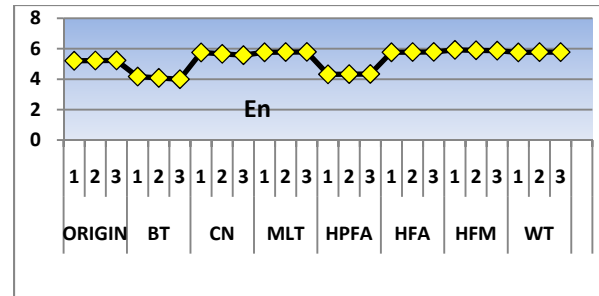


Fig. 1b: Chart Representation of En of Fused Images

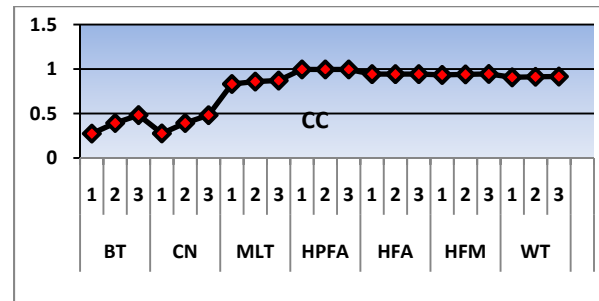


Fig. 1c: Chart Representation of CC of Fused Images

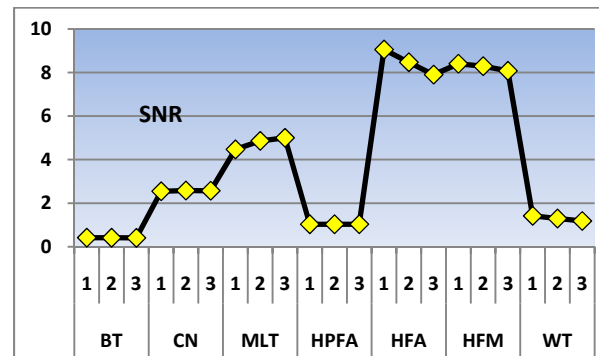


Fig. 1d: Chart Representation of SNR of Fused Images

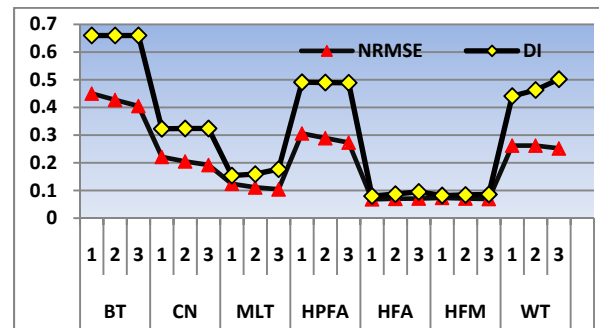


Fig. 1e: Chart Representation of NRMSE & DI of Fused Images

Fig. 1: Chart Representation of SD , En , CC ,NRMSE & DI of Fused Images





Fig.2a. Original Panchromatic Fig.2b.Original Multispectral



Fig. 2c. BT Fig.2d. CN



Fig. 2f. MUL



Fig. 2g. HPF



Fig. 2e. HFA



Fig. 2f. HFM



Fig. 2i. WT

**Table 2: Quantitative Analysis of Original MS and Fused Image Results Through the Different Methods**

Method	Band	SD	En	SNR	NRMSE	DI	CC
<b>ORIGIN</b>	1	51.018	5.2093				
	2	51.477	5.2263				
	3	51.983	5.2326				
<b>BT</b>	1	13.185	4.1707	0.416	0.45	0.66	0.274
	2	13.204	4.0821	0.413	0.427	0.66	0.393
	3	12.878	3.9963	0.406	0.405	0.66	0.482
<b>CN</b>	1	39.278	5.7552	2.547	0.221	0.323	0.276
	2	39.589	5.6629	2.579	0.205	0.324	0.393
	3	38.633	5.5767	2.57	0.192	0.324	0.481
<b>MLT</b>	1	37.009	5.7651	4.468	0.124	0.154	0.832
	2	37.949	5.7833	4.858	0.111	0.159	0.859
	3	38.444	5.7915	4.998	0.104	0.177	0.871
<b>HPFA</b>	1	25.667	4.3176	1.03	0.306	0.491	0.996
	2	25.869	4.3331	1.032	0.289	0.49	0.996
	3	26.121	4.3424	1.033	0.273	0.489	0.996
<b>HFA</b>	1	52.793	5.7651	9.05	0.068	0.08	0.943
	2	53.57	5.7833	8.466	0.07	0.087	0.943
	3	54.498	5.7915	7.9	0.071	0.095	0.943
<b>HFM</b>	1	52.76	5.9259	8.399	0.073	0.082	0.934
	2	53.343	5.8979	8.286	0.071	0.084	0.94
	3	54.136	5.8721	8.073	0.069	0.086	0.945
<b>WT</b>	1	37.666	5.7576	1.417	0.262	0.441	0.907
	2	37.554	5.7754	1.296	0.262	0.463	0.913
	3	37.875	5.7765	1.182	0.252	0.502	0.916

Fig.2: The Representation of original and Fused Images

## 6. Conclusion

Image Fusion aims at the integration of disparate and complementary data to enhance the information apparent in the images as well as to increase the reliability of the interpretation. This leads to more accurate data and increased utility in application fields like segmentation and classification. In this paper, the comparative studies undertaken by using two types of pixel-based image fusion techniques Arithmetic Combination and Frequency Filtering Methods of Pixel-Based Image Fusion Techniques as well as effectiveness based image fusion and the performance of these methods. The fusion procedures of the first type, which includes (BT; CN; MLT) by using all PAN band, produce more distortion of spectral characteristics because such methods depend on the degree of global correlation between the PAN and multispectral bands to be enhanced. Therefore, these fusion techniques are not adequate to preserve the spectral characteristics of original multispectral. But those methods enhance the spatial quality of the imagery except BT. The fusion procedures of the second type includes HPFA; HFA; HFM and the WT based fusion method by using selected (or Filtering) PAN band frequencies including HPF, HFA, HFM and WT algorithms. The preceding analysis shows that the HFA and HFM methods maintain the spectral integrity and enhance the spatial quality of the imagery. The HPF method does not maintain the spectral integrity and does not enhance the spatial quality of the imagery. The WTF method has been shown in many published papers as an efficient image fusion. In the present work, the WTF method has shown low results.

In general types of the data fusion techniques, the use of the HFM & HFA could, therefore, be strongly recommended if the goal of the merging is to achieve the best representation of the spectral information of multispectral image and the spatial details of a high-resolution panchromatic image.

## References

- [1] Wenbo W., Y. Jing, K. Tingjun, 2008. "Study Of Remote Sensing Image Fusion And Its Application In Image Classification" The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XXXVII. Part B7. Beijing 2008, pp.1141-1146.
- [2] Pohl C., H. Touron, 1999. "Operational Applications of Multi-Sensor Image Fusion". International Archives of Photogrammetry and Remote Sensing, Vol. 32, Part 7-4-3 w6, Valladolid, Spain.
- [3] Steinnocher K., 1999. "Adaptive Fusion Of Multisource Raster Data Applying Filter Techniques". International Archives of Photogrammetry and Remote Sensing, Vol. 32, Part 7-4-3 W6, Valladolid, Spain, 3-4 June, pp.108-115.
- [4] Dou W., Chen Y., Li W., Daniel Z. Sui, 2007. "A General Framework for Component Substitution Image Fusion: An Implementation Using the Fast Image Fusion Method". Computers & Geosciences 33 (2007), pp. 219-228.
- [5] Zhang Y., 2004. "Understanding Image Fusion". Photogrammetric Engineering & Remote Sensing, pp. 657-661.
- [6] Wald L., 1999a, "Some Terms Of Reference In Data Fusion". IEEE Transactions on Geosciences and Remote Sensing, 37, 3, pp.1190-1193.
- [7] Hall D. L. and Llinas J., 1997. "An introduction to multisensor data fusion," (invited paper) in Proceedings of the IEEE, Vol. 85, No 1, pp. 6-23.
- [8] Pohl C. and Van Genderen J. L., 1998. "Multisensor Image Fusion In Remote Sensing: Concepts, Methods And Applications". (Review Article), International Journal Of Remote Sensing, Vol. 19, No.5, pp. 823-854.
- [9] Zhang Y., 2002. "PERFORMANCE ANALYSIS OF IMAGE FUSION TECHNIQUES BY IMAGE". International Archives of Photogrammetry and Remote Sensing (IAPRS), Vol. 34, Part 4. Working Group IV/7.
- [11] Ranchin, T., L. Wald, M. Mangolini, 1996a, "The ARSIS method: A General Solution For Improving Spatial Resolution Of Images By The Means Of Sensor Fusion". Fusion of Earth Data, Proceedings EARSel Conference, Cannes, France, 6- 8 February 1996 (Paris: European Space Agency).
- [12] Ranchin T., L. Wald, M. Mangolini, C. Penicand, 1996b. "On the assessment of merging processes for the improvement of the spatial resolution of multispectral SPOT XS images". In Proceedings of the conference, Cannes, France, February 6-8, 1996, published by SEE/URISCA, Nice, France, pp. 59-67
- [13] Wald L., 1999b, "Definitions And Terms Of Reference In Data Fusion". International Archives of Photogrammetry and Remote Sensing, Vol. 32, Part 7-4-3 W6, Valladolid, Spain, 3-4 June.
- [14] Pohl C., 1999. "Tools And Methods For Fusion Of Images Of Different Spatial Resolution". International Archives of Photogrammetry and Remote Sensing, Vol. 32, Part 7-4-3 W6, Valladolid, Spain, 3-4 June.
- [15] Hsu S. H., Gau P. W., I-Lin Wu I., and Jeng J. H., 2009, "Region-Based Image Fusion with Artificial Neural Network". World Academy of Science, Engineering and Technology, 53, pp 156 -159.
- [16] Zhang J., 2010. "Multi-source remote sensing data fusion: status and trends", International Journal of Image and Data Fusion, Vol. 1, No. 1, pp. 5-24.
- [17] Ehlers M., S. Klonusa, P. Johan A °strand and P. Rosso, 2010. "Multi-sensor image fusion for pansharpening in remote sensing". International Journal of Image and Data Fusion, Vol. 1, No. 1, March 2010, pp. 25-45
- [18] Vijayaraj V., O'Hara C. G. And Younan N. H., 2004. "Quality Analysis Of Pansharpened Images". 0-7803-8742-2/04/(C) 2004 IEEE, pp.85-88

- [19] ŠVab A. and Oštir K., 2006. "High-Resolution Image Fusion: Methods To Preserve Spectral And Spatial Resolution". *Photogrammetric Engineering & Remote Sensing*, Vol. 72, No. 5, May 2006, pp. 565–572.
- [20] Parcharidis I. and L. M. K. Tani, 2000. "Landsat TM and ERS Data Fusion: A Statistical Approach Evaluation for Four Different Methods". 0-7803-6359-0/00/ 2000 IEEE, pp.2120-2122.
- [21] Ranchin T., Wald L., 2000. "Fusion of high spatial and spectral resolution images: the ARSIS concept and its implementation". *Photogrammetric Engineering and Remote Sensing*, Vol.66, No.1, pp.49-61.
- [22] Prasad N., S. Saran, S. P. S. Kushwaha and P. S. Roy, 2001. "Evaluation Of Various Image Fusion Techniques And Imaging Scales For Forest Features Interpretation". *Current Science*, Vol. 81, No. 9, pp.1218
- [23] Alparone L., Baronti S., Garzelli A., Nencini F. , 2004. " Landsat ETM+ and SAR Image Fusion Based on Generalized Intensity Modulation". *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 42, No. 12, pp. 2832-2839
- [24] Dong J., Zhuang D., Huang Y., Jingying Fu, 2009. "Advances In Multi-Sensor Data Fusion: Algorithms And Applications ". Review , ISSN 1424-8220 *Sensors* 2009, 9, pp.7771-7784.
- [25] Amarsaikhan D., H.H. Blotvogel, J.L. van Genderen, M. Ganzorig, R. Gantuya and B. Nergui, 2010. "Fusing high-resolution SAR and optical imagery for improved urban land cover study and classification". *International Journal of Image and Data Fusion*, Vol. 1, No. 1, March 2010, pp. 83–97.
- [26] Vrabel J., 1996. "Multispectral imagery band sharpening study". *Photogrammetric Engineering and Remote Sensing*, Vol. 62, No. 9, pp. 1075-1083.
- [27] Vrabel J., 2000. "Multispectral imagery Advanced band sharpening study". *Photogrammetric Engineering and Remote Sensing*, Vol. 66, No. 1, pp. 73-79.
- [28] Gangkofner U. G., P. S. Pradhan, and D. W. Holcomb, 2008. "Optimizing the High-Pass Filter Addition Technique for Image Fusion". *Photogrammetric Engineering & Remote Sensing*, Vol. 74, No. 9, pp. 1107–1118.
- [29] Wald L., T. Ranchin and M. Mangolini, 1997. 'Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images', *Photogrammetric Engineering and Remote Sensing*, Vol. 63, No. 6, pp. 691–699.
- [30] Li J., 2001. "Spatial Quality Evaluation Of Fusion Of Different Resolution Images". *International Archives of Photogrammetry and Remote Sensing*. Vol. XXXIII, Part B2, Amsterdam 2000, pp.339-346.
- [31] Aiazzi, B., L. Alparone, S. Baronti, I. Pippi, and M. Selva, 2003. "Generalised Laplacian pyramid-based fusion of MS + P image data with spectral distortion minimization". URL: <http://www.isprs.org/commission3/proceedings02/papers/paper083.pdf> (Last date accessed: 8 Feb 2010).
- [32] Hill J., C. Diemer, O. Stöver, Th. Udelhoven, 1999. "A Local Correlation Approach for the Fusion of Remote Sensing Data with Different Spatial Resolutions in Forestry Applications". *International Archives Of Photogrammetry And Remote Sensing*, Vol. 32, Part 7-4-3 W6, Valladolid, Spain, 3-4 June.
- [33] Carter, D.B., 1998. "Analysis of Multiresolution Data Fusion Techniques". Master Thesis Virginia Polytechnic Institute and State University, URL: <http://scholar.lib.vt.edu/theses/available/etd-32198-21323/unrestricted/Etd.pdf> (last date accessed: 10 May 2008).
- [34] Aiazzi B., S. Baronti , M. Selva, 2008. "Image fusion through multiresolution oversampled decompositions". in *Image Fusion: Algorithms and Applications* "Edited by: Stathaki T. "Image Fusion: Algorithms and Applications". 2008 Elsevier Ltd.
- [35] Lillesand T., and Kiefer R. 1994. "Remote Sensing And Image Interpretation". 3rd Edition, John Wiley And Sons Inc.,
- [36] Gonzales R. C, and R. Woods, 1992. "Digital Image Processing". Addison-Wesley Publishing Company.
- [37] Umbaugh S. E., 1998. "Computer Vision and Image Processing: A practical Approach Using CVIP tools". Prentice Hall.
- [38] Green W. B., 1989. *Digital Image processing A system Approach*" 2<sup>nd</sup> Edition. Van Nostrand Reinhold, New York.
- [39] Sangwine S. J., and R.E.N. Horne, 1989. "The Colour Image Processing Handbook". Chapman & Hall.
- [40] Gross K. and C. Moulds, 1996. *Digital Image Processing*. (<http://www.net/DigitalImageProcessing.htm>). (last date accessed: 10 Jun 2008).
- [41] Jensen J.R., 1986. "Introductory Digital Image Processing A Remote Sensing Perspective". Englewood Cliffs, New Jersey: Prentice-Hall.
- [42] Richards J. A., and Jia X., 1999. "Remote Sensing Digital Image Analysis". 3rd Edition. Springer - verlag Berlin Heidelberg New York.
- [43] Cao D., Q. Yin, and P. Guo, 2006. "Mallat Fusion for Multi-Source Remote Sensing Classification". *Proceedings of the Sixth International Conference on Intelligent Systems Design and Applications (ISDA'06)*
- [44] Hahn M. and F. Samadzadegan, 1999. " Integration of DTMS Using Wavelets". *International Archives Of Photogrammetry And Remote Sensing*, Vol. 32, Part 7-4-3 W6, Valladolid, Spain, 3-4 June. 1999.
- [45] King R. L. and Wang J., 2001. "A Wavelet Based Algorithm for Pan Sharpening Landsat 7 Imagery". 0-7803-7031-7/01/ 02001 IEEE, pp. 849- 851
- [46] Kumar Y. K., "Comparison Of Fusion Techniques Applied To Preclinical Images: Fast Discrete Curvelet Transform Using Wrapping Technique & Wavelet Transform". *Journal Of Theoretical And Applied Information Technology*. © 2005 - 2009 Jatit, pp. 668-673
- [47] Malik N. H., S. Asif M. Gilani, Anwaar-ul-Haq, 2008. "Wavelet Based Exposure Fusion". *Proceedings of the World Congress on Engineering 2008 Vol I WCE 2008*, July 2 - 4, 2008, London, U.K
- [48] Li S., Kwok J. T., Wang Y., 2002. "Using The Discrete Wavelet Frame Transform To Merge Landsat TM And SPOT Panchromatic Images". *Information Fusion* 3 (2002), pp.17–23.

- [49] Garzelli, A. and Nencini, F., 2006. "Fusion of panchromatic and multispectral images by genetic Algorithms". IEEE Transactions on Geoscience and Remote Sensing, 40, 3810–3813.
- [50] Aiazzi, B., Baronti, S., and Selva, M., 2007. "Improving component substitution pan-sharpening through multivariate regression of MS+Pan data". IEEE Transactions on Geoscience and Remote Sensing, Vol.45, No.10, pp. 3230–3239.
- [51] Das A. and Revathy K., 2007. "A Comparative Analysis of Image Fusion Techniques for Remote Sensed Images". Proceedings of the World Congress on Engineering 2007, Vol. I, WCE 2007, July 2 – 4, London, U.K.
- [52] Pradhan P.S., King R.L., 2006. "Estimation of the Number of Decomposition Levels for a Wavelet-Based Multi-resolution Multi-sensor Image Fusion". IEEE Transaction of Geosciences and Remote Sensing, Vol. 44, No. 12, pp. 3674-3686.
- [53] Hu Deyong H. L., 1998. "A fusion Approach of Multi-Sensor Remote Sensing Data Based on Wavelet Transform". URL: <http://www.gisdevelopment.net/AARS/ACRS1998/Digital Image Processing> (last date accessed: 15 Feb 2009).
- [54] Li S., Li Z., Gong J., 2010. "Multivariate statistical analysis of measures for assessing the quality of image fusion". International Journal of Image and Data Fusion Vol. 1, No. 1, March 2010, pp. 47–66.
- [55] Böhler W. and G. Heinz, 1998. "Integration of high Resolution Satellite Images into Archaeological Documentation". Proceeding International Archives of Photogrammetry and Remote Sensing, Commission V, Working Group V/5, CIPA International Symposium, Published by the Swedish Society for Photogrammetry and Remote Sensing, Goteborg. (URL: <http://www.i3mainz.fh-mainz.de/publicat/cipa-98/sat-im.html>) (Last date accessed: 28 Oct. 2000).

has More than 60 scientific papers published in scientific journals in several scientific conferences.

## AUTHORS

**Mrs. Firouz Abdullah Al-Wassai.** Received the B.Sc. degree in, Physics from University of Sana'a, Yemen, Sana'a, in 1993. The M.Sc. degree in, Physics from Bagdad University, Iraq, in 2003, Research student. Ph.D in the department of computer science (S.R.T.M.U), India, Nanded.

**Dr. N.V. Kalyankar.** B.Sc. Maths, Physics, Chemistry, Marathwada University, Aurangabad, India, 1978. M Sc. Nuclear Physics, Marathwada University, Aurangabad, India, 1980. Diploma in Higher Education, Shivaji University, Kolhapur, India, 1984. Ph.D. in Physics, Dr.B.A.M. University, Aurangabad, India, 1995. Principal Yeshwant Mahavidyalaya College, Membership of Academic Bodies, Chairman, Information Technology Society State Level Organization, Life Member of Indian Laser Association, Member Indian Institute of Public Administration, New Delhi, Member Chinmay Education Society, Nanded. He has one publication book, seven journals papers, two seminars Papers and three conferences papers.

**Dr. Ali A. Al -Zuky.** B.Sc Physics Mustansiriyah University, Baghdad, Iraq, 1990. M Sc. In 1993 and Ph. D. in 1998 from University of Baghdad, Iraq. He was supervision for 40 post-graduate students (MSc. & Ph.D.) in different fields (physics, computers and Computer Engineering and Medical Physics). He



# Using Fuzzy Decision-Making in E-tourism Industry: A Case Study of Shiraz city E-tourism

Zohreh Hamedi<sup>1</sup>, Shahram Jafari<sup>2</sup>

<sup>1</sup> Faculty of E-Learning , Shiraz University, Iran

<sup>2</sup> School of Electrical and Computer Science, Shiraz University, Iran

## Abstract

In recent years, e-commerce has had great impacts on various industries by developing new approaches. Its benefits include faster and easier access to information and the possibility to coordinate for any task before attempting it. In the tourism sector, e-commerce is playing a great role to develop the industry and improve services. On the other hand, combining e-commerce technology with mathematics and the other basic sciences has provided special facilities for flexibility in complying human needs. These include the use of fuzzy knowledge in this technology. Tourism combined with fuzzy knowledge and e-commerce technology will create further expansion of this industry especially in better addressing customers' needs and tastes. The aim of this project is to introduce an electronic tourism system (e-tourism) based on fuzzy knowledge for the city of Shiraz, as a case study. This electronic system is in the form of a website, which tourists can use to find an appropriate accommodation by inputting data related to their interests and needs.

**Keywords:** *E-Tourism, Fuzzy, Decision-Making, Internet, Shiraz*

## 1. Introduction

The introduction of the internet in the early 1990s has changed the way of doing business in the tourism industry dramatically [1]. The Internet is already the primary source of tourist destination information for travelers [2]. Nowadays, most people who plan a trip or a day-out will first initiate a search through the internet. More and more people realize the advantages of the new technologies for planning leisure activities as an increasing number of companies and institutions offer tourist information which is easily accessible through web services [3]. The Internet has improved hotel reservation process and facilitated extensive services for online distribution and bookings, which are reliable, diverse and rapid. Hotels can develop their presence and partnership with distributors.

Reservation through the Internet provides effective and efficient communication mechanism, particularly for their frequent customers [4].

Using efficient and helpful techniques to suggest better options to tourists will result in customer satisfaction, which in turn attracts tourists and promote tourism industry. There are many factors affecting tourists' decision making, but the main factor is to suggest options that better address customers' interests, needs and preferences. Various techniques have been developed to facilitate this task. Fuzzy logic improves classification and decision support systems by allowing the use of overlapping class definitions and improves the interpretability of the results by providing more insight into the classifier structure and decision making process [5]. In tourism, selecting suitable accommodation considering costs, facilities and distance to the tourists' destination are very important. Therefore, this project was a new attempt to use fuzzy knowledge and its inference method for e-tourism in Shiraz, providing an electronic system to suggest a list of accommodation to the tourists based on their interests and priorities.

## 2. Fuzzy decision-making structure

The overall structure of decision-making in a fuzzy environment is presented in figure 1.

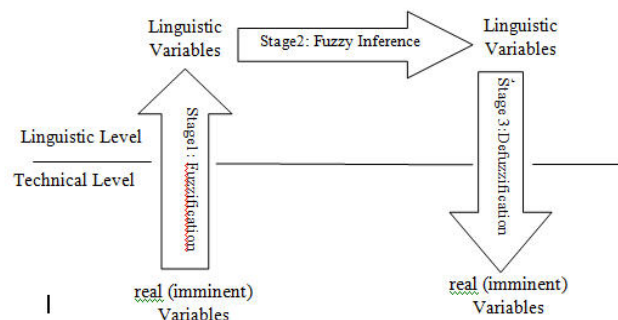


Fig. 1. Fuzzy Decision-making structure[6]

## 2-1. Step 1: Fuzzification

The first step in fuzzy decision-making process is making fuzzy real (imminent) variables, in which absolute variables are converted to linguistic variables. This step is called Fuzzification, since fuzzy sets are used to convert real (imminent) variables to fuzzy variables [6]. Fuzzy membership functions are needed for this purpose.

### 2-1-1. Chart and fuzzy membership functions

A fuzzy set is described by its membership functions. A more precise way to define a membership function is expressing it as a mathematical formula. Several different classes of membership parametric function are introduced, and in real world of fuzzy sets applications, membership function shapes usually are restricted to a definite class of functions that can clarify with few parameters [7]. Most famous shapes are Triangular, Trapezoidal and Gaussian shaped as shown in figure 2.

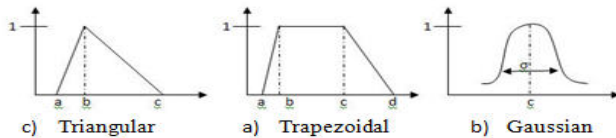


Fig. 2. Fuzzy Decision-making structure[6]

## 2-2. Step 2: Fuzzy inference

In this step, the behavior of a system is defined using a set of “if → then” rules. Result of this inference will be a linguistic value for a linguistic variable[6]. In the case study section, you can find more explanation about Fuzzy inference way.

## 2-3. Step 3: De-Fuzzification

In the third step (making definite), linguistic values will be changed to definite numbers in order to do decision-making [6].

## 3. Case Study

This is a case study on Shiraz e-tourism. The goal is to suggest the best accommodation to tourists, based on their preferences. The tourists enter their preferences for accommodation, including their budget, residence facilities and desired distance from sights of visit. System does the Fuzzification of tourists' desired values, including budget, facilities and distance to the spots, and then

prioritizes the accommodations, using Fuzzy inference methods.

## 3-1. Tourist decision-making criteria (system input)

In decision models, criteria selection process is accomplished according to decision objectives [9, 10]. In this case study, the following factors are considered:

- Accommodation cost for one night in Shiraz, using fuzzy charts, are converted to linguistic values of “cheap”, “moderate” and “expensive”.
- Importance of each accommodation facilities. In this case study, we considered eight facilities. This criteria is converted to linguistic values of “low”, “medium” and “high”.
- Distance from historical sight in downtown Shiraz (origin: Arg-e Karimkhan)
- Distance from business center (origin: Setareh-Fars shopping center)
- Distance from cultural attractions (origin: Hafeziyeh)
- Distance from pilgrimage center (origin: Shah-Cheragh)
- Distance from academic center (origin: Shiraz University - Faculty of Engineering, building No. 1)

All of distances are converted to linguistic parameters of “far”, “average” and “near,” using the relations of fuzzy charts.

## 3-2. Data Collection

One of the required data for calculation in fuzzy decision-making is accommodations' price list, facilities and distance from the desired origin. Membership matrix is derived from these data and is kept for the next calculation. In this case study, all distances are calculated based on the newest map of Shiraz and in kilometers. Figure 3 shows the price of accommodations and distance from different spots in the website admin panel, where this information can be easily modified using related forms.

Trade center	Historical	Cultural	University	Hotels	Hotels Degree	Hotels Cost(\$)	Hotels Names
1.48	5	0.49	2.47	2.64	4	90	هنل پارک Park
1.57	4.29	0.58	3.79	1.98	2	41	هنل کوکزار Koozar
1.73	4.12	0.74	3.63	1.81	3	57	هنل ارام Eram
3.38	4.62	2.39	2.14	2.31	5	151	هنل هوما Homa
3.21	2.64	2.23	3.88	0.33	4	128	هنل پارس Pars
1.48	5.36	0.49	2.31	3.46	3	80	هنل پارسه Parsesh
2.8	6.19	1.81	0.41	3.88	4	118	هنل پرسپولیس Pirsopolice
3.46	7.26	2.47	1.98	3.03	3	52	هنل آرژ Arzi
2.14	4.62	1.15	3.38	2.31	3	60	هنل آپادانا Apadana
2.72	6.1	1.73	0.49	3.79	3	60	هنل اطلس Atlas
1.81	4.29	0.82	3	1.98	1	30	هنل ساسان Sasian
0.66	5.11	0.25	2.31	3.79	1	30	هنل سینا Sina
1.48	5.36	0.49	1.98	3.05	2	43	هنل حافظ Hafez
6.5	3	5.5	7	2.97	5	147	هنل چمران Chamran



Fig. 3. Reporting of real hotels cost and their distances from different centers in admin panel of website

### 3-3. Fuzzy decision-making steps

#### 3-3-1. Step 1: Data Fuzzification

As previously mentioned, in order to fuzzify data, membership functions and fuzzy charts are used. Boundary values of function can be managed by website administrator, and given values are just samples used for the purpose of this project and based on the current prices, distances and the researcher's interests.

##### 3-3-1-1. Membership functions for Shiraz Case Study

In Shiraz case study, membership functions are considered as triangular. Triangular membership function formula for three different scales of linguistic variables are "low", "Average" and "high".

- Membership Function for linguistic variable: "Low" :

$$\mu_{cheap}(x) = \begin{cases} 1 & \text{if } x = 0 \\ \frac{c-x}{c} & \text{if } 0 \leq x < c \\ 0 & \text{if } x \geq c \end{cases} \quad (1)$$

- Membership Function for linguistic variable: "Average" :

$$\mu_{medium}(x) = \begin{cases} 0 & \text{if } x \leq a_m \\ \frac{x - a_m}{b_m - a_m} & \text{if } a_m \leq x \leq b_m \\ \frac{c_m - x}{c_m - b_m} & \text{if } b_m \leq x \leq c_m \\ 0 & \text{if } x \geq c_m \end{cases} \quad (2)$$

- Membership Function for linguistic variable: "High" :

$$\mu_{expensive}(x) = \begin{cases} 0 & \text{if } x \leq a_h \\ \frac{x - a_h}{b_h - a_h} & \text{if } a_h \leq x < b_h \\ 1 & \text{if } x \geq b_h \end{cases} \quad (3)$$

These membership functions are configured for different parameters of accommodation budget, distances, and facilities, with different bordering values. Figure 4, 5 and 6 show these functions, drawn in MatLab.

##### 3-3-1-2. Shiraz tourism fuzzy charts

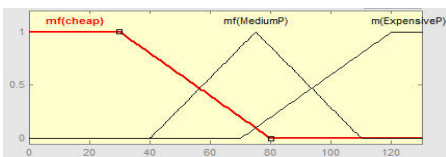


Fig 4: Membership function charts for 1 night stay in hotels.

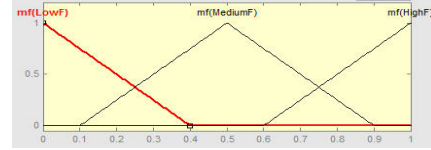


Fig 6: Membership function charts for hotels facilities

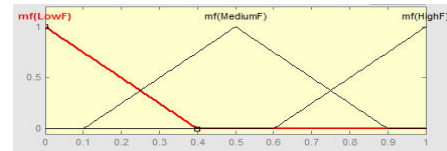


Fig 6: Membership function charts for hotels facilities

##### 3-3-1-3. Tourist entry in the system

Tourist input form is shown blow. As you see in the form, criteria stated in Section 3-1 are entered as input by the tourists.

Fig. 7. Request form of tourist

To increase options for distance, at first users can choose from the categories of visiting spots including cultural, historical, pilgrimage, commercial or academic spots. Then, the users select distance from these spots. Minimum and maximum distance values and prices are adjusted by the website administrator. Users have to register to the website to be able to access to this information and follow their requests.

##### 3-3-2. step 2: Fuzzy inference

In this step, the system should be able to suggest suitable accommodations to the users based on input values and tourist interests and priorities. In the usual fuzzy method,

at first it is needed to define a set of “if → then” rules, which are defined as follow in this research.

### 3-3-2-1. Fuzzy Rules

Fuzzy rules are defined and adjusted by the administrator or a skilled expert. Website software has the ability to create, modify or delete the conditions. Figure 8 shows a part of adjusted fuzzy conditions report in admin panel of the website.



Fig 8. Reporting of adjusted fuzzy conditions in admin panel of website

### 3-3-2-2. Inference Method 1: Min-Max or Max-Prod method

One way to compare two fuzzy sets is using the criteria of necessity and possibility. The possibility criteria of fuzzy set A in relation to fuzzy set B is defined as Pos (A, B), which is as follows:

$$Pos(A, B) = \max[\min(A(x), B(x))] \quad x \in X \quad (4)$$

Criterion and measurement possibility shows the overlap amount of A and B [7]. So, using the fuzzy max-min or max-prod relation, fuzzy rules and tourist entry, we can obtain the priorities.

### 3-3-2-3. Inference Method2: Distance computing method

In addition to standard fuzzy method, we can use another method called “Euclidean distance” to compare the user's requests with existing accommodations and determine their priority. In this method, a three-dimensional space of cost-distance-facilities based on fuzzy values are considered and marked on each accommodation using distance, cost and facilities membership functions. The location of tourist entry is determined in this space, and then the distance of this point to each accommodation will be computed. The accommodations with less distance to

user's request have more priority. The advantage of this method is independency to fuzzy rules as well as simplicity, and priority of all hotels will be computed. It should be noted that all three different parameters (cost, distance and facilities) must be changed to fuzzy-value and also normalized. In other words, fuzzification is required and since the number of decision criteria is more than one, we cannot apply it on non-fuzzy values.

In general, if A and B are from a world debate X, we can define the distance between A and B using Minkovsky rule as follows:

$$D(A, B) = \left( \sum |A(x) - B(x)|^P \right)^{1/P} \quad x \in X \quad (5)$$

Considering  $P \geq 1$  [7].

There are different states for P in applications. The best state that matches our problem is when  $P = 2$ , which is called “Euclidean distance”.

### 3-3-3. Step 3: De-Fuzzification

It is clear that in system output, suggested accommodations are shown to the users. In addition, system can reply as an e-commercial system with fee.

Since the membership functions are not included in the output of fuzzy inference system in our case study, and the output is only a list of suggested accommodation, the output values are constant and no-fuzzy and then, our problem is from Fuzzy-Crisp type and hence there isn't the third step or de-fuzzification.

### 3-4. Comparing method 1 and 2

With a brief comparison between the methods of inference 1 and 2, they can be expressed as bellow:

1. Euclidean distance method determines priorities of all accommodations, while in Max-Prod or Max-Min method just the nearest accommodations are shown and the users have no idea about them unless select their features. But, in distance computing method, using a simple mathematical formula, the priority of all accommodations will be determined. On the other hand, providing priorities of all accommodations does not seem to be very interesting, especially when there are enormous accommodations, which makes users confused.

2. In standard fuzzy, when the numbers of decision-making factors increase, the number of fuzzy rule must be increased. For example, for three 3-state factors, a maximum of 81 rules are needed, and if the number of

factors increases, the number of fuzzy rules will be much more. However, according to skilled experts' opinion, many of these rules may not be used, but the rules will increase and need to be analyzed.

3. Computational complexity of the method of calculating the distance is much less than the standard method.

4. In terms of development capability and generalization, if the number of factors increases, calculating the distance method will be quite responsive and easy and just by putting in a formula, priorities will be determined. But, if we want to generalize the users' input values on an interval, then Max-Min or Max-Prod method has more capability and flexibility.

## Conclusion

In this research, as a new work, we tried to apply fuzzy decision-making in e-tourism industry and in order to increase research integrity, we focused on Shiraz city. Our goal was to find a simple and applicable way. So, we used two methods; one of them is the usual method for fuzzy decision-making, and the other is Euclidean distance method, which is very simple in calculation. After inspecting both methods, we selected the usual method as the main method of fuzzy inference in our website. Distance method also can present a complete list of all accommodations and their priorities to the users. The most important result in this research is providing a system of e-tourism in which tourists can enter their interests and needs without conflicting binary systems and receive an appropriate suggestion to plan their traveling to Shiraz. The results of this research show that in an e-tourism system using fuzzy decision-making is more efficient.

## References

- [1] M. Moharrer, and T. Tahayori, "Drivers of customer convenience in electronic tourism industry", Canadian Conference on Electrical and Computer Engineering (CCECE) IEEE, p. 836, 2007.
- [2] Waralak V. and Siricharoen, "E-commerce adaptation using ontologies for E-tourism", IEEE International Symposium on Communication and information Technologies, 2007.
- [3] Laura Sebastia and Inma Garcia and Eva Onaindia and Cesar Guzman, "e-Tourism: a tourist recommendation and planning application", 20th IEEE International Conference on Tools with Artificial Intelligence, 2008.
- [4] Mazyar Yari, and Hosein Vazifehdust, "Electronic tourism: the interaction between e-commerce and tourism industry", Iran, 4th national conference of e-commerce. 2007
- [5] Valente de, and Oliveria J., "Semantic constraints for membership function optimization", IEEE Transaction on Fuzzy System 19, 128-138, 1999.
- [6] Adel Azar, and Hojjat Faraji, Fuzzy Management Science, Iran, Ketab Mehraban Nashr Institute. 2008 [In Farsi]

[7] Mehmed Kantardzik, Data mining, translated by Amir Alikhanzadeh, Iran, Oloom Rayaneh Publication. 2006 [In Farsi].

[8] Zadeh, L.A. and Bellman, R.E. "Decision-making in a fuzzy environment", management science, Vol. 17, No.4, pp.141-164,1970.

[9] Ahmad Jamali, and Mahmud Saremi, "Using Fuzzy Multi Attribute decision-making model for selection of foreign investment method in the high managerial deputies of Oil Industry in Iran.", Quarterly research of trade. No 29. P. 167-188, 2003.

**Z. Hamedi** was born in Iran. She receives B.Sc degree in computer engineering from Shahid Bahonar Kerman university in 2000. She is currently a M.Sc. student in Information Technology at Shiraz University. Her research interests include Information Technology, telecommunications and computer networks.

**S. Jafari** received the PhD degree in Computer Systems Engineering from Monash University, Australia in 2006. He is currently a lecturer in the Electrical and Computer Engineering School, Shiraz University, Shiraz, Iran. His research interests include Artificial Intelligence, especially Expert Systems Robotics, Hybrid Systems: Fuzzy logic, Certainty Factor, Neural Network combinations: Neuro-Fuzzy, Fuzzy neural Networks and Image Processing.

# A Reliable routing algorithm for Mobile Adhoc Networks based on fuzzy logic

Arash Dana<sup>1</sup>, Golnoosh Ghalavand<sup>2</sup>, Azadeh Ghalavand<sup>3</sup> and Fardad Farokhi<sup>4</sup>

<sup>1</sup> Electrical Engineering Department, Islamic Azad University Central Tehran Branch  
Tehran, Iran

<sup>2</sup> Computer Engineering Department, Islamic Azad University Science And Research Branch  
Tehran, Iran

<sup>3</sup> Computer Engineering Department, Islamic Azad University South Tehran Branch  
Tehran, Iran

<sup>4</sup> Electrical Engineering Department, Islamic Azad University Central Tehran Branch  
Tehran, Iran

## Abstract

By growing the use of real-time application on mobile devices, there is a constant challenge to provide reliable and high quality routing algorithm among these devices. In this paper, we propose a reliable routing algorithm based on fuzzy-logic (RRAF) for finding a reliable path in Mobile Ad Hoc Networks. In this scheme for each node we determine two parameters, trust value and energy value, to calculate the lifetime of routes. Every node along route discovery, records its trust value and energy capacity in RREQ packet. In the destination with the aid of fuzzy logic, a new parameter is generated from inputs trust value and energy value of each route which is called "Reliability Value". The path with more reliability value is selected as a stable route from source to destination. Simulation results show that RRAF has significant reliability improvement in comparison with AODV.

**Keywords:** Mobile AdHoc Networks, Routing, Reliability, Fuzzy logic, RRAF.

## 1. Introduction

A mobile ad hoc network is an independent group of mobile users which communicate over unstable wireless links. Because of mobility of nodes, the network topology may change rapidly and unpredictably over time. All network activity, including delivering messages and discovering the topology must be executed by the nodes themselves. Therefore routing functionality, the act of moving information from source to a destination, will have to be incorporated into the mobile nodes. Hence routing is one of the most important issue in MANET. Routing protocols in MANETs are generally classified as proactive and reactive [1]. Reactive routing protocols [2,3,4,5,6,7], which also called on demand routing

protocols, start to establish routes when required. These kind of protocols are based on broadcasting RREQ and RREP messages. The duty of RREQ message is to discover a route from source to destination node. When the destination node gets a RREQ message, it sends RREP message along the established path. On demand protocols minimize the whole number of hops of the selected path and also they are usually very good on single rate networks. There are many reactive routing protocols, such as ad hoc on-demand distance vector (AODV) [6], dynamic source routing (DSR) [4], temporally order routing algorithm (TORA)[5], associativity-based routing (ABR) [7], signal stability-based adaptive (SSA) [3], and relative distance microdiscovery ad hoc routing (RDMAR) [2]. In contrast, in table-driven or pro-active routing protocols [8,9,10,11,12], each node maintains one or more routing information table of all the participating nodes and updates their routing information frequently to maintain latest view of the network. In proactive routing protocols when there is no actual routing request, control messages transmit to all the nodes to update their routing information. Hence proactive routing protocols bandwidth become deficient. The major disadvantage of pro-active protocols is the heavy load caused from the need to broadcast control messages in the network [3]. There are many proactive routing protocols, such as destination sequenced distance vector (DSDV) [12], wireless routing protocol (WRP) [9], clusterhead gateway switch routing (CGSR) [10], fisheye state routing (FSR) [11], and optimized link state routing (OLSR) [8]. Many of the work reported on routing protocols have focused only on shortest path, power aware and minimum cost. However much less attention has been paid in making the routing protocol to choose a more reliable route. In critical environment like military operation, data packets

are forwarded to destination through reliable intermediate nodes[13].In this paper, we propose a reliable routing algorithm based on fuzzy logic. In this scheme for each node we determine two parameters , trust value and energy value, to calculate the lifetime of routes . During route discovery, every node inserts its trust value and energy value in RREQ packet .In the destination , based on a new single parameter which is called reliability value , is decided which route is selected. The route with higher reliability value is candidated to route data packets from source to destination.

The rest of the paper is organized as follows: In Section 2, we briefly describe the related work. Section 3 describes our proposed routing algorithm and its performance is evaluated in Section 4.Finally,Section 5 concludes the paper.

## 2. Related Works

We can classify all the works that have been done in reliable routing, in three categories: GPS-aided protocols ,energy aware routing ,and trust evaluation methods .In this section, we will overview some proposed protocols that have been given to designing reliable routing protocols.

A reliable path has more stability than a command path. Some of reliable routing protocols propose a GPS-aided process and use route expiration time to select a reliable path. In [14] Nen-chung Wang et al, propose a stable weight-based on-demand routing protocol (SWORP) for MANETs. The proposed scheme uses the weight-based route strategy to select a stable route in order to enhance system performance .The weighth of a route is decided by three factors:the route expiration time, the error count , and the hop count . Route discovery usually first finds multiple routes from the source node to the destination node. Then the path with the largest weighth value for routing is selected .

In [15], Nen-Chung Wang and Shou-Wen Chang also propose a reliable on-demand routing protocol (RORP) with mobility prediction. In this scheme, the duration of time between two connected mobile nodes is determined by using the global positioning system (GPS) and a request region between the source node and the destination node is discovered for reducing routing overhead. the routing path with the longest duration of time for transmission is selected to increase route reliability. In [16], Neng-Chung Wang etal, propose a reliable multi-path QoS routing (RMQR) protocol for MANETs by constructing multiple QoS paths from a source node to a destination node. The proposed protocol is an on-demand QoS aware routing scheme. They examine the QoS routing problem associated with searching for a reliable multi- path (or uni-path) QoS route

from a source node to a destination node in a MANET. This route must also satisfy certain bandwidth requirements. They determine the route expiration time (RET) between two connected mobile nodes by using global positioning system (GPS). Then use two parameters, the route expiration time and the number of hops, to select a routing path with low latency and high stability.

some other proposed protocols are considering energy and trust evaluation as a factor of reliability . In [17], an approach has been proposed in which the intermediate nodes calculate cost based on battery capacity. The intermediate node take into consideration whether they can forward RREQ packet or not . This protocol improves packet delivery ratio and throughput and reduces nodes energy consumption[13].In [18], Gupta Nishant and Das Samir had proposed a method to make the protocols energy aware .They were using a new function of the remaining battery level in each node on a route and number of neighbours of the node. This protocol gives significant benefits at high traffic but at low mobility scenarios[13].In [19], a novel method has been discussed for maximizing the life span of MANET by integrating load balancing and transmission power control approach. The simulation results of this mechanism showed that the average required transmission energy per packet was reduced in comparison with the standard AODV. In [20] Pushpalatha & Revathy have proposed a trust model in DSR protocol that categorize trust value as friend, acquaintance and stranger based on the number of packets transferred successfully by each node[13].The most trusted path was determined from source to destination.Results indicated that the proposal had a minimum packet loss when compared to the conventional DSR.Huafeng Wu & Chaojian Shi [21] has proposed the trust management model to get the trust rating in peer to peer systems, and aggregation mechanism is used to indirectly combine and obtain other node's trust rating[13]. The result shows that the trust management model can quickly detect the misbehaviour nodes and limit the impacts of them in a peer to peer file sharing system[13].all above papers used the separate parameters such as battery power ,trust of a node or route expiration time individually as a factor for measuring reliability of route. In this paper, we consider both energy capacity and trust of nodes for route discovery .

## 3. Proposed Model

In this section we propose our novel reliable routing algorithm which is improved version of [22].

### 3.1 RRAF Mechanism

Trust value and battery capacity are the two main parameters in this method that make the routing algorithm more reliable. Before explaining the algorithm, trust estimation and power consumption mechanism are described below.

**Trust Evaluation:** Trust value of each node is measured based on the various parameters like length of the association, ratio of number of packets forwarded successfully by the neighbors to the total number of packets sent to that neighbor and average time taken to respond to a route request [13,20]. Based on the above parameters trust level of a node  $i$  to its neighbor node  $j$  can be any of the following types:

a) Node  $i$  is a stranger to neighbor node  $j$

Node  $i$  have never sent/received message to/from node  $j$ . Their trust levels between each other will be low. Every new node which is entering an ad hoc network will be a stranger to all its neighbors.

b) Node  $i$  is an acquaintance to neighbor node  $j$

Node  $i$  have sent/received few messages from node  $j$ . Their trust levels are neither too low nor too high to be reliable.

c) Node  $i$  is a friend to neighbor node  $j$

Node  $i$  have sent/received a lot of messages to/ from node  $j$ . The trust levels between them are reasonably high.

The above relationships are represented in Fig.1 as a membership function.

**Energy Evaluation:** We defined that every node is in high level which means it has full capacity (100%). The node will not be a good router to forward the packets If the energy of it falls below 50%.

**FuzzyLogic Controller:** A useful tool for solving hard optimization problems with potentially conflicting objectives is fuzzy logic.

In fuzzy logic, values of different criteria are mapped into linguistic values that characterize the level of satisfaction with the numerical value of the objectives. The numerical values are chosen typically to operate in the interval  $[0, 1]$  according to the membership function of each objective. Fig.1 represents the trust value membership function. According to three types of trust value: friend, acquaintance and stranger, we define three fuzzy sets: high, medium and low, respectively. we also determined three fuzzy sets for node's energy. For energy capacity between 50% to 100% of total capacity, we define high set, for 0% to 100% we define medium set and for 50% to 100% we define low set. The above relationships are represented in Fig.2 as energy value

membership function and Fig.3 shows the membership function of reliability value.

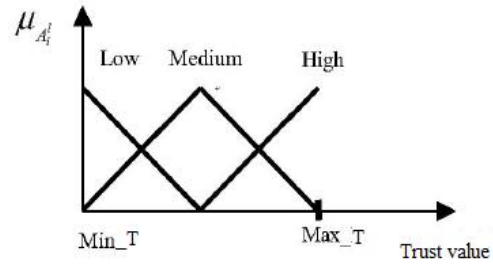


Fig. 1 Membership function for trust value.

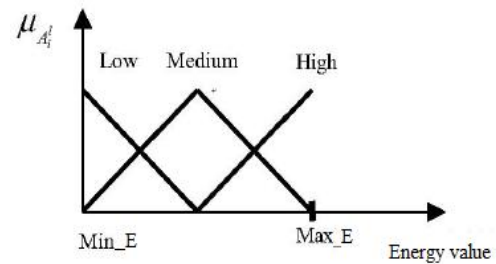


Fig. 2 Membership function for energy value.

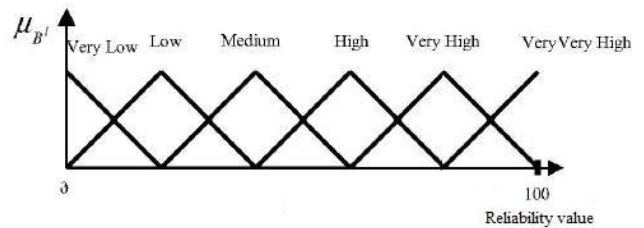


Fig. 3 Membership function for Reliability value.

**Reliability Evaluation:** Reliability factor take different values based on six rules that dependent upon varied input metric values i.e. energy and trust values. A fuzzy system decides for each two input values which values appear in output.

The fuzzy system with product inference engine, singleton fuzzifier and center average defuzzifier are of the following form:

$$f(x) = \frac{\sum_{i=1}^6 \bar{y}^i \left( \prod_{i=1}^2 \mu_{A_i}(x_i) \right)}{\sum_{i=1}^6 \left( \prod_{i=1}^2 \mu_{A_i}(x_i) \right)} \quad (1)$$



In Eq.1,  $x_i$  represents crisp input  $i^{th}$  (energy or trust values),  $\mu_{A_i}(x_i)$  represents fuzzy membership function for input  $i^{th}$ , and  $\bar{y}^j$  is center average of output fuzzy set  $l^{th}$ .

The rules are as follows:

**Rule1:** if trust value is high and energy value is high then reliable value is very high.

**Rule2:** if trust value is medium and energy value is high then reliable value is very high.

**Rule3:** if trust value is high and energy value is medium then reliable value is high.

**Rule4:** if trust value is medium and energy value is medium then reliable value is medium.

**Rule5:** if trust value is low and energy value is medium then reliable value is low.

**Rule6:** if trust value is anything and energy value is low then reliable value is very low.

### 3.1.1. Route discovery procedure

**Step1:** A source node starts to flood RREQ packets to its neighboring nodes in a MANET until they arrive at their destination node. Each RREQ consists of sourceid,destinationid,energy value and trust value of nodes along the path.

**Step2:** If the intermediate node N receives a RREQ packet and it is not the destination, then the information of node N is added to the RREQ packet which is appended to packet fields. After that, node N reforward the packet to all the neighboring nodes of itself.

**Step 3:** If node N receives a RREQ packet and node N is the destination, it waits a period of time. therefore, the destination node may receive many different RREQ packets from the source. Then it calculates the value of reliability value for each path from source to the destination using the information in each RREQ packet. Finally, destination node sends a route reply(RREP) packet along the path which has a maximum reliable value.

## 4. Simulation and results

The simulation environment is constructed by an 1500m x300m rectangular simulation area and 50 nodes, distributed over the area. Initial energy of a battery of each node is 4 Watts which is mapped to 100%. Simulation results have been compared with AODV. Simulation study has been performed for packet delivery ratio, throughput and end to end delay evaluations.

**Packet delivery ratio:** The fraction of successfully received packets, which survive while finding their destination. This performance measure also determines the completeness and correctness of the routing protocols[23].

**End to End Delay :** Average end to end delay is the delay experienced by the successfully delivered packets in reaching their destinations. This is a good metric for comparing protocols and denotes how efficient the underlying routing algorithm is, because delay primarily depends on optimality of path chosen[23].

**Throughput:** It is defined as rate of successfully transmitted data per second in the network during the simulation. Throughput is calculated such that, it is the sum of successfully delivered payload sizes of data packets within the period, which starts when a source opens a communication port to a remote destination port, and which ends when the simulation stops. Average throughput can be calculated by dividing total number of bytes received by the total end to end delay[23].

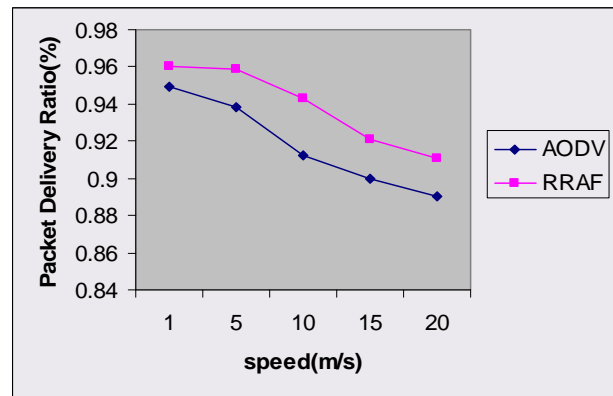


Fig.4 packet delivery ratio at different speed.

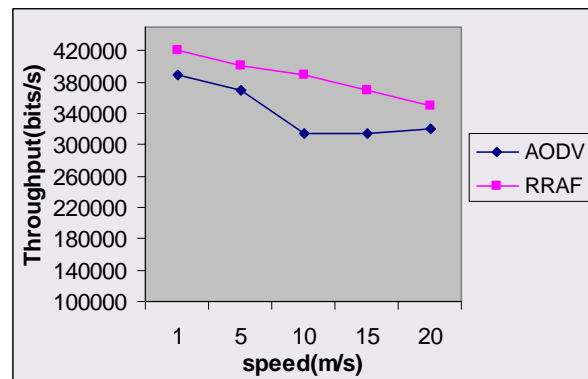


Fig.5 Throughput at different speed.

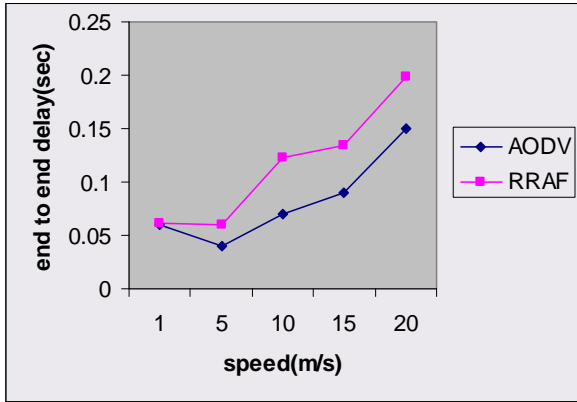


Fig.6 end to end delay at different speed.

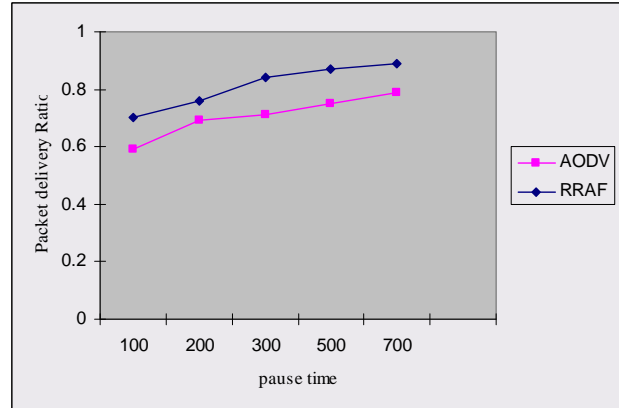


Fig.7 Packet delivery ratio at different pause time.

Fig.4 shows the packet delivery ratio with different mobility speeds. When mobile nodes moved at higher mobility speeds, both protocols decreased the packet delivery ratio. The reason is that the routing path was easy to break when the mobility speed increased, but we can see that RRAF transmits and receives more data packets than AODV. This is because RRAF always chooses the most stable route for transmission packets along the path instead of choosing the shortest path.

In Fig.5 the simulation result shows that throughput of both methods reduces when the speeds increase. When the speed of the mobile node increased, the routing path was more unreliable. The reason is that there were more chances for routes to break when the speed of the mobile node was faster. Thus, the number of rebroadcasts increased. Since RRAF has chosen more reliable route than AODV, we can see that it has performed better at all speeds.

Fig.6 shows average end to end delay with speed as a function. Here it is clear that AODV has less delays than RRAF. Higher delay in the proposed method is because of the time it has wasted for discovering the route with longer life, so the packets would in the meanwhile stay in the buffer until a valid route is found. This takes some time and will, therefore, increase the average delay while AODV chooses the shortest path as a valid path.

Fig.7 shows the performance of packet delivery ratio under various pause times. The results in Fig. 7 illustrate that packet delivery ratio in RRAF is better compared to AODV, and the results in Fig. 8 show that RRAF experiences a high end to end delay because route selection is based on trust and energy level not on the minimum number of hops.

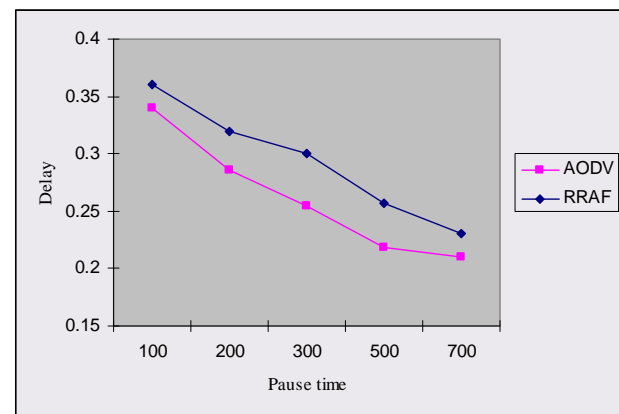


Fig. 8 End to end delay at different pause time.

## 5. Conclusions

Since in MANET, mobile nodes are battery powered and nodes behaviour are unpredictable, wireless links may be easily broken. Hence it is important to find a route that endures a longer time. In this paper, we have proposed a reliable routing algorithm based on fuzzy logic approach. In this scheme, we determine three parameters: trust value, energy value and reliability value that are used for finding a stable route from source to destination. During route discovery, every node records its trust value and energy capacity in RREQ packet. In the destination, based on reliability value, is decided which route is selected. The path with more reliability value is candidate to route data packets from source to destination. The simulation results show that the proposed method has significant reliability improvement in comparison with AODV.

## References

- [1] D. Remondo, "Tutorial on wireless ad hoc networks", Second International Conference in Performance Modeling and Evaluation of heterogeneous networks, July 2004.

- [2] G. Aggelou, R. Tafazolli, RDMAR: a bandwidth-efficient routing protocol for mobile ad hoc networks, Proceedings of the Second ACM International Workshop on Wireless Mobile Multimedia (WoWMoM), August, 1999 pp. 26–33.
- [3] R. Dube, C.D. Rais, K.Y. Wang, S.K. Tripathi, Signal stability-based adaptive routing (SSA) for ad hoc mobile networks, IEEE Personal Communications 4 (1997) 36–45.
- [4] D.B. Johnson, D.A. Maltz, Dynamic Source Routing in Ad Hoc Wireless Networks, Kluwer, 1996.
- [5] V. Park, M.S. Corson, A highly adaptive distributed routing algorithm for mobile wireless networks, Proceedings of the 1997 IEEE INFOCOM, Kobe, Japan, April, 1997 pp. 1405–1413.
- [6] C.E. Perkins, E. Royer, Ad-hoc on-demand distance vector routing, Proceedings of the Second IEEE Workshop on Mobile Computing Systems and Applications, New Orleans, LA, USA, February, 1999 pp. 90–100.
- [7] C.K. Toh, A novel distributed routing protocol to support ad-hoc mobile computing, Proceedings of the fifteenth IEEE Annual International Phoenix Conference on Computers and Communications, March, 1996 pp. 480–486.
- [8] P. Jacquet, P. Muhlethaler, T. Clausen, A. Laouiti, A. Qayyum, L. Viennot, Optimized link state routing protocol for ad hoc networks, Proceedings of the 2001 IEEE INMIC, December, 2001 pp. 62–68.
- [9] S. Murthy, J.J. Garcia-Luna-Aceves, A routing protocol for packetradio networks, Proceedings of ACM First International Conference on Mobile Computing and Networking, Berkeley, CA, USA, November, 1995 pp. 86–95.
- [10] S. Murthy, J.J. Garcia-Luna-Aceves, An efficient routing protocol for wireless networks, ACM Mobile Networks and Applications, Special Issue on Routing in Mobile Communication Networks 1 (2) (1996) pp. 183–197.
- [11] G. Pei, M. Gerla, T.W. Chen, Fisheye state routing: a routing scheme for ad hoc wireless networks, Proceedings of the 2000 IEEE International Conference on Communications (ICC), New Orleans, LA, June, 2000 pp. 70–74.
- [12] C.E. Perkins, P. Bhagwat, Highly dynamic destination sequenced distance-vector routing (DSDV) for mobile computers, Proceedings of the ACM Special Interest Group on Data Communication, London, UK, September, 1994 pp. 234–244.
- [13] M. Pushpalatha, R. Venkataraman, and T. Ramarao, Trust based energy aware reliable reactive protocol in mobile ad hoc networks, World Academy of Science, Engineering and Technology 56 2009.
- [14] N.-C. Wang, Y.-F. Huang, J.-C. Chen, A stable weight-based on-demand routing protocol for mobile ad hoc networks, Information Sciences 2007 pp 5522–5537.
- [15] N.-C. Wang, S.-W. Chang, A reliable on-demand routing protocol, Computer Communications 2005, pp 123–135.
- [16] N.-C. Wang, C.-Y. Lee, A reliable QoS aware routing protocol with slot assignment for mobile ad hoc network, Journal of Network and Computer Applications Vol. 32, Issue 6, November 2009, Pages 1153–1166.
- [17] R. Patil and A. Damodaram, “Cost Based Power Aware Cross Layer Routing Protocol For Manet”, IJCSNS International Journal of Computer Science and Network Security, VOL.8 No.12, December 2008.
- [18] G. Nishant and D. Samir, “Energy-aware on-demand routing for mobile Ad Hoc networks,” Lecture notes in computer science ISSN: 0302-743, Springer, International workshop in Distributed Computing, 2002.
- [19] M. Tamilarasi, T.G. Palani Velu, “Integrated Energy-Aware Mechanism for MANETs using On-demand Routing”, International Journal of Computer, Information, and Systems Science, and Engineering 2;3 © www.waset.org Summer 2008.
- [20] M. Pushpalatha, Revathi Venkatraman, “Security in Ad Hoc Networks: An extension of Dynamic Source Routing”, 10th IEEE Singapore International conference on Communication Systems Oct 2006, ISBN No:1-4244-0411-8, Pg1-5.
- [21] H. Wu and C. Shi, “A Trust Management Model for P2P File Sharing System”, International Conference on Multimedia and Ubiquitous Engineering, IEEE Explore 78-0-7695-3134-2/08, 2008.
- [22] G. Ghalavand, A. Dana, A. Ghalavand, and M. Reza Hoseini, Reliable routing algorithm based on fuzzy logic for mobile ad hoc networks, International Conference on Advanced Computer Theory and Engineering (ICACTE), 2010.
- [23] V. Rishiwal, A. Kush and S. Verma, Backbone nodes based Stable routing for mobile ad hoc network, UBICC Journal, Vol2, No.3, 2007, pp34-39.

# A Knowledge Driven Computational Visual Attention Model

Amudha J<sup>1</sup>, Soman. K.P<sup>2</sup> and Padmakar Reddy. S<sup>3</sup>

<sup>1</sup> Department of computer science, Amrita School of Engineering  
Bangalore, Karnataka, India

<sup>2</sup> Department of Computational Engineering and Networking, Amrita School of Engineering  
Coimbatore, Tamilnadu, India

<sup>3</sup> Department of computer science, Amrita School of Engineering  
Bangalore, Karnataka, India

## Abstract

Computational Visual System face complex processing problems as there is a large amount of information to be processed and it is difficult to achieve higher efficiency in par with human system. In order to reduce the complexity involved in determining the saliency region, decomposition of image into several parts based on specific location is done and decomposed part is passed for higher level computations in determining the saliency region with assigning priority to the specific color in RGB model depending on application. These properties are interpreted from the user using the Natural Language Processing and then interfaced with vision using Language Perceptual Translator (LPT). The model is designed for a robot to search a specific object in a real time environment without compromising the computational speed in determining the Most Salient Region.

**Keywords:** Visual Attention, Saliency, Language Perceptual Translator, Vision.

## 1. Introduction

Visual attention is a mechanism in human perception which selects relevant regions from a scene and provides these regions for higher-level processing as object recognition. This enables humans to act effectively in their environment despite the complexity of perceivable sensor data. Computational vision systems face the same problem as humans as there is a large amount of information to be processed. To achieve computational efficiency, may be even in real-time Robotic applications, the order in which a scene is investigated must be determined in an intelligent way. The term attention is common in everyday language and familiar to everyone. Visual attention is an important biological mechanism which can rapidly help human to

capture the interested region within eye view and filter out the minor part of image. By means of visual attention, checking for every detail in image is unnecessary due to the property of selective processing. Computational Visual Attention (CVA) is an artificial intelligence for simulating this biometric mechanism. With this mechanism, the difference feature between region centre and surround would be emphasized and integrated in a conspicuity map. Given the complexity of natural language processing and computer vision, few researchers have attempted to integrate them under one approach. Natural language can be used as a source of disambiguation in images since natural language concepts guide the interpretation of what humans can see. Interface between natural language and vision is through a noun phrase recognition systems. A noun phrase recognition system is a system that given a noun phrase and an image is able to find an area in an image where what the noun phrase refers to is located. One of the main challenges in developing a noun phrase recognition system is to transform noun phrases (low level of natural language description) in to conceptual units of a higher level of abstraction that are suitable for image search. The goal is to understand how linguistic information can be used to reduce the complexity of the task of object recognition. However, integrating natural language processing and vision might be useful for solving individual tasks like resolving ambiguous sentences through the use of visual information.

The various related works in the field of computational visual attention model are discussed in Section 2. Section 3 explains the system architecture and Language Processing model. The Section 4 gives the implementation details with analysis of the model followed by conclusion in section 5.

## 2. Related Work

The various models which identify the salient region are analyzed in this section. Frintrop proposed a Visual Attention System for Object Detection and Goal directed search (VOCUS) [1]. Laurent Itti, Christof Koch and Ernst Niebur [5] proposed an algorithm to identify the saliency region in an image using linear filtering. The authors describe in detail how the feature maps for intensity, orientation, and colour are computed. All computations are performed on image pyramids that enable the detection of features on different scales. Additionally, they propose a weighting function for the weighted combination of the different feature maps by promoting feature maps with few peaks and suppressing those with many ones. Simone Frintrop, Maria Klodt and Erich Rome [6] proposed a bottom-up approach algorithm for detection of region of interest (ROI) in a hierarchical way. The method involves smart feature computation techniques based on integral images without compromise on computational speed. Simone Frintrop, Gerriet Bracker and Erich Rome [2] proposed an algorithm where both top-down and bottom-up approaches are combined in detection of ROI by enabling the weighting of features. The weights are derived from both target and back ground properties. The task is to build a map of the environment and to simultaneously stay localized within the map which serves as visual landmarks for the Robot. Simone Frintrop and Markus Kessel proposed a model for Most Salient Region tracking [10] and Ariadna Quattoni [3] has proposed a model for detection of object using natural language processing, which is used in system discussed here.

In psychophysics, top-down influences are often investigated by so called cuing experiments. In these experiments, a “cue” directs the attention to the target. Cues may have different characteristics: they may indicate *where* the target will be, or *what* the target will be. A cue speeds up the search if it matches the target exactly and slows down the search if it is invalid. Deviations from the exact match slow down search speed, although they lead to faster speed compared with a neutral cue or a semantic cue. This is the main motivation behind integrating the verbal cues to the attention model to enhance the search speed which is experimentally verified.

## 3. System Architecture

The block diagram in Fig.1 describes the flow of the system. The system architecture describes two major modules. 1) Language Perceptual Translator (LPT) [3] 2) Visual Attention Model (VAM) [1, 4, 7, 8, 9].

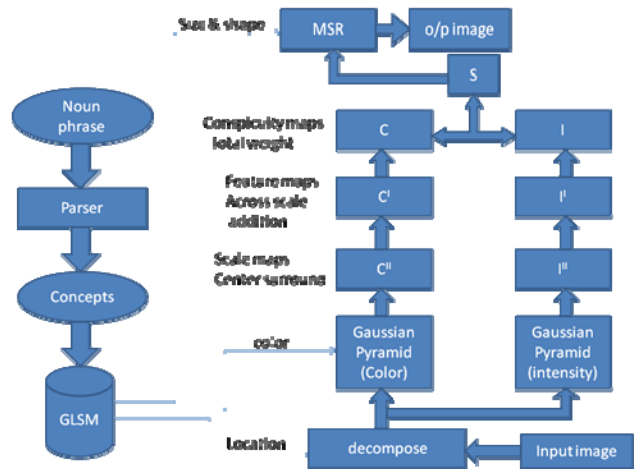


Fig. 1 Visual Attention Model with NLP.

1) LPT: One of the main challenges in developing a noun phrase recognition system is to transform noun phrases (low level of natural language description) into conceptual units of a higher level of abstraction that are suitable for image search. That is, the challenge is to come up with a representation that mediates between noun phrases and low-level image input. The Parser processes the sentence and it outputs the corresponding properties like location, Color, Size, Shape and for the Thing (object). We must construct a “grounded” lexicon semantic memory that includes perceptual knowledge about how to recognize the things that words refer to in the environment. A “grounded” lexical semantic memory would therefore connect concepts to the physical world enabling machines to use that knowledge for object recognition. A GLSM (Grounded Lexical Semantic Memory) is a data-structure that stores knowledge about words and their relationships. Since the goal of LPT is to transform a noun-phrase into perceptual constraints that can be applied to visual stimuli to locate objects in an image. The outputs of GLSM is given to the VAM at different processing levels like location property at decomposition level, Color property at Gaussian pyramid construction and Size and Shape property after detecting of salient region to identify the required object in an image.

2) The Visual Attention model (VAM) identifies the most attended region in the image. The following sections present the algorithm in detail.

### 3.1 Visual Attention Model

The 1<sup>st</sup> level of bottom-up visual attention shown in fig.1 is decomposition of an image based on location property. We divided the image based on index method as shown in Fig. 2 as Top, Left, Right, etc.



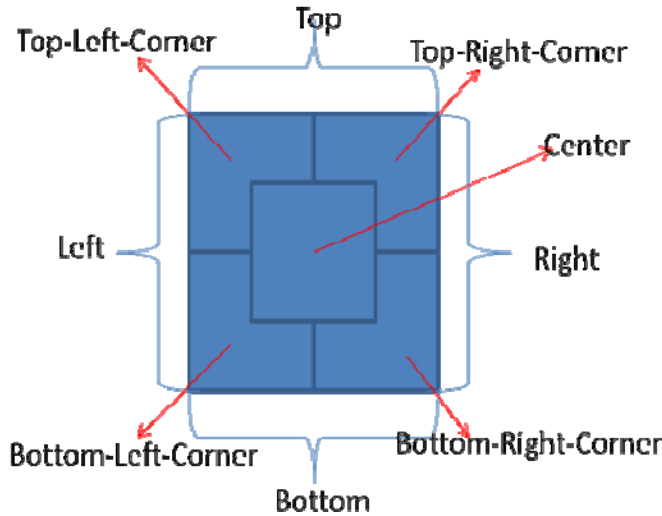


Fig. 2 Dividing Image by Index Method.

The image  $I$  is divided into 9 different parts and the default option is an entire image. Here the location cue property determines the search region to detect an object which reduces the possibility to shift the focus of attention to other objects in an entire image due to intensity or color is reduced when we crop the image based on location property.

In our approach we used to detect the sign boards which uses the prior knowledge of location has Top-Left-Corner or Top-Right-Corner. Before decomposing the image based on location cue matrix is converted in to  $N \times N$  square Matrix by resizing the image  $I$ .  $I$  is divided in to 9 parts with different Location Cues are shown in Fig.3.

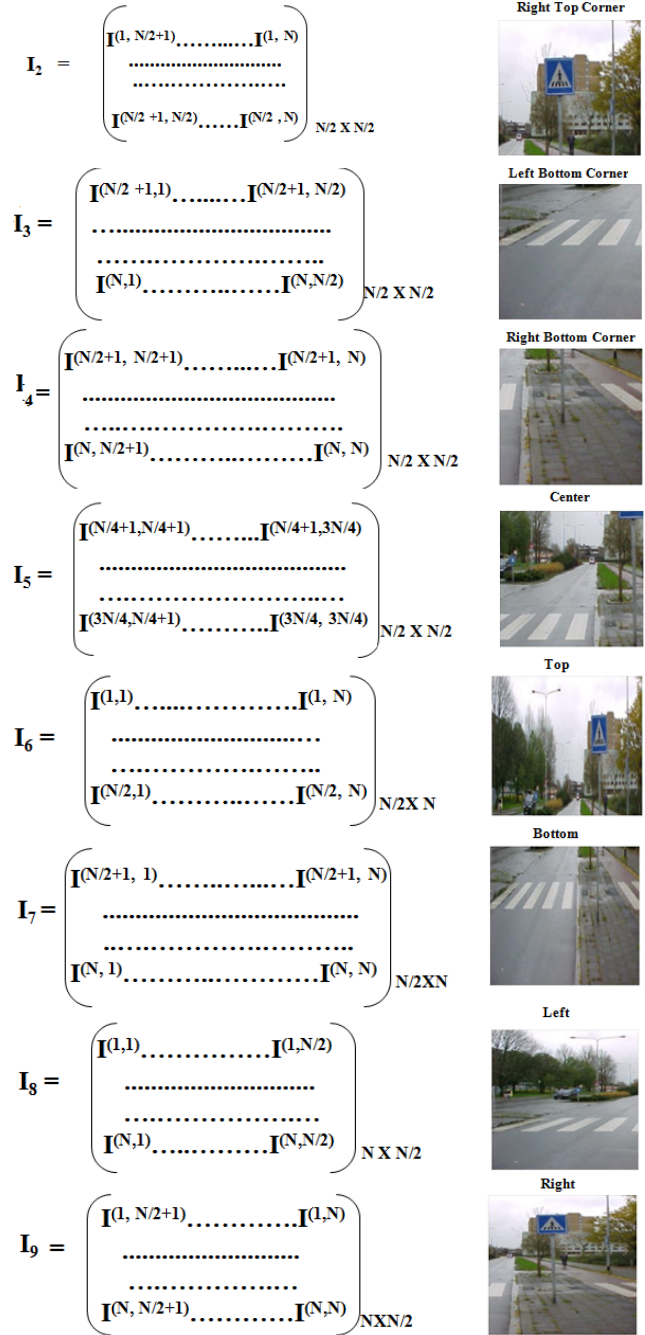
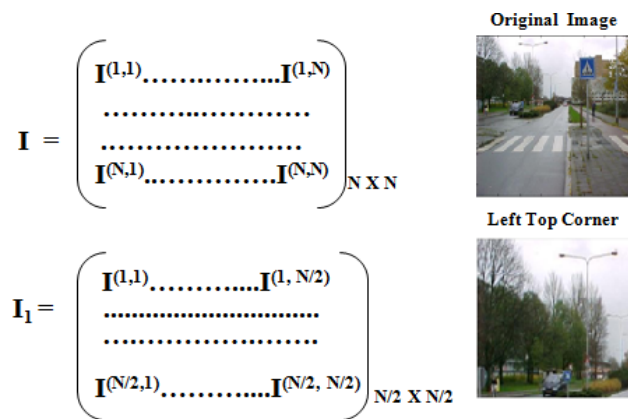


Fig. 3 Different Locations of Image (I) with respective Matrices.

The input image  $I$  is sub-sampled into a Gaussian pyramid on 4 different scales, and each pyramid level is decomposed into channels for red(R), green (G), blue (B), yellow (Y), intensity (I) using (1), (2), (3), (4) and (5) .



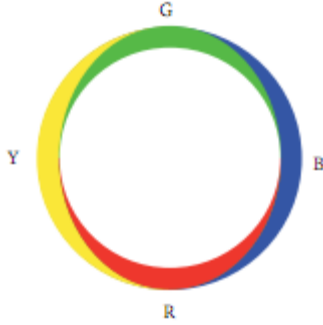


Fig. 4 Opponent colors.

$$I = (r + g + b)/3 \quad (1)$$

$$R = r - (g + b)/2 \quad (2)$$

$$G = g - (r + b)/2 \quad (3)$$

$$B = b - (r + g)/2 \quad (4)$$

$$Y = r + g - 2(|r - g| + b) \quad (5)$$

Depending on color property cue from the GLSM the priority of which color is high and which is low is set on different color channels of red(R), green (G), Blue (B), and Yellow(Y). The Color opponent process is a color theory that states that the human visual system interprets information about color by processing signals from cones and rods [in an antagonistic manner]. Opponency is thought to reduce redundant information by de-correlating the photoreceptor signals. It suggests that there are three opponent channels Red Vs Green, Blue Vs Yellow, Dark Vs White. Response to one color of an opponent channel are antagonistic to those to the other color, i.e. one color produces an excitatory effect and the other produces an inhibitory effect, the opponent colors are never perceived at the same time (the visual system can't be simultaneously excited and inhibited).The decision on which color channel to be used is based on the color cue. The output of the feature maps are then fed to the center-surround. These 5 channels are fed to the center surround differences after resizing all the surround images to the center image. Center-Surround operations are implemented in the model as difference between a fine and a coarse scale for a given feature. The center of the receptive feature corresponds to the pixel at the level  $c \in \{2, 3\}$  in the pyramid and the surround corresponds to the pixel at the level  $s = c+1$ . Hence we compute three feature maps in general case. One feature type encodes for on/off image intensity contrast, two encodes for red/green and blue/yellow double component channels. The intensity feature type encodes for the modulus of image luminance contrast. That is the absolute value of the difference

between the intensity at the center and the intensity in the surround as given in (6).

$$I''_{(I,C,S)} = N(|I(c) \ominus I(s)|) \quad (6)$$

The quantity corresponding to the double opponency cells in primary visual context are then computed by center surround differences across the normalized color channels. Each of the three-red /green Feature map is created by first computing (red-green) at the center, then subtracting (green-red) from the surround and finally outputting the absolute value. Accordingly maps  $RG(c,s)$  are created in the model to simultaneously account for red/green and green/red double opponency and  $BY(c,s)$  for blue/yellow and yellow/blue double opponency using (7) and(8).

$$C^I_{RG,C,S} = N(|R(c) - G(c) \ominus (R(s) - G(s))|) \quad (7)$$

$$C^I_{BG,C,S} = N(|B(c) - Y(c) \ominus (B(s) - Y(s))|) \quad (8)$$

The feature maps are then combined into two conspicuity maps, intensity  $\bar{I}$  (9), color  $\bar{C}$  (10), at the saliency map's scale ( $\sigma=4$ ). These maps are computed through across-scale addition ( $\oplus$ ), where each map is reduced to scale four and added point-by-point:

$$I = \oplus_{c=2}^4 \oplus_{s=c+3}^{c+4} N(I(c,s)) \quad (9)$$

$$C = \oplus_{c=2}^4 \oplus_{s=c+3}^{c+4} [N(RG(c,s)) + N(BY(c,s))] \quad (10)$$

The two conspicuity maps are then normalized and summed into the input  $S$  to the saliency map (11).

$$S = (N(I) + N(C)) \quad (11)$$

The  $N(\cdot)$  represents the non-linear Normalization operator. From the saliency map the most attention region is identified by finding the maximum pixel value in the salient region. The identification of the segmented region can be made based on size and shape property.

## 4. Results and Analysis

The system developed is tested on a dataset where the attention object is a signboard. The various signs in the dataset are bike, crossing and pedestrian symbols. The number of testing samples used for analysis is as shown in Table 1. The cues that are used in the dataset are the location cues, the color cue, the size and shape cue pertaining to the object signboard. In table 2 the verbal cues that mostly suit for the chosen dataset is shown.

Table 1: Testing samples for signboard detection

Type of Image	Total No. of Images
Bike	16
Crossing	16
Pedestrian	16

Table 2: Cues for data set

Location	Color	Size	Shape	Thing
Right top Corner	Red	Large/Small	Triangle	Sign board
Right top Corner	Blue	Large/Small	Rectangle	Sign board
Right top Corner	Blue	Large/Small	Circle	Sign board

The analysis is done with and without cues. Visual attention model without cues has  $N \times N$  i.e.  $N^2$  computations at each level, where as with cues depending on Location Property the number of computations is reduced to  $N^2/4$  or  $N^2/2$  at each level to get Region of Interest. Priority for color is chosen by trial and error method with different combinations of inhibiting and exhibiting channels. The system developed is tested under various cases scenarios like

- No verbal cues are given to the system.
- Only the color property is obtained.
- Only the location (region information available).
- Both color and location information.

VAM is tested and compared with the different combination of cues like only color, only location, both color and location and without cues as shown in Table 3.

Table 3: VAM with different combinations of Cues.

Images	Total No. of Images	No. of images Detected with different combinations			
		No Cues	Only Color	Only Location	Both Color and Location
Bike	16	3	10	4	15
Crossing	16	8	7	9	12
pedestrian	16	3	15	5	15

In Table 4 VAM is tested with both location and color cues for the same data set with varying the color priority. The VAM decides excite the weights to frame channels to enhance the color information in the image in the following ways. For identifying the Red color Signboards.

- Increment Red and decrement Green component by a factor of 0.5.
- Increment Red and Green component by a factor of 0.7.
- Increment Red component by 0.7 and decrement Green, Blue, and Yellow components by a factor of

0.3.

- Double the Red component and decrement Green, Blue and Yellow by a factor of 0.3.

For identifying the Blue color Signboards replace the Red color with Blue and Blue color with the Red and repeat the above 4 steps and the same as shown in Table 4.

Table 4: Testing sign board data set with different Priority levels.

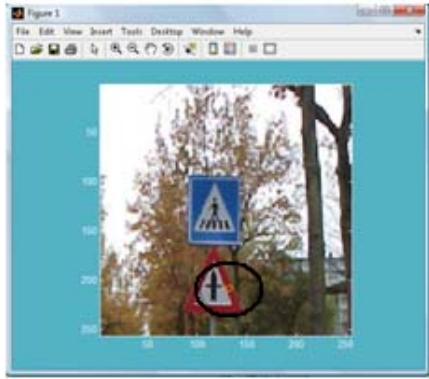
Images	No. of Correctly detected images with color priority.			
	R_i & G_d by 50%	R_i by 70% & G_d by 30%	R_i by 70% & (G,B,Y)_d by 30%	Double R & (G, B,Y)_d by 30%
Crossing Priority RED color	6	4	10	12
	B_i & Y_d by 50%	B_i by 70% & Y_d by 30%	B_i by 70% & (R,G,Y)_d by 30%	Double B & (R,G,Y)_d by 30%
Bike Priority BLUE color	10	13	12	15
Pedestrian Priority BLUE color	10	12	13	14

The Symbol's R/B/G/Y\_i indicates Red/Blue/Green/Yellow color priority increased and R/G/B/Y\_d indicates Red/Green/Blue/Yellow color priority decreased.

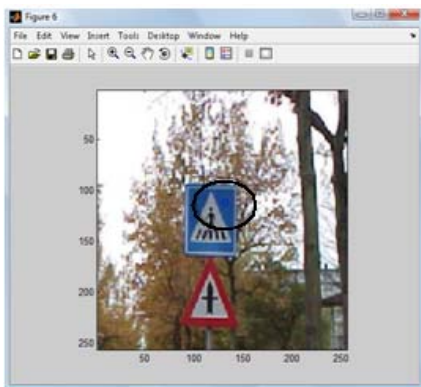
To the VAM system the input Sign board image shown in Fig.4 (a) is given as input to the VAM and input to the LPT is noun phrase which is "Find the Red color Sign board on "Right\_top\_corner". So, here the desired color cue is Red, location cue is Right\_top\_corner and the object is Sign board. The result of VAM is shown in Fig.4 (b) and when the color cue is Blue is shown in Fig.4(c). The performance with different priority levels shown in Table 4 and for the same color cue is shown in Fig (5).



(a)



(b)



(c)

Fig.5 Image with both Crossing and pedestrian sign boards (a) Input Image to the system (b) Output of VAM with Color and Location cues. (c) Result of VAM with Color and Location cues.

In Fig. 6 **Type 1** indicates, increment R, B and decrement G, Y by a factor of 0.5. **Type 2** indicates, increment R, B by a factor of 0.7 and decrement G, Y by a factor of 0.3. **Type 3** indicates, increment R/B by a factor of 0.7 decrement G, B/R, Y by a factor of 0.3. **Type 4** indicates double R/B component and decrement G, B/R, Y by a factor of 0.3.

The **Type 4** system performance is much better than other systems, hence the system assigns color cue weightage based on Type 4. Comparison between the various visual attention models on computation of the number of maps computed for identifying the salient region is shown in Table 5. The statistics clearly depict the computation of the map which is less in case of VAM in comparison with VOCUS and Itti's Model. In case of VAM with verbal cue color it is only 52 maps. In case of VAM with location cue the computation of the number of maps remains the same but the image size is reduced to half or quarter of the original size which reduces the time taken for computation.

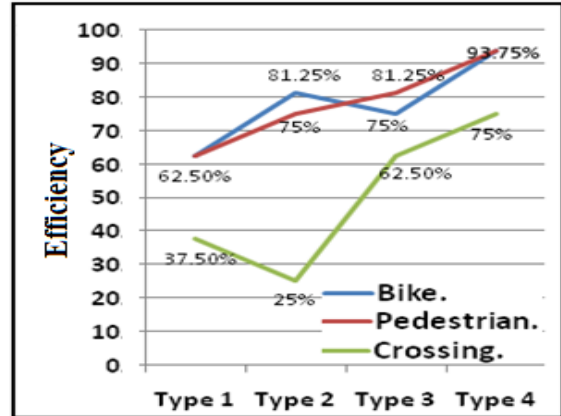


Fig.6 Performance with different sign board images and with different types of priority color Cues.

Table.5. Comparison of Maps in various models

Various Maps in the Architecture	Various Visual Attention Architecture Maps at different levels.				
	Itti's	VOCUS	VAM	VAM with verbal cue color	VAM with verbal cue location
Pyramid Maps	24	28	45	30	45
Scale Maps	42	48	14	12	14
Feature Maps	7	10	7	6	7
Conspicuity Maps	3	3	3	3	3
Saliency Map	1	1	1	1	1
Total Maps	77	100	65	52	65

Comparison between the various visual attention models on computation of the number of maps computed for identifying the salient region is shown in table5. The statistics clearly depict the computation of the map which is less in case of VAM in comparison with VOCUS and Itti's Model. In case of VAM with verbal cue color it is only 52 maps. In case of VAM with location cue the computation of the number of maps remains the same but the image size is reduced to half or quarter of the original size which reduces the time taken for computation

## 5. Conclusion

The computation of saliency region is determined with and without decomposing the image and the time taken to compute the most salient region with decomposition takes less time in comparison without decomposition. The verbal cue also reduces the number of maps computed for determining the saliency. The other cues for size and shape which reduces the time taken to identify the object hasn't been implemented and is left for future scope of the system. The various other issues like

combination of the verbal cues which will result in a flexible architecture for visual attention has to be studied extensively with a language interface.

## 6. References

- [1] Frintrop, S. VOCUS: A Visual Attention System for Object Detection and Goal directed Search. PhD thesis Rheinische Friedrich-Wilhelms-University at Bonn Germany (2005). Published 2006 in Lecture Notes in Artificial Intelligence (LNAI), Vol. 3899, Springer Verlag Berlin/ Heidelberg.
- [2] Frintrop, S., Backer, G. and Rome, E. Goal-directed Search with a Top-down Modulated Computational Attention System. In: Proc. of the Annual meeting of the German Association for Pattern Recognition DAGM 2005 Lecture Notes in Computer Science (LNCS) Springer (2005) 117–124.
- [3] Ariadna Quattoni, Using Natural Language Descriptions to aid object Recognition. PhD thesis University of Massachusetts, Amherst Massachusetts, 2003.
- [4] Frintrop, S., Jensfelt, P. and Christensen, H. Attentional Landmark selection for Visual SLAM. In: Proc. of the International Conference on Intelligent Robots and Systems (IROS '06) (2006).
- [5] Itti, L., Koch, C. and Niebur, E. A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20 (11, 1998) 1254–1259.
- [6] Simon Frintrop, Maria Klodt, and Erich Rome. A Real time Visual Attention System Using Integral Images, in proc of the 5<sup>th</sup> international conference on ICVS 2007, Bielefeld, Germany, March 2007.
- [7] Wei-song Lin and Yu-Wei Huang. Intention-oriented Computational Visual Attention Model for learning and seeking image Content. Department of Electrical engineering National Tiwan University . 2009 IEEE Transaction.
- [8] Simone Frintrop, Patric jensfelt and Henrik Christensen. Attentional Robot Localization and Mapping at the ICVS Workshop on Computational Attention and Applications, (WCAA), Bielefeld, Germany, March 2007.
- [9] Cairong Zhao, ChuanCai Liu, Zhihui Lai, Yue Sui, and Zuoyong Li. Sparse Embedding Visual Attention Model IEEE Transaction 2009.
- [10] Simone Frintrop and Markus Kessel, "Most Salient Region Tracking", IEEE 2009 international conference on Robotics and Automation (ICRA'09), Kobe, Japan, May 2009.

and Electronics Engineering from Bharathiyar University, Coimbatore, India in 1998 and M.E. Computer Science and Engineering , Anna University, Chennai, India in 2003. Her research interests include image processing, computer vision and soft computing.

**Dr. K.P Soman** is the head, CEN, Amrita Vishwa Vidyapeetham Amrita Vishwa Vidyapeetham, Ettimadai, Coimbatore-641105. His qualifications include B.Sc. Engg. in Electrical engineering from REC, Calicut.P.M. Diploma in SQC and OR from ISI, Calcutta.M.Tech (Reliability engineering) from IIT, KharagpurPhD (Reliability engineering) from IIT, Kharagpur.Dr. Soman held the first rank and institute silver medal for M.Tech at IIT Kharagpur. His areas of research include optimization, data mining, signal and image processing, neural networks, support vector machines, cryptography and bio-informatics. He has over 55 papers in national and international journals and proceedings. He has conducted various technical workshops in India and abroad.

**Padmakar Reddy.S** is a Post graduate student in Amrita School of Engineering, Bangalore, Karnataka. His qualifications include B.Tech. in Electronics and Communication Engineering in Madanapalli Institute of Technology & Sciences, Madanapalli, Andhra Pradesh, India. His research interests include image processing and Embedded Systems.

**Amudha Joseph** is an assistant Professor in Amrita School of Engineering, Bangalore. Her qualifications include B.E., Electrical



# A Frame Work for Frequent Pattern Mining Using Dynamic Function

Sunil Joshi<sup>1</sup>, R S Jadon<sup>2</sup> and R C Jain<sup>3</sup>

<sup>1</sup> Computer Applications Department, Samrat Ashok Technological Institute  
Vidisha, M.P. , India

<sup>2</sup> Computer Applications Department, Madhav Institute of Technology and Science  
Gwalior, M.P. , India

<sup>3</sup> Computer Applications Department, Samrat Ashok Technological Institute  
Vidisha, M.P. , India

## Abstract

Discovering frequent objects (item sets, sequential patterns) is one of the most vital fields in data mining. It is well understood that it require running time and memory for defining candidates and this is the motivation for developing large number of algorithm. Frequent patterns mining is the paying attention research issue in association rules analysis. Apriori algorithm is a standard algorithm of association rules mining. Plenty of algorithms for mining association rules and their mutations are projected on the foundation of Apriori Algorithm. Most of the earlier studies adopted Apriori-like algorithms which are based on generate-and-test candidates theme and improving algorithm approach and formation but no one give attention to the structure of database. Several modifications on apriori algorithms are focused on algorithm Strategy but no one-algorithm emphasis on least transaction and more attribute representation of database. We presented a new research trend on frequent pattern mining in which generate Transaction pair to lighten current methods from the traditional blockage, providing scalability to massive data sets and improving response time. In order to mine patterns in database with more columns than rows, we proposed a complete framework for the frequent pattern mining. A simple approach is if we generate pair of transaction instead of item id where attributes are much larger then transaction so result is very fast. Newly, different works anticipated a new way to mine patterns in transposed databases where there is a database with thousands of attributes but merely tens of stuff. We suggest a novel dynamic algorithm for frequent pattern mining in which generate transaction pair and for generating frequent pattern we find out by longest

common subsequence using dynamic function. Our solutions give result more rapidly. A quantitative investigation of these tradeoffs is conducted through a wide investigational study on artificial and real-life data sets.

**Keywords:** *Longest Common Subsequence, Frequent Pattern mining, dynamic function, candidate, transaction pair, association rule, vertical mining*

## 1. Introduction

Frequent Pattern Mining is most dominant problem in association mining. Plenty of algorithms for mining association rules and their mutations are projected on the foundation of Apriori Algorithm. Most of the earlier studies adopted Apriori-like algorithms which are based on generate-and-test candidates theme and improving algorithm approach and formation but always focus on item id instead of transaction id. Several modifications on apriori algorithm are focused on algorithm Strategy but no one-algorithm emphasis on least transaction more attribute representation of database.

Most of the preceding work on mining frequent patterns is based on the horizontal representation. However, recently a number of vertical mining algorithms have been projected for mining frequent itemsets. Mining algorithms using the vertical representation have shown to be effective and usually do better than horizontal approaches [11]. This benefit stems from the fact that frequent patterns can be counted via tidset intersections, instead of using complex interior data structures like the hash/search trees that the horizontal algorithms need [10]. Also in the vertical mining, the candidate creation and counting phases are done in a single

step. This is done because vertical mining offers usual pruning of unrelated transactions as a result of an intersection. Another characteristic of vertical mining is the utilization of the autonomy of classes, where each frequent item is a class that contains a set of frequent  $k$ -itemsets (where  $k > 1$ ) [6]. The vertical arrangement appears to be a usual choice for achieving association rule mining's purpose of discovering associated items. Computing the supports of itemsets is simpler and quicker with the vertical arrangement since it involves only the intersections of tid-lists or tid-vectors, operations that are well-supported by the current database systems. In difference, complex hash-tree data structures and functions are required to perform the same function for flat layouts. There is an automatic reduction of the database before each scan for those itemsets that are significant to the following scan of the mining process are accessed from disk. In the horizontal arrangement, however, irrelevant information that happens to be part of a row in which useful information is present is also transferred from disk to memory. This is because database reductions are moderately hard to implement in the horizontal arrangement. Further, still if reductions were possible, the irrelevant information can be removed only in the scan following the one in which its irrelevance is exposed. Therefore, there is always a reduction delay of at least one scan in the horizontal layout.

A simple approach is if we generate pair of transaction instead of item id where attributes are much larger than transaction then result is very fast. Recently, different works proposed a new way to mine patterns in transposed databases where a database with thousands of attributes but only tens of objects [15]. In this case, mining the transaction pair runs through a smaller search space. None algorithm filters or reduces the database in each pass of apriori algorithm to count the support of prune pattern candidate from database. Most of the preceding work on vertical mining concentrates on intersection of transaction [12]. This is based on intersection of perpendicular tid-vector where it is a set of columns with each column storing an IID and a bit-vector of 1's and 0' to represent the occurrence or nonexistence, respectively, of the item in the set of customer transactions. If we use list-based layout then it takes much less space than the bit-vector approach (which has the overhead of openly representing absence) in sparse databases. We make the case in this paper and use list-based layout [16]. To find intersection we use dynamic technique instead of traditional approach. We suggest a novel dynamic algorithm for frequent pattern mining in which we generate transaction pair and for generating frequent pattern we find out by longest common subsequence using dynamic function.

The rest of this paper is structured as follows. Section II introduces the problem and reviews some efficient related works. The projected method is described in section III. Section IV explains in details the projected FPMDF

algorithm. A justification with Example is given in Section V. The investigational results and assessment show in section VI. Finally Section VII contains the conclusions and upcoming works

## 2. Frequent Pattern Mining

Frequent Itemset Mining came from efforts to determine valuable patterns in customers' transaction databases. A customers' transaction database is a series of transactions ( $T = t1. . . tn$ ), where each transaction is an itemset ( $t_i \subseteq I$ ). An itemset with  $k$  elements is known as  $k$ -itemset. In the rest of the paper we make the (practical) assumption that the items are from a prearranged set, and transactions are stored as sorted itemsets. The support of an itemset  $X$  in  $T$ , denoted as  $\text{supp}T(X)$ , is the number of those transactions that hold  $X$ , i.e.  $\text{supp}T(X) = |\{t_j : X \subseteq t_j\}|$ . An itemset is frequent if its support is larger than a support threshold, originally denoted by  $\text{min supp}$ . The frequent itemset mining problem is to discover all frequent itemset in a given transaction database.

The primary Algorithm Proposed for finding frequent itemsets, is the APRIORI Algorithm [1]. This algorithm was enhanced later to obtain the frequent pattern quickly [2]. The Apriori algorithm employs the downhill closure property—if an itemset is not frequent, any superset of it cannot be frequent either. The Apriori Algorithm performs a breadth-first search in the search Space by generating candidate  $k+1$  itemsets from frequent  $k$ -itemsets. The occurrence of an itemset is computed by counting its happening in each transaction. Numerous variants of the Apriori algorithm have been developed, like AprioriTid, AprioriHybrid, direct hashing and pruning (DHP), Partition algorithm, dynamic itemset counting (DIC) etc.[3]. FP-growth [4] is a well-known algorithm that uses the FP-tree data structure to get a condensed representation of the database transactions and employs a divide-and conquer approach to decompose the mining problem into a set of smaller problems. In spirit, it mines all the frequent itemsets by recursively determining all frequent 1-itemsets in the restrictive pattern base that is proficiently constructed with the help of a node link structure. In algorithm FP-growth-based, recursive production of the FP-tree affects the algorithm's complexity. Most of the preceding work on association mining has utilized the conventional horizontal transactional database arrangement. However, a number of vertical mining algorithms have been proposed recently for association mining [5, 6, 9, 11, 12]. In a vertical database each item is associated with its equivalent tidset, the set of all transactions (or tids) where it appears. Mining algorithms using the vertical format have shown to be very valuable and usually do better than horizontal approaches. This advantage stems from the fact that frequent patterns can be counted via tidset intersections, instead of using complex internal data



structures (candidate generation and counting happens in a single step). The horizontal approach on the other hand needs complex search/hash trees. Tidsets offer ordinary pruning of extraneous transactions as a result of an intersection (tids not relevant drop out). Furthermore, for databases with lengthy transactions it has been shown using a simple cost model, that the vertical approach reduces the number of I/O operations [7]. In a current study on the integration of database and mining, the Vertical algorithm [8] was shown to be the best approach (better than horizontal) when forcefully integrating association mining with database systems. Eclat [9] is the primary algorithm to find frequent patterns by a depth-first search and it has been shown to execute fine. They use vertical database representation and count the support of itemset by using the intersection of tids. However, pruning used in the Apriori algorithm is not applicable during the candidate itemsets generation due to depth-first search. VIPER [5] uses the vertical database layout and the intersection to accomplish an excellent performance. The only difference is that they use the compacted bitmaps to represent the transaction list of each itemset. However, their compression method has limitations especially when tids are uniformly distributed. Zaki and Gouda [10] developed a new approach called dEclat using the vertical database representation. They store the difference of tids, called diffset, between a candidate k-itemset and its prefix k-1 frequent itemsets, instead of the tids intersection set, denoted here as tidset. They calculate the support by subtracting the cardinality of diffset from the support of its prefix k-1 frequent itemset. This algorithm has been exposed to gain significant performance improvements over Eclat. However, diffset will drop its advantage over tidset when the database is sparse.

Most of the preceding work on mining frequent patterns is based on the horizontal illustration. However, recently a number of vertical mining algorithms have been projected for mining frequent itemsets. Mining algorithms using the vertical representation have exposed to be effective and usually do better than horizontal approaches [11]. This advantage stems from the fact that frequent patterns can be counted via tidset intersections, instead of using complex internal data structures like the hash/search trees that the horizontal algorithms require [10]. The candidate generation and counting phases are done in a single step in vertical mining. This is done because vertical mining offers ordinary pruning of irrelevant transactions as a result of an intersection.

Another characteristic of vertical mining is the utilization of the autonomy of classes, where each frequent item is a class that contains a set of frequent k-itemsets (where  $k > 1$ ) [6]. The vertical arrangement appears to be a natural choice for achieving association rule mining's objective of discovering correlated items. Computing the supports of itemsets is simpler and faster with the vertical arrangement since it

involves only the intersections of tid-lists or tid-vectors, operations that are well-supported by existing database systems. In contrast, complex hash-tree data structures and functions are required to perform the same function for horizontal layouts. There is an automatic reduction of the database before each scan in that only those itemsets that are significant to the following scan of the mining process are accessed from disk. In the horizontal layout, however, irrelevant information that happens to be part of a row in which useful information is present is also transferred from disk to memory. This is because database reductions are comparatively hard to implement in the horizontal arrangement. Further, even if reduction were promising, the irrelevant information can be removed only in the scan following the one in which its irrelevance is discovered. Therefore, there is always a reduction delay of at least one scan in the horizontal layout.

Most of the preceding work on vertical mining concentrates on intersection of transaction [12]. This is based on intersection of perpendicular tid-vector where it is a set of columns with each column storing an IID and a bit-vector of 1's and 0' to represent the occurrence or nonexistence, respectively, of the item in the set of customer transactions. If we use list-based layout then it takes much less space than the bit-vector approach (which has the overhead of openly representing absence) in sparse databases. We make the case in this paper and use list-based layout [16]. To find intersection we use dynamic technique instead of traditional approach. We suggest a novel dynamic algorithm for frequent pattern mining in which we generate transaction pair and for generating frequent pattern we find out by longest common subsequence using dynamic function

### 3. Dynamic Function

The longest common subsequence problem is one of the frequent problems which can be solved powerfully using dynamic programming. "The Longest common subsequence problem is, we are given two sequences  $X = \langle x_1, x_2, \dots, x_n \rangle$  and  $Y = \langle y_1, y_2, \dots, y_m \rangle$  and wish to find a maximum length

common subsequence of X and Y" for example : if  $X = \langle A, B, C, B, D, A, B \rangle$  and  $Y = \langle B, D, C, A, B, A \rangle$  then The sequence  $\langle B, C, B, A \rangle$  longest common subsequence. Let us define  $CC[i, j]$  to be the length of an LCS of the sequences  $x_i$  and  $y_j$ . If either  $i=0$  or  $j=0$ , one of the sequence has length 0, so the LCS has length 0. The Optimal substructure of the LCS Problem gives the recursive formula in fig.1

$$C(i, j) = \begin{cases} 0 & \text{if } i = 0 \text{ or } j = 0 \\ C(i-1, j-1)+1 & \text{if } i, j > 0 \text{ and } x_i = y_j \\ \max(c(i, j-1), c(i-1, j)) & \text{if } i, j > 0 \text{ and } x_i \neq y_j \end{cases}$$

Figure 1. Longest Common Subsequence Recursive Formula

records, it means that an itemset that is supported by at least two transactions is a frequent set and output shown in fig.2

#### 4. Algorithm

The Novel algorithm works over the entire database file, now apply Apriori like Algorithm in which first we generate transaction pair with longest common subsequence of item id instead of item id pair. For each Iteration we apply following sequence of operation until condition occurred. First generate the transaction pair and prune with empty longest common subsequence by dynamic function. To count the support , instead of whole database for each pruned pattern we find all subset and display it and also stored new transaction pair and its attribute common subsequence so that next iteration we trace above subsequence. To find longest common subsequence we used dynamic function which faster then traditional function. Write pruned transaction pair list with attribute common subsequence so that in next pass we used this pair list instead of all pair list. An advantage of This approach is in each iteration database filtering and reduces, so each iteration is faster then previous iteration

#### Algorithm FPMDF ( Frequent Patterns Mining Using Dynamic Function)

- I. Given Database T with  $\partial$ (Min. no. of transaction)
- II. K: =2.
- III. While Lk-1 $\neq$  { } do
- IV. Ck=Compute each pair of each previous transaction pair .
- V. Computer LCS of Item id for each previous transaction pair.
- VI. Lk=Prune Transaction Pair having empty LCS.
- VII. If  $\partial \leq k$  then Fk=All\_Subset(Lk)
- VIII. K:=K+1

#### 5. Explanation with example which support the arguments

Study the following transaction database .T={T1,T2,T3,T4,T5 >, Assume  $\sigma=40\%$ , Since T contains 5

TABLE I. GIVEN DATASET T (C1)

TI d	Attributes (Item Id)
1	1,2,5,6,7,9,10,15
2	1,3,14
3	2,3,5,6,7,8,9,12
4	4,10,15
5	2,4,5,7,9,11,13

Now Apply Algorithm

#### Iteration 1

Generate Transaction Pair with two elements with Longest Common Subsequence (LCS) By Dynamic Function of Attributes

TABLE II. C2

TId	Attributes (Item Id)
1,2	1
1,3	2,5,6,7,9
1,4	10,15
1,5	2,5,7,9
2,3	3
2,4	NIL
2,5	NIL
3,4	NIL
3,5	2,5,7,9
4,5	4

Prune C2 by removing Transaction pair having Empty LCS of attributes.

```

C:\WINDOWS\system32\command.com
D:\TF>java TF tp.dat 60
Total Transaction is 5
Maximum Frequent Items are -----2 5 7 9,
Execution time is: 31ms
D:\TF>
    
```

Figure 2. Frequent Pattern with support 60%

TABLE III. L2

TId	Attributes (Item Id)
1,2	1
1,3	2,5,6,7,9
1,4	10,15
1,5	2,5,7,9
2,3	3
3,5	2,5,7,9
4,5	4

If  $\sigma=40\%$ , frequent set support record=2 then

$$F2 = \text{All\_Subset}(L2)$$

$$F2 := \{ 1,2, 3, 4, 5, 6, 7,9,10,15,(2,5),(2,6),(2,7),(2,9),(5,6),(5,7),(5,9),(6,7),(6,9),(7,9),(10,15),(2,5,6),(2,5,7),(2,5,9),(2,6,7),(2,6,9),(2,7,9),(2,5,6,7),(2,5,6,9),(2,5,7,9), (2,6,7,9),(5,6,7,9),(2,5,6,7,9) \}$$

### Iteration 2

Generate Transaction Pair with 3 elements with Longest Common Subsequence (LCS) By Dynamic Function of Attributes

TABLE IV. C3

TId	Attributes (Item Id)
1,3,5	2,5,7,9

1,2,3	NIL
1,2,4	NIL
1,2,5	NIL
1,3,4	NIL
1,3,5	2,5,7,9
1,4,5	NIL

Prune C2 by removing Transaction pair having Empty LCS of attributes.

TABLE V. L3

TId	Attributes (Item Id)
1,3,5	2,5,7,9

If  $\sigma=40\%$ , frequent set support record=2 then

$$F3 = F2 \cup \text{All\_Subset}(L3)$$

$$F3 := \{ 1,2,3,4,5,6, 7,9,10,15,(2,5),(2,6),(2,7),(2,9),(5,6),(5,7),(5,9),(6,7),(6,9),(7,9),(10,15),(2,5,6),(2,5,7),(2,5,9),(2,6,7),(2,6,9),(2,7,9),(2,5,6,7),(2,5,6,9),(2,5,7,9),(2,6,7,9),(5,6,7,9),(2,5,6,7,9) \}$$

If  $\sigma=60\%$ , frequent set support record=3 then

$$F3 = \text{All\_Subset}(L3)$$

$$F3 := \{ 2,5, 7,9,(2,5), ,(2,7),(2,9), (5,7),(5,9),(7,9), (2,5,7),(2,5,9), ,(2,7,9),(5,7,9),(2,5,7,9), \}$$

## 6. Experimental results

In this section we performed a set of experiments to evaluate the effectiveness of the frequent pattern mining using dynamic function method. The algorithm DFPMT was executed on a Pentium 4 CPU, 2.26GHz, and 1 GB of RAM computer. It was implemented in Java. The experiment database sources are T40I4D100K, provided by the QUEST generator of data generated from IBM's Almaden lab. The experimental dataset contains data whose records are set to 10. The testing results of experiments are showed in Fig.3. In the Fig.3, the horizontal axis represents the number of support in database and the vertical axis represents mining time. The three curves denote different time cost of the algorithm Apriori, FP Growth and FPMDF with different minsup.

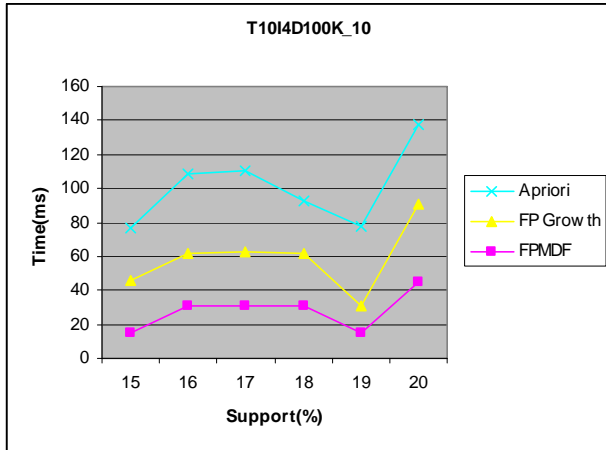


Figure 3. The test results of apriori, FP Growth and DFPMT

## 6. Conclusion

Discovering frequent objects (item sets, sequential patterns) is one of the most vital fields in data mining. It is well understood that it requires running time and memory for defining candidates and this is the motivation for developing large number of algorithms. We presented a new research trend on frequent pattern mining in which if the number of transactions are very less as compared to attributes or items specially in medical fields then instead of generating item id pair we generate pair of transactions with longest common subsequence of item ids. Then we gave an approach to use this framework to mine all the itemsets satisfying. We used a dynamic function which is superior to conventional functions for finding longest common subsequences. We also presented a new research trend on filtering the database in all iterations. Further investigations are required to clear the possibilities of this method.

## Acknowledgments

We thank Sh. R. S. Thakur and Sh. K. K. Shrivastava for discussing and giving us advice on its implementation.

## References

[1] R. Agrawal, T. Imielinski, and A.N. Swami, "Mining Association Rules between Sets of Items in Large Databases," Proc. ACM SIGMOD Int'l Conf. Management of Data, pp. 207-216, May 1993.  
 [2] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules," Proc. 20th Int'l Conf. Very Large Data Bases, pp. 487-499, 1994.

[3] B. Goethals, "Survey on Frequent Pattern Mining," manuscript, 2003.  
 [4] J. Han, J. Pei, and Y. Yin, "Mining Frequent Patterns without Candidate Generation," Proc. ACM SIGMOD Int'l Conf. Management of Data, pp. 1-12, May 2000.  
 [5] P. Shenoy, J.R. Haritsa, S. Sudarshan, G. Bhalotia, M. Bawa, and D. Shah. Turbo-charging vertical mining of large databases. In *ACM SIGMOD Int'l Conf. Management of Data*, May 2000.  
 [6] M. J. Zaki. Scalable algorithms for association mining. *IEEE Transactions on Knowledge and Data Engineering*, 12(3):372-390, May-June 2000.  
 [7] B. Dunkel and N. Soparkar. Data organization and access for efficient data mining. In *15th IEEE Intl. Conf. on Data Engineering*, March 1999.  
 [8] S. Sarawagi, S. Thomas, and R. Agrawal. Integrating association rule mining with databases: alternatives and implications. In *ACM SIGMOD Int'l Conf. Management of Data*, June 1998.  
 [9] M.J. Zaki, S. Parthasarathy, M. Ogihara, and W. Li, "New Algorithms for Fast Discovery of Association Rules," Proc. Third Int'l Conf. Knowledge Discovery and Data Mining, pp. 283-286, 1997.  
 [10] M.J. Zaki and K. Gouda, "Fast Vertical Mining Using Diffsets," Proc. Ninth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, pp. 326-335, 2003.  
 [11] M. Song, S. Rajasekaran. (2006). "A Transaction Mapping Algorithm for Frequent Itemsets Mining", IEEE Transactions on Knowledge and Data Engineering, Vol.18, No.4, pp. 472-481, April 2006.  
 [12] M. Jamali, F. Taghiyareh (2005) "Generating Frequent Pattern through Intersection between Transactions"  
 [13] Sunil Joshi, Dr. R. C. Jain: accepted and published research paper in The IEEE 2010 International Conference on Communication software and Networks (ICCSN 2010) on "A Dynamic Approach for Frequent Pattern Mining Using Transposition of Database" from 26 - 28 February 2010  
 [14] Finding Longest Increasing and Common Subsequences in Streaming Data David Liben-Nowell\_y dln@theory.lcs.mit.edu Erik Vee\_z env@cs.washington.edu An Zhu\_x anzhu@cs.stanford.edu November 26, 2003  
 [15] B. Jedy and F. Rioult, Database transposition for constrained closed pattern mining, in: Proceedings of Third International Workshop on Knowledge Discovery in Inductive Databases (KDID) co-located with ECML/PKDD, 2004.  
 [16] M. J. Zaki and C. J. Hsiao. CHARM: An efficient algorithm for closed itemset mining. In Proc. 2002 SIAM Int. Conf. Data Mining (SDM'02), pages 457-473, Arlington, VA, April 2002.

**Sunil Joshi** is presently working as an Ass. Professor, Computer Applications at Samrat Ashok Technological Institute Vidisha (M.P). He has 9 years teaching experience and 2 years research experience. His research areas include Data mining.

**R S Jadon** is presently working as a Head, Computer Applications at Madhav Institute of Technology and Science, Gwalior. He has 12 years research experience. He has presented research papers in more than 30 national and international conferences and published more than 30 papers in national and international journals. His research areas include Video Data Processing.

**R C Jain** is presently working as a Director and Head; Computer Applications at Samrat Ashok Technological Institute Vidisha He has 30 years teaching experience and 15 years research experience. He has presented research papers in more than 100 national and international Conferences and published more than

150 papers in national and international journals. His research areas include Data mining and Network security.

# Decision Support System for Medical Diagnosis Using Data Mining

D.Senthil Kumar<sup>1</sup>, G.Sathyadevi<sup>2</sup> and S.Sivanesh<sup>3</sup>

<sup>1</sup> Department of Computer Science and Engineering, Anna University of Technology,  
Tiruchirappalli, Tamil Nadu, India

<sup>2,3</sup> Department of Computer Science and Engineering, Anna University of Technology,  
Tiruchirappalli, Tamil Nadu, India

## Abstract

The healthcare industry collects a huge amount of data which is not properly mined and not put to the optimum use. Discovery of these hidden patterns and relationships often goes unexploited. Our research focuses on this aspect of Medical diagnosis by learning pattern through the collected data of diabetes, hepatitis and heart diseases and to develop intelligent medical decision support systems to help the physicians. In this paper, we propose the use of decision trees C4.5 algorithm, ID3 algorithm and CART algorithm to classify these diseases and compare the effectiveness, correction rate among them.

**Keywords:** Active learning, decision support system, data mining, medical engineering, ID3 algorithm, CART algorithm, C4.5 algorithm.

## 1. Introduction

The major challenge facing the healthcare industry is the provision for quality services at affordable costs. A quality service implies diagnosing patients correctly and treating them effectively. Poor clinical decisions can lead to disastrous results which is unacceptable. Even the most technologically advanced hospitals in India have no such software that predicts a disease through data mining techniques. There is a huge amount of untapped data that can be turned into useful information. Medical diagnosis is known to be subjective; it depends on the physician making the diagnosis. Secondly, and most importantly, the amount of data that should be analyzed to make a good prediction is usually huge and at times unmanageable. In this context, machine learning can be used to automatically infer diagnostic rules from descriptions of past, successfully treated patients, and help specialists make the diagnostic process more objective and more reliable.

The decision support systems that have been developed to assist physicians in the diagnostic process often are based on static data which may be out of date. A decision support system which can learn the relationships between patient history, diseases in the population, symptoms, pathology of a disease, family history and test results, would be useful to physicians and hospitals. The concept of Decision Support System (DSS) is very broad because of many diverse approaches and a wide range of domains in which decisions are made. DSS terminology refers to a class of computer-based information systems including knowledge based systems that support decision making activities. In general, it can say that a DSS is a computerized system for helping make decisions. A DSS application can be composed of the subsystems. However, the development of such system presents a daunting and yet to be explored task. Many factors have been attributed but inadequate information has been identified as a major challenge. To reduce the diagnosis time and improve the diagnosis accuracy, it has become more of a demanding issue to develop reliable and powerful medical decision support systems (MDSS) to support the yet and still increasingly complicated diagnosis decision process. The medical diagnosis by nature is a complex and fuzzy cognitive process, hence soft computing methods, such as decision tree classifiers have shown great potential to be applied in the development of MDSS of heart diseases and other diseases.

The aim is to identify the most important risk factors based on the classification rules to be extracted. This section explains how well data mining and decision support system are integrated and also describes the datasets undertaken for this work. In the next section relevant related works referred to the exploitation of classification technology in the medical field are surveyed. Section III outlines the results, explaining the decision tree



algorithms devised for the purposes outlined above. Section IV illustrates conclusions.

Decision support systems are defined as interactive computer based systems intended to help decision makers utilize data and models in order to identify problems, solve problems and make decisions. They incorporate both data and models and they are designed to assist decision makers in semi-structured and unstructured decision making processes. They provide support for decision making, they do not replace it. The mission of decision support systems is to improve effectiveness, rather than the efficiency of decisions [19]. Chen argues that the use of data mining helps institutions make critical decisions faster and with a greater degree of confidence. He believes that the use of data mining lowers the uncertainty in decision process [20]. Lavrac and Bohanec claim that the integration of dm can lead to the improved performance of DSS and can enable the tackling of new types of problems that have not been addressed before. They also argue that the integration of data mining and decision support can significantly improve current approaches and create new approaches to problem solving, by enabling the fusion of knowledge from experts and Knowledge extracted from data [19].

## 2. Overview of related work

Up to now, several studies have been reported that have focused on medical diagnosis. These studies have applied different approaches to the given problem and achieved high classification accuracies, of 77% or higher, using the dataset taken from the UCI machine learning repository [1]. Here are some examples:

Robert Detrano's [6] experimental results showed correct classification accuracy of approximately 77% with a logistic-regression-derived discriminant function.

The John Gennari's [7] CLASSIT conceptual clustering system achieved 78.9% accuracy on the Cleveland database.

L. Ariel [8] used Fuzzy Support Vector Clustering to identify heart disease. This algorithm applied a kernel induced metric to assign each piece of data and experimental results were obtained using a well known benchmark of heart disease.

Ischemic -heart:-disease (IHD) -Support .Vector Machines serve as excellent classifiers and predictors and can do so with high accuracy. In this, tree based: classifier uses non-linear proximal support vector machines.(PSVM).

Polat and Gunes [18] designed an expert system to diagnose the diabetes disease based on principal component analysis. Polat *et al.* also developed a cascade learning system to diagnose the diabetes.

Campos-Delgado *et al.* developed a fuzzy-based controller that incorporates expert knowledge to regulate the blood glucose level. Magni and Bellazzi devised a stochastic model to extract variability from a self-monitoring blood sugar level time series [17].

Diaconis, P. & Efron, B. (1983) developed an expert system to classify hepatitis of a patient. They used Computer-Intensive Methods in Statistics.

Cestnik, G., Kononenko, I., & Bratko, I. designed a Knowledge-Elicitation Tool for Sophisticated Users in the diagnosis of hepatitis.

## 3. Analysis and results

### 3.1 About the Datasets

The Aim of the present study is the development and evaluation of a Clinical Decision Support System for the treatment of patients with Heart Disease, diabetes and hepatitis. According to one survey, heart disease is the leading cause of death in the world every year. Just in the United States, almost 930,000 people die and its cost is about 393.5 billion dollars. Heart disease, which is usually called coronary artery disease (CAD), is a broad term that can refer to any condition that affects the heart. Many CAD patients have symptoms such as chest pain (angina) and fatigue, which occur when the heart isn't receiving adequate oxygen. Nearly 50 percent of patients, however, have no symptoms until a heart attack occurs.

Diabetes mellitus is a chronic disease and a major public health challenge worldwide. According to the International Diabetes Federation, there are currently 246 million diabetic people worldwide, and this number is expected to rise to 380 million by 2025. Furthermore, 3.8 million deaths are attributable to diabetes complications each year. It has been shown that 80% of type 2 diabetes complications can be prevented or delayed by early identification of people at risk. The American Diabetes Association [2] categorizes diabetes into type-1 diabetes [17], which is normally diagnosed in children and young adults, and type-2 diabetes, i.e., the most common form of diabetes that originates from a progressive insulin secretory defect so that the body does not produce adequate insulin or the insulin does not affect the cells. Either the fasting plasma glucose (FPG) or the 75-g oral glucose tolerance test (OGTT [19]) is generally appropriate to screen diabetes or pre-diabetes.

Hepatitis, a liver disorder requires continuous medical care and patient self-management education to prevent acute complications and to decrease the risk of long-term complications. This is caused due to the condition of anorexia (loss of appetite) and increased level of alkaline phosphate. The disease can be classified in to Hepatitis a,

b, etc.,. All these datasets used in this study are taken from UCI KDD Archive [1].

### 3.2 Experimental Data

We have used three medical datasets namely, heart disease, diabetes and hepatitis datasets. All these datasets are obtained from UC-Irvine archive of machine learning datasets [1]. The aim is to classify the diseases and to compare the attribute selection measure algorithms such as ID3, C4.5 and CART. The heart disease dataset [1] of 473 patients is used in this experiment and has 76 attributes, 14 of which are linear valued and are relevant as shown in table 1. The hepatitis disease dataset [1] has 20 attributes, and there are 281 instances and 2 classes which are described in table 2. The diabetic dataset [1] of 768 patients with 9 attributes is as shown in table 3.

Table 1: Description of the features in the heart disease dataset

No	Name	Description
1	Age	age in years
2	Sex	1 = male ; 0 = female
3	Cp	chest pain type (1 = typical angina; 2 = atypical angina ; 3 = non-anginal pain; 4 = asymptomatic)
4	Trestbps	resting blood pressure(in mm Hg on admission to the hospital)
5	Chol	serum cholestoral in mg/dl
6	Fbs	(fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
7	Restecg	resting electrocardiographic results (0 = normal; 1 = having ST-T wave abnormality; 2 = showing probable or definite left ventricular hypertrophy by Estes' criteria)
8	Thalach	maximum heart rate achieved
9	Exang	exercise induced angina (1 = yes; 0 = no)
10	Oldpeak	ST depression induced by exercise relative to rest
11	Slope	the slope of the peak exercise ST segment ( 1 = upsloping; 2 = flat ; 3= downsloping)
12	Ca	number of major vessels (0-3) colored by flourosopy
13	Thal	( 3 = normal; 6 = fixed defect; 7 = reversible defect)
14	Num	Diagnosis classes (0 = healthy; 1 = patient who is subject to possible heart disease)

Table 2: Description of the features in the hepatitis dataset

1	Class	DIE, LIVE
2	Age	10, 20, 30, 40, 50, 60, 70,80
3	Sex	male, female

4	Steroid	no, yes
5	Antivirals	no, yes
6	Fatigue	no, yes
7	Malaise	no, yes
8	Anorexia	no, yes
9	Liver Big	no, yes
10	Liver Firm	no, yes
11	Spleen Palpable	no, yes
12	Spiders	no, yes
13	Ascites	no, yes
14	Varices	no, yes
15	Bilirubin	0.39, 0.80, 1.20, 2.00, 3.00, 4.00
16	Alk Phosphate	33, 80, 120, 160, 200, 250
17	SGOT	13, 100, 200, 300, 400, 500,
18	Albumin	2.1, 3.0, 3.8, 4.5, 5.0, 6.0
19	Protime	10, 20, 30, 40, 50, 60, 70, 80, 90
20	Histology	no, yes

Table 3: description of the features in the diabetes dataset

No	Attribute Name	Description
1	Number of times pregnant	Numerical values
2	Plasma glucose concentration	glucose concentration in a 2 hours in an oral glucose tolerance test
3	Diastolic blood pressure	In mm Hg
4	Triceps skin fold thickness	Thickness of skin in mm
5	2-Hour serum insulin	Insulin (mu U/ml)
6	Body mass index	(weight in kg/(height in m)^2)
7	Diabetes pedigree function	A function – to analyse the presence of diabetes
8	Age	Age in years
9	Class	1 is interpreted as “tested positive for diabetes and 0 as negative

### 3.3 Attributes Selection Measures

Many different metrics are used in machine learning and data mining to build and evaluate models. We have implemented the ID3, C4.5 CART algorithm and tested them on our experimental datasets. The accuracy of these

algorithms can be examined by confusion matrix produced by them. We employed four performance measures: precision, recall, F-measure and ROC space [5]. A distinguished confusion matrix (sometimes called contingency table) is obtained to calculate the four measures. Confusion matrix is a matrix representation of the classification results. It contains information about actual and predicted classifications done by a classification system. The cell which denotes the number of samples classified as true while they were true (i.e., TP), and the cell that denotes the number of samples classified as false while they were actually false (i.e., TN). The other two cells denote the number of samples misclassified. Specifically, the cell denoting the number of samples classified as false while they actually were true (i.e., FN), and the cell denoting the number of samples classified as true while they actually were false (i.e., FP). Once the confusion matrixes were constructed, the precision, recall, F-measure are easily calculated as:

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}) \quad (1)$$

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \quad (2)$$

$$\text{F\_measure} = (2 * \text{TP}) / (2 * \text{TP} + \text{FP} + \text{FN}) \quad (3)$$

Less formally, precision measures the percentage of the actual patients (i.e. true positive) among the patients that got declared disease; recall measures the percentage of the actual patients that were discovered; F-measure balances between precision and recall. A ROC (receiver operating characteristic [5]) space is defined by false positive rate (FPR) and true positive rate (TPR) as x and y axes respectively, which depicts relative tradeoffs between true positive and false positive.

$$\text{TPR} = \text{TP} / (\text{TP} + \text{FN}) \quad (4)$$

$$\text{FPR} = \text{FP} / (\text{FP} + \text{TN}) \quad (5)$$

### ID3 Algorithm

Itemized Dichotomizer 3 algorithm or better known as ID3 algorithm [13] was first introduced by J.R Quinlan in the late 1970's. It is a greedy algorithm that selects the next attributes based on the information gain associated with the attributes. The information gain is measured by entropy, ID3 algorithm [13] prefers that the generated tree is shorter and the attributes with lower entropies are put near the top of the tree. The three datasets are run against ID3 algorithm and the results generated by ID3 are as shown in tables 4, 5, 6 respectively.

Table 4: Confusion matrix of id3 algorithm- heart disease dataset

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0.686	0.281	0.66	0.686	0.673	0.68	No
0.719	0.314	0.742	0.719	0.73	0.719	Yes

Table 5: Confusion matrix of id3 algorithm- hepatitis dataset

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0.686	0.281	0.66	0.686	0.673	0.68	No

0.719	0.314	0.742	0.719	0.73	0.719	Yes
-------	-------	-------	-------	------	-------	-----

Table 6: confusion matrix of id3 algorithm- diabetes dataset

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0.582	0.154	0.67	0.582	0.623	0.767	Yes
0.846	0.418	0.791	0.846	0.817	0.767	No

### C4.5 Algorithm

At each node of the tree, C4.5 [15] chooses one attribute of the data that most effectively splits its set of samples into subsets enriched in one class or the other. Its criterion is the normalized information gain (difference in entropy) that results from choosing an attribute for splitting the data. The attribute with the highest normalized information gain is chosen to make the decision. C4.5 [16] made a number of improvements to ID3. Some of these are:

- Handling both continuous and discrete attributes – creates a threshold and then splits the list into those whose attribute value is above the threshold and those that are less than or equal to it.
- Handling training data with missing attribute values
- Handling attributes with differing costs.
- Pruning trees after creation – C4.5 [16] goes back through the tree once it's been created and attempts to remove branches that do not help by replacing them with leaf nodes.

When the three medical datasets are run against the C4.5 algorithm and the results are indicated in the tables 7, 8, 9 respectively.

Table 7: confusion matrix of c4.5 algorithm- heart disease dataset

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0.596	0.364	0.586	0.596	0.591	0.636	No
0.636	0.404	0.646	0.636	0.641	0.636	Yes

Table 8: Confusion matrix of c4.5 algorithm-hepatitis dataset

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0.97	0.615	0.89	0.97	0.929	0.669	Live
0.385	0.03	0.714	0.385	0.5	0.669	Die

Table 9: Confusion matrix of c4.5 algorithm-diabetes dataset

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0.597	0.186	0.632	0.597	0.614	0.751	Yes
0.814	0.403	0.79	0.814	0.802	0.751	No

### CART Algorithm

Classification and regression trees (CART [14]) is a non-parametric technique that produces either classification or regression trees, depending on whether the dependent variable is categorical or numeric, respectively. Trees are formed by a collection of rules based on values of certain variables in the modelling data set. Rules are selected based on how well splits based on variables' values can differentiate observations based on the dependent variable. Once a rule is selected and splits a node into two, the same logic is applied to each "child" node (i.e. it is a recursive procedure). Splitting stops when CART detects no further gain can be made, or some pre-set stopping rules are met. The basic idea of tree growing is to choose a split among all the possible splits at each node so that the resulting child nodes are the "purest". In this algorithm, only univariate splits are considered. That is, each split depends on the value of only one predictor variable. All possible splits consist of possible splits of each predictor. CART innovations include:

- solving the "how big to grow the tree"- problem;
- using strictly two-way (binary) splitting;
- incorporating automatic testing and tree validation, and;
- Providing a completely new method for handling missing values.

The result of CART algorithm for the medical datasets are described in the following tables 10, 11, 12 respectively

Table 10: Confusion matrix of CART algorithm-heart disease dataset

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0.702	0.258	0.702	0.702	0.702	0.726	No
0.742	0.298	0.742	0.742	0.742	0.726	Yes

Table 11: Confusion matrix of CART algorithm- hepatitis dataset

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0.91	0.769	0.859	0.91	0.884	0.541	Live
0.231	0.09	0.933	0.831	0.273	0.541	Die

Table 12: Confusion matrix of CART algorithm- diabetes dataset

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0.534	0.132	0.884	0.934	0.6	0.727	Yes
0.868	0.466	0.776	0.868	0.82	0.727	No

### 3.4 Classification Rules

Significant rules [20] are extracted which are useful for understanding the data pattern and behaviour of experimental dataset. The following pattern is extracted by applying CART decision tree algorithm [14]. Some of the rules extracted for heart disease dataset are as follows,

- Heartdisease(absence):-  
 Thal=fixed\_defect,Number\_Vessels=0, Cholestorl =126-213.
- Heart\_disease(presence):-  
 Thal=normal,Number\_Vessels=0, Old\_Peak=0-1.5, Max\_Heart\_Rate=137-169, Cholestorl=126-213.
- Heart\_disease(absence):-  
 Thal=normal,Number\_Vessels=0, Old\_Peak=0-1.5, Max\_Heart\_Rate=137-169,Cholestorl=214-301, Rest=0, Pressure=121-147.

The rules for Hepatitis datasets are extracted and some of them are as follows

- Ascites = Yes AND Histology = No: Live (46.0/1.0)
- Anorexia = Yes ANDProtime > 47 AND Fatigue = No: Live (8.0)
- Anorexia = Yes AND Malaise = Yes AND Ascites = Yes: Live (10.0/2.0)
- Anorexia = Yes: Die (10.0) : Live (6.0)

Some classification rules for diabetes datasets are as follows,

- Age <= 28 AND Triceps skin fold thickness > 0 AND Triceps skin fold thickness <= 34 AND Age > 22 AND No.timespreg <= 3 AND Plasma gc(2) <= 127: No (61.0/7.0)
- Plasma gc(2) <= 99 AND 2-Hour serum insulin <= 88 AND 2-Hour serum insulin <= 18 AND Triceps skin fold thickness <= 21: No (26.0/1.0)
- Age <= 24 AND Triceps skin fold thickness > 0 AND Body MI <= 33.3: No (37.0) Diastolic blood pressure <= 40 AND Plasma gc(2) > 130: Yes (10.0)
- Plasma gc(2) <= 107 AND Diabetespf <= 0.229 AND Diastolic blood pressure <= 80: No (23.0)
- No.timespreg <= 6 AND Plasma gc(2) <= 112 AND Diastolic blood pressure <= 88 AND Age <= 35: No (44.0/8.0)
- Age <= 30 AND Diastolic blood pressure > 72 AND Body MI <= 42.8: No (41.0/7.0)

### 3.5 Comparison Of ID3, C4.5 and CART Algorithm

Algorithm designers have had much success with greedy, divide-and-conquer approaches to building class descriptions. It is chosen decision tree learners made popular by ID3, C4.5 (Quinlan1986) and CART (Breiman, Friedman, Olshen, and Stone 1984 [14] ) for this survey, because they are relatively fast and typically they produce competitive classifiers. On examining the confusion matrices of these three algorithms, we observed that among the attribute selection measures C4.5 performs better than the ID3 algorithm, but CART performs better both in respect of accuracy and time complexity. When



compared with C4.5, the run time complexity of CART is satisfactory.

Table 13: Prediction accuracy table

S.No	Name of algorithm	Accuracy %
1	CART Algorithm	83.2
2	ID3 Algorithm	64.8
3	C4.5 Algorithm	71.4

We have done this research and we have found 83.184% accuracy with the CART algorithm which is greater than previous research of ID3 and C4.5 as indicated in the table XVIII.

#### 4. Conclusions

The decision-tree algorithm is one of the most effective classification methods. The data will judge the efficiency and correction rate of the algorithm. We used 10-fold cross validation to compute confusion matrix of each model and then evaluate the performance by using precision, recall, F measure and ROC space. As expected, bagging algorithms, especially CART, showed the best performance among the tested methods. The results showed here make clinical application more accessible, which will provide great advance in healing CAD, hepatitis and diabetes. The survey is made on the decision tree algorithms ID3, C4.5 and CART towards their steps of processing data and Complexity of running data. Finally it can be concluded that between the three algorithms, the CART algorithm performs better in performance of rules generated and accuracy. This showed that the CART algorithm is better in induction and rules generalization compared to ID3 algorithm and C4.5 algorithm. Finally, the results are stored in the decision support repository. Since, the knowledge base is currently focused on a narrow set of diseases. The approach has been validated through the case study, it is possible to expand the scope of modeled medical knowledge. Furthermore, in order to improve decision support, interactions should be considered between the different medications that the patient is on.

#### References

[1] UCI Machine Learning Repository  
<http://www.ics.uci.edu/~mllearn/MLRepository.html> .  
 [2] American Diabetes Association, "Standards of medical care in diabetes—2007," *Diabetes Care*, vol. 30, no. 1, pp. S4-S41, 2007.  
 [3] J. Du and C.X. Ling, "Active Learning with Generalized Queries," Proc. Ninth IEEE Int'l Conf. Data Mining, pp. 120-128, 2009  
 [4] Jiawei Han and Micheline Kamber, "Data Mining Concepts and techniques", 2nd ed., Morgan Kaufmann Publishers,

San Francisco, CA, 2007.  
 [5] H.W. Ian, E.F., "Data mining: Practical machine learning tools and techniques," 2005: Morgan Kaufmann.  
 [6] R. Detrano, A.J., W. Steinbrunn, M. Pfisterer, J.J. Schmid, S. Sandhu, K.H.Guppy, S. Lee, and V. Froelicher, "International application of a new probability algorithm for the diagnosis of coronary artery disease," *American Journal of Cardiology*, 1989. 64: p. 304-310.  
 [7] G. John, "Models if incremental concept formation," *Journal of Artificial Intelligence*, 1989: p. 11-61.  
 [8] A. L. Gamboa, M.G.M., J. M. Vargas, N. H. Gress, and R. E. Orozco, "Hybrid Fuzzy-SV Clustering for Heart Disease Identification," in *Proceedings of CIMCA-IAWTIC'06*. 2006.  
 [9] D. Resul, T.I., S. Abdulkadir, "Effective diagnosis of heart disease through neural networks ensembles," Elsevier, 2008.  
 [10] Z. Yao, P.L., L. Lei, and J. Yin, "R-C4.5 Decision tree modeland its applications to health care dataset, in roceedings of the 2005 International Conference on Services Systems and Services Management," 2005. p. 1099-1103.  
 [11] K. Gang, P.Y., S. Yong, C. Zhengxin, "Privacy-preserving data mining of medical data using data separation-based techniques," *Data science journal*, 2007. 6.  
 [12] L. Cao, "Introduction to Domain Driven Data Mining," *Data Mining for Business Applications*, pp. 3-10, Springer, 2009.  
 [13] Quinlan, J.R., "Induction of Decision Trees," *Machine Learning*. Vol. 1. 1986. 81-106.  
 [14] L. Breiman, J. Friedman, R. Olshen, and C. Stone. *Classification and Regression Trees*. Wadsworth Int. Group, 1984.  
 [15] S. R. Safavin and D. Landgrebe. A survey of decision tree classifier methodology. *IEEE Trans. on Systems, Man and Cybernetics*, 21(3):660-674, 1991.  
 [16] Kusriani, Sri Hartati, "Implementation of C4.5 algorithm to evaluate the cancellation possibility of new student applicants at stmik amikom yogyakarta." *Proceedings of the International Conference on Electrical Engineering and Informatics Institut Technologic Bandung, Indonesia June 17-19, 2007*.  
 [17] P. Magni and R. Bellazzi, "A stochastic model to assess the variability of blood glucose time series in diabetic patients self-monitoring," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 6, pp. 977-985, Jun. 2006.  
 [18] K. Polat and S. Gunes, "An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease," *Dig. Signal Process.*, vol. 17, no. 4, pp. 702-710, Jul. 2007.  
 [19] J.Friedman, "Fitting functions to noisy data in high dimensions", in *Proc.20<sup>th</sup> Symp. Interface Amer. Statistical .Assoc. , E.J.Wegman.D.T.Gantz, and I.J. Miller.Eds.1988 pp.13-43*  
 [20] T.W.simpson, C.Clark and J.Grelbsh , "Analysis of support vector regression for appreciation of complex engineering analyses ", presented as the ASME 2003.  
 [21] L. B. Goncalves, M. M. B. R. Vellasco, M. A. C. Pacheco, and F. J. de Souza, "Inverted hierarchical neuro-fuzzy BSP system: A novel neuro-fuzzy model for pattern classification and rule extraction in LEE AND WANG: FUZZY EXPERT SYSTEM FOR DIABETES DECISION SUPPORT

APPLICATION 153 databases,” IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., vol. 36, no. 2, pp. 236–248, Mar. 2006.

**First Author** D. Senthil Kumar is an Assistant Professor in the Department of Computer Science and Engineering in Anna University of Technology, Tiruchirappalli, India. He has completed 10 years of Teaching in various courses in the Undergraduate and Postgraduate Engineering & MBA program. He received a Master of Science in Mathematics from Presidency College, University of Madras and Master of Engineering in Systems Engineering And Operations Research from College of Engineering, Anna University (both located in Chennai, India). He received Prof. T.R. Natesan Endowment Award (Instituted by Operational Research Society Of India – Chennai Chapter). He is a member of IEEE and his research interest includes Optimization, Security and Data Mining.

**Sathyadevi** received the B.E degree in computer science and Engineering from Coimbatore Institute of Engineering and Information Technology in 2009. She is currently a M.E. candidate in the Department of Computer Science at Anna University of Technology, Tiruchirappalli. Her research interests include data mining, machine learning, and related real-world applications.

**Third Author** S.Sivanesh is an Assistant Professor in Computer Science and Engineering in Anna University of Technology, Tiruchirappalli, India. His research interests include Internet routing, routing security, network management and measurement.



# Internet and political communication – Macedonian case

MSc. Sali Emruli<sup>1</sup>, Prof. dr. sc. Miroslav Bača<sup>2</sup>

Faculty of Organization and Informatics, University of Zagreb,  
Varaždin, 42000, Croatia

Faculty of Organization and Informatics, Department for biometrics, University of Zagreb,  
Varaždin, 42000, Croatia

## Abstract

Analysis how to use Internet influence to the process of political communication, marketing and the management of public relations, what kind of online communication methods are used by political parties, and to assess satisfaction, means of communication and the services they provide to their party's voters (people) and other interest groups and whether social networks can affect the political and economic changes in the state, and the political power of one party.

**Keywords:** *Network Analysis, Political parties, Complexity, Scale Free Network, Social Network Analysis, Non-Profit Organization, Capacity, Public relations, marketing, Interne, Facebook, YouTube, Twitter, Blogs, MySpace, and Forum.*

## 2. Introduction

The analysis will be done in a way that will create a list of largest political parties in the Republic of Macedonia, their communications infrastructure through ICT (website), content and the manner in which they placed their information and receive feedback from voters.

Internet, social networking, Web 2.0, Facebook, YouTube, blog ... All these are relatively new word in the political vocabulary, new concepts, new media and new opportunities for the transmission of ideas and messages are not enough channels used to communicate with the public. Although the practice of using the Internet in local political advertising goes back to the nineties, only in recent years the advent of new tools and social networks demonstrates true strength of this medium.

Besides direct access to the public, political ideas, it provides full force confrontation, but also provides a relatively convenient ground for review of public attitudes, research and development of certain ideas. Using such a change in social communication, transmission of political

messages through the transition from traditional forms of communication and finding new paths to the recipients. Professional and political public for years following the development of the Internet as a medium, but he showed the greatest strength in the last U.S. presidential election.

Political power depends on the satisfaction of the people towards a particular party and party connections with other parties or organizations. Well-developed social network provides further prestige and power of the party and its direct channel of communication with voters and other influential interests groups.

## 3. Political Communication through Internet

Internet and politics in the modern world have become inseparable and thus gradually eliminating barriers to free flow of information between the political decision-makers and those in whose name the benefits they bring (the public). Countries in transition must follow the contemporary trends of fitting of the Internet in the area of political communication, which simultaneously causes the change to the model that is still the dominant, of political communication based on secrecy and lack of transparency.

In Macedonia, the network and politics are still not together, except in the case of international organizations. Internet is not fully incorporated into political communication (or, more precisely, it is not done properly). A key condition (requirement) for this is application of technology and simultaneous transformation of consciousness. This change requires the rejection of the principle of confidentiality as a condition of political activity of government and party, because it is absolutely contrary to the nature of the Internet. It is necessary also to strengthen the awareness of the importance of on-line crystallization of public opinion, and more intensive and better connection of on-line and off-line political stage.

## 4. Impact of Social network analysis in politics

Political communication is a new and exciting area of research and teaching that is located at the crossroads of the study of communication, political parties and electoral behavior. As well as profiling the changing nature of the media system such an approach invariably leads us onto what we term the 'new political communication' - that based around the new Information and Communication Technologies (ICTs). We examine the work that has been done on the uses of the new media by parties and politicians across a range of democratic contexts and offer some insights into the strong challenges they introduce for the established manufacturers of political communication. One of the key uses of the Internet is to build databases of voter data and access that through different applications for different purposes. Because data entry can be easily done automatically by scanners or by hand more campaigns and political operatives are recognizing the importance of capturing, storing, analyzing and using voter information. What used to take days of analyzing can now take minutes by using computers to analyze important information. That data can also be used offline or online for a number of different ways and the usage of these systems have become key components of the political system.

Throughout history political campaigns have evolved around the advancing technologies that are available to candidates. As technology develops, candidates are able to permeate the lives of citizens on a daily basis. Television, radio, newspapers, magazines, billboards, yard signs, bumper stickers, and Internet websites all create a means of spreading political platforms.

While the traditional forms of media are still an integral portion of campaign strategy, the availability of the Internet opens the door of campaign tools waiting for candidate's attention. The Internet provides numerous opportunities for politicians to reach the polity. Among those is a new phenomenon called social networking websites. Social networking sites have gained popularity in the last few years. These sites are growing popular particularly on college campuses nationwide. Specifically social networking websites such as MySpace and Facebook have provided users with a new form of communication. When new forms of communication are made available, political candidates begin to use the new technology to their advantage. What social networking websites allow politicians to do is to create a sense of personalized communication with their constituents. This personalization of politics enables voters and politicians alike to feel as though a connection is made. The Internet can make direct communication possible among

government officials, candidates, parties, and citizens. As history shows us, when new technologies are made available, they begin to reshape the personalization factor between the candidate and the voter. This increase in interpersonal interactivity has shown to offer opportunities and increase success for political campaigns.

## 5. Political Parties in Republic of Macedonia

### 5.1. Overview of the political system

Macedonia is a Republic having multi-party parliamentary democracy and a political system with strict division into legislative, executive and judicial branches. From 1945 Macedonia had been a sovereign Republic within Federal Yugoslavia and on September 8, 1991, following the referendum of its citizens, Macedonia was proclaimed a sovereign and independent state. The Constitution of the Republic of Macedonia was adopted on November 17, 1991, by the first multiparty parliament. The basic intention was to constitute Macedonia as a sovereign and independent, civil and democratic state and also to create an institutional framework for the development of parliamentary democracy, guaranteeing human rights, civil liberties and national equality.

The Assembly is the central and most important institution of state authority. According to the Constitution it is a representative body of the citizens and the legislative power of the Republic is vested in it. The Assembly is composed of 120 seats.

The President of the Republic of Macedonia represents the Republic, and is Commander-in-Chief of the Armed Forces of Macedonia. He is elected in general and direct elections, for a term of five years, and two terms at most.

Executive power of the Republic of Macedonia is bicephalous and is divided between the Government and the President of the Republic. The Government is elected by the Assembly of the Republic of Macedonia by a majority vote of the total number of Representatives, and is accountable for its work to the Assembly. The organization and work of the Government is defined by a law on the Government.

In accordance with its constitutional competencies, executive power is vested in the Government of the Republic of Macedonia. It is the highest institution of the state administration and has, among others, the following responsibilities: it proposes laws, the budget of the Republic and other regulations passed by the Assembly, it determines the policies of execution of laws and other regulations of the Assembly and is responsible for their execution, decides on the recognition of states and governments, establishes diplomatic and consular relations

with other states, proposes the Public Prosecutor, proposes the appointment of ambassadors and representatives of the Republic of Macedonia abroad and appoints chiefs of consular offices, and also performs other duties stipulated by the Constitution and law.

In Macedonia there are more political parties participating in the electoral process at national and local level.

## 5.2. Current Structure

Parties of traditional left and right:

Coalition VMRO – DPMNE (63 mandates, right oriented Macedonian party)

Democratic Party of the Albanians (12 mandates, right oriented Albanian party)

Coalition “SONCE” – SDSM (27 mandates left oriented Macedonian party)

Democratic Union for Integration (18 mandates left oriented Albanian party)

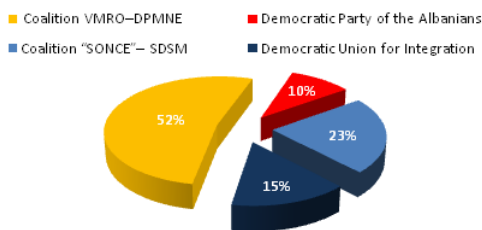


Fig. 1 Current political structure in Republic of Macedonia and Political Parties mandate percentage win in latest parliamentary elections.

## 6. Political party web sites and use of social media

Party websites represent an application of a technology which has led those dealing in votes to invest considerable amounts of time and money. This part presents a survey of the websites of Macedonian political parties. It examines the individual parties on the web and the party system on the web as virtual counterparts of the ordinary parties and party system.

Because of the importance of applying information and communication technologies (ICT) in the work of central and local government and in terms of facilitating the life and work of citizens, as well as the major role in the development of information society in EU integration process, an analysis has been conducted on websites and social media usage by political parties in Macedonia.

The research includes analysis of websites, analysis of the use of social media and online activities compared in terms of seats obtained by political parties.

In account were taken only websites of the major political parties that can certainly be determined to be guided by their info centers.

In order to level the differences between the parties and a common base for comparison, two indexes were created: Internal platform which are the party's official website and External platforms which are websites that defined as social media platforms.

While reviewing the websites I have searched for features that can be considered as social media features and that were incorporated in parties' websites.

Contrary to traditional content analysis where texts are the subject of a thorough analysis, here a content analysis was made, but on a more general level of website sections and less on their content.

Platforms that are not owned by the parties and considered as social media platforms. On these platforms the party has an official profile/user that is uploading the content and has the permissions to monitor moderate the other users' activity.

The results had showed that the relatively popular parties, The VMRO-DMPNE, Social Democrats, DPA, DUI and ND had been using more social media features and platforms than the other parliament parties.

Despite these findings there are no clear signs of an established use of social media as a political communication strategy. There is not enough correlation between number of parliament members and website's users. Some of major parties are the worst in translating voters into website's users.

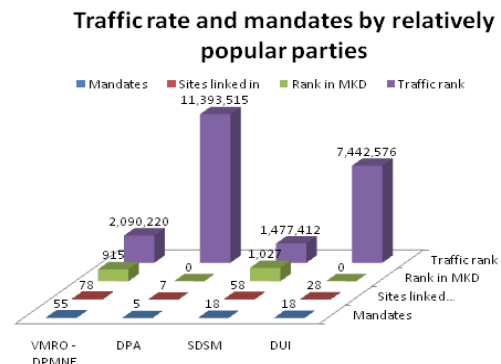


Fig. 2 Represents Traffic rate and mandates by relatively popular parties

## 6.1. Purpose of research

Due to the increased use of Internet in Macedonia and the increased influence of the same as medium, research has been conducted to determine how political parties use the influence of Internet.

The study aims to evaluate several levels of Internet usage and social networks in everyday political events:

- To determine the level of quality and implementation of the standards of the political parties websites
- To determine the level of use of social media for their promotion
- To determine the level of online communication of political parties with public
- To compare the number of online supporters with conquered mandate of the last elections.

One area of running a political campaign is Internet, and also the use of social media. The analysis will include internal platform which are the party's official website their social content and does this web pages have links to the external platforms which are websites that defined as social media platforms. Also, the technical characteristics of the sites.

### 6.1.1. Research questions and hypotheses

The analysis of political parties' websites is based on clearly defined issues that are divided into several categories, in order to evaluate every aspect of the content and making the website of the political party. Questionnaire for this section is designed to address the following main questions:

- In what languages website is available
- What type of content is offered on website (text, multimedia, transparency ...)
- Applying the standard for usability of website
- Usage of Social Media

### 6.1.2. Limits

In conducting the research and when creating the list of political parties whose Web sites will be analyzed, there were taken in consideration only those whose identity could be confirmed. It means that it is evident that website is managed by the political party.

## 6.1.3. Methodology

To implement this part of the research content analysis methodology was used. Drawing up the list of web sites that will be use, their analysis is processed by strictly defined form, with concrete questions and directions.

The form consists of three main issues, which contain additional questions about obtaining the necessary information and conclusions.

The main issues are:

- In which language versions websites are accessible
- What type of content is offered by websites
- What kind of social media are used by political party

## 6.2. Language version

The websites analysis will include Web sites of relatively popular political parties in Macedonia. From 19 political parties, only 13 (70%) have their own web sites for promotion and marketing of their political activity. From analyzed Web sites , only 4 offer bilingual accessibility (30%), and others offer information only in the language of own ethnicity (unilingual content).

Language versions of political party websites	Political party	Percentage
Macedonian	10	77%
Albanian	3	23%
English	4	30%
Turkish	1	8%
Serbian	1	8%

Table 1: Language Versions

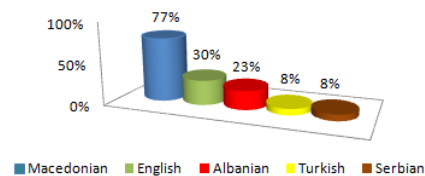


Fig. 3 Language versions

As I mentioned above from analyzed websites, only 4 offer bilingual accessibility (30%), and nine other political parties (70%) offer only unilingual content.

Multilanguage Usage by political party websites	Political party	Percentage
Multilanguage	4	31%
Only Macedonian	6	46%
Only Albanian	2	15%
Only Turkish	1	8%

Table 2: Multilanguage Usage

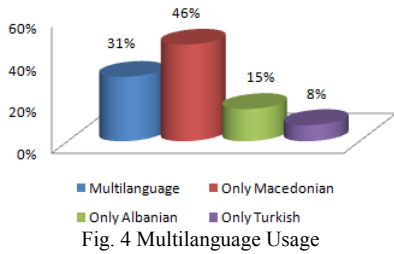


Fig. 4 Multilanguage Usage

### 6.3. Offered content

The second area for which websites analysis was performed is content offered by political party websites to their readers. The type of content is divided into text and multimedia.

	Mandates	Textual content			
		About Us	Events/Calendar	Notifications	Out Links
1 VMRO - DPMNE	55	1	1	1	1
2 Social - Democratic Union of Macedonia	18	1	1	1	1
3 Democratic Union for Integration	18	0	1	1	1
4 Democratic Party of the Albanians	5	1	1	1	0
5 Liberal - Democratic Party	4	1	1	1	0
6 New Democracy	4	1	1	1	1
7 New Social-Democratic Party	3	1	1	1	1
8 Socialist Party	1	1	1	1	1
9 Democratic reconstruction of Macedonia	1	1	1	1	1
10 Democratic union	1	1	1	1	1
11 Democratic Party of Serbs in Macedonia	1	1	1	1	1
12 Democratic Party of Turks in Macedonia	1	1	1	1	0
13 Liberal Party	1	1	1	1	0

Table 3-1: Content offered by political parties websites

	Mandates	Multimedia			
		Photos	Photo gallery	Audio clips	Video Clips
1 VMRO - DPMNE	55	1	0	0	0
2 Social - Democratic Union of Macedonia	18	1	1	1	1
3 Democratic Union for Integration	18	1	1	0	1
4 Democratic Party of the Albanians	5	1	0	0	1
5 Liberal - Democratic Party	4	1	1	0	1
6 New Democracy	4	1	1	1	1
7 New Social-Democratic Party	3	1	1	0	1
8 Socialist Party	1	1	0	0	0
9 Democratic reconstruction of Macedonia	1	1	1	0	1
10 Democratic union	1	1	0	0	0
11 Democratic Party of Serbs in Macedonia	1	1	1	1	1
12 Democratic Party of Turks in Macedonia	1	1	0	0	0
13 Liberal Party	1	1	0	0	0

Table 3-2: Content offered by political parties' websites

The results of this part of the research shows that most of the websites of political parties are filled with textual content, but that the textual content is not linked to the outside source (Out Links) 31% of websites, while regarding the multimedia content nearly 46% of the websites of political parties have no photo gallery, 77% of websites of political parties have no audio clips and 38% of websites have no video clips. Also a lack in all the websites of political parties is informing the guests for

future activities and events. Most political parties use their websites to archive articles from the media, rather than used to inform their supporters.

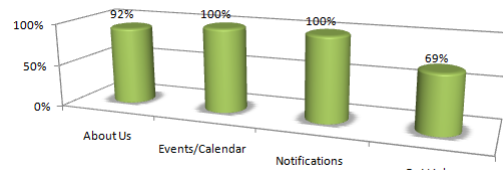


Fig.5 Textual content of all political party web sites

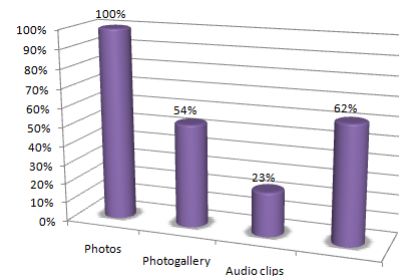


Fig. 6 Multimedia content of all political party web sites

Frequently political parties put banners to external WebPages, and not applying the concept of linking in the text. Although most of the texts offered by political parties at their web sites are excerpts from articles in the media, although they cite the source from where the content is downloaded, they not publish the link to the original article, not even the online edition of the medium. Besides textual content almost all political parties are offering and multimedia content. Most of political parties have placed videos of nearly any report or television interview. Besides the video clips, several political parties offer galleries of their activities. Only audio clips are missing from multimedia content on political party websites. Only three political parties have offered this type of content, and they have offered several songs (hymns of the party) for download, but they were not taken into account.

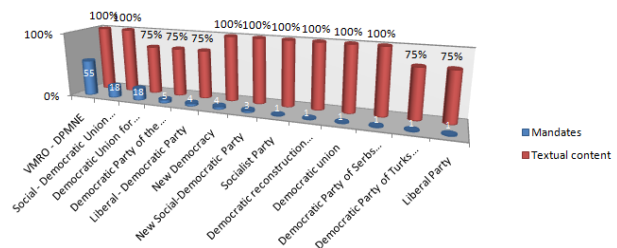


Fig.7 Textual content of each political party websites



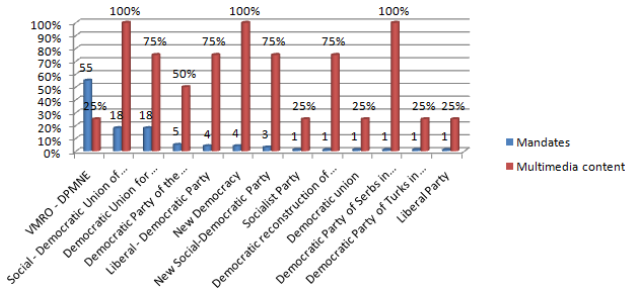


Fig.8 Multimedia content of each political party website.

Besides the types of content (text and multimedia) in this part of the analysis were explored also the ways of communication which are offered by political parties through their websites.

Surprising fact that political parties do not use their sites for the opportunity to communicate and besides the possibility of sending e-mails almost there is no other possible way to contact them, except a very small number of political parties which offers email and other contact information as phone number, address and the like.

Very few political parties have set forum on their website, also very small number of political parties offers an opportunity for asking questions and publishing answers online.

	Mandates	Public e-mail	Form	Phone	ZIP address	Mailing list	Poll	Discussion	Forum
VMRO - DPMNE	55	0	1	1	1	0	0	1	1
Social - Democratic Union of Macedonia	18	1	1	1	1	1	1	1	1
Democratic Union for Integration	18	1	1	1	1	1	0	0	0
Democratic Party of the Albanians	5	1	1	1	0	1	0	0	0
Liberal - Democratic Party	4	1	1	1	1	0	0	0	0
New Democracy	4	1	1	0	0	0	0	1	1
New Social-Democratic Party	3	1	1	1	1	1	0	0	0
Socialist Party	1	1	0	1	1	0	0	0	0
Democratic reconstruction of Macedonia	1	1	1	1	1	1	0	0	0
Democratic union	1	1	1	0	0	0	0	0	0
Democratic Party of Serbs in Macedonia	1	1	1	1	1	0	1	0	0
Democratic Party of Turks in Macedonia	1	1	1	1	1	0	0	0	0
Liberal Party	1	1	1	1	1	1	0	0	0

Table 4: Ways of communication offered by political parties

Table shows that in terms of interaction, political parties are not handled and did not use the opportunities of new media field. Besides basic information such as postal address, phone and email address, no other method is used. Sites of some political parties have disabled the opportunity to contact them via e-mail or form, but they offer only the traditional ways of communication (telephone and letter).

For transparency of the website is necessary to enable seeing the number of visitors on the site, which was also left out of more websites of the political parties. A very small part of the political parties had included counters on their websites, whether public or just used by the administrators of the website. This means that political

parties do not take care of attendance (visitors) of their websites.

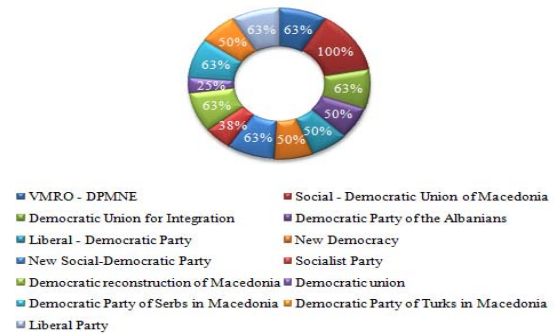


Fig.9 Opportunities for contact / interaction / transparency offered by each political party in Macedonia

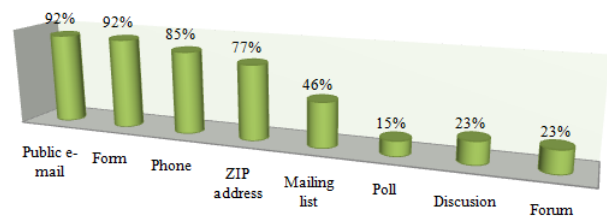


Fig.10 Opportunities for contact / interaction / transparency offered by all political party in Macedonia by category

	Mandates	Visitors Counter
VMRO - DPMNE	55	0
Social - Democratic Union of Macedonia	18	0
Democratic Union for Integration	18	0
Democratic Party of the Albanians	5	0
Liberal - Democratic Party	4	0
New Democracy	4	0
New Social-Democratic Party	3	0
Socialist Party	1	0
Democratic reconstruction of Macedonia	1	0
Democratic union	1	0
Democratic Party of Serbs in Macedonia	1	1
Democratic Party of Turks in Macedonia	1	0
Liberal Party	1	0
<b>Total</b>	<b>1</b>	<b>1</b>
<b>Percentage</b>	<b>8%</b>	<b>8%</b>

Table 5: Visitors Counter on political web pages

From the table we can conclude that almost 92% of political parties have no counters on their websites, as a consequence of lack of counters we cannot say with certainty about the attendance (visitors) of website of certain political parties.

Almost all political parties have used CMS (Content Management System) for making their websites, so they meet the basic rules for usability of the website. However astonishing fact that despite meeting the technical specifications for usability, they have errors that are not inherent for the platforms that are used, for example the search box which does not work properly and the like. In terms of recommendations for visibility of search engines,



many political parties does not satisfy the conditions, which means that their site search will not be among the first results and will not be easily accessible to readers.

### 6.4. Technical Specifications of political parties' websites

This part of research covers the technical characteristics of the websites of political parties, respectively hosting and platform on which websites are set up and registration of the domain are shown in the table below.

	Mandates	Software solution	Platform	Server	Registered in	Hosted in
VMRO - DPMNE	55	ASP	Windows 2003	Microsoft-IIS/6.0	MK	MK
Social - Democratic Union of Macedonia	18	ASPX	Windows 2000	Microsoft-IIS/5.0	MK	MK
Democratic Union for Integration	18	PHP	Windows 2003	Microsoft-IIS/6.0	MK	MK
Democratic Party of the Albanians	5	PHP	Unknown	Apache/2.2.X OVH	France	France
Liberal - Democratic Party	4	ASP	Windows 2000	Microsoft-IIS/5.0	MK	MK
New Democracy	4	ASP	Windows 2003	Microsoft-IIS/6.0	US	US
New Social-Democratic Party	3	PHP	Unknown	Unknown	MK	MK
Socialist Party	1	PHP	Unknown	Unknown	MK	NL
Democratic reconstruction of Macedonia	1	PHP	Unknown	Unknown	MK	US
Democratic union	1	PHP	Windows 2000	Microsoft-IIS/5.0	MK	MK
Democratic Party of Serbs in Macedonia	1	PHP	Unknown	Unknown	US	US
Democratic Party of Turks in Macedonia	1	PHP	Unknown	Unknown	US	US
Liberal Party	1	PHP	Unknown	Unknown	US	US

Table 6: Technical Specifications of political parties web sites

If we do comparison of software solutions which is more used we can conclude that 69% of political parties websites use the PHP programming language, 23% use ASP programming language and only 8% use ASPX programming language. Platforms that are used by political parties websites are 23% Windows Server 2003, 23% of them use Windows 2000 platform and 54% were Unknown platforms.

Software solution	Platform
ASPX	8%
ASP	23%
PHP	69%

Table 7: Software solutions and platforms used by political parties websites

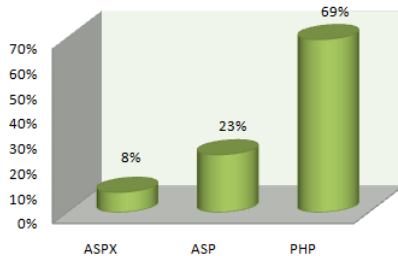


Fig.11 Software solution used by political parties' websites in Macedonia

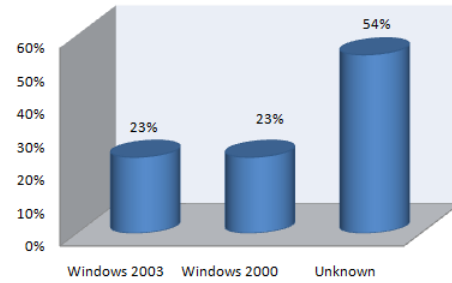


Fig.12 Platforms used by political parties' websites in Macedonia

Server	Hosted In
Microsoft-IIS/6.0	23%
Microsoft-IIS/5.0	23%
Apache/2.2.X OVH	8%
Unknown	46%

Table 8: Servers used by political parties' websites in Macedonia

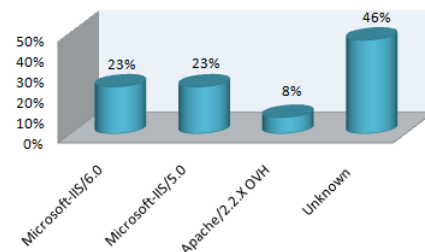


Fig.13 Servers used by political parties' websites in Macedonia

Most of political parties websites are hosted in Macedonia 46% of the total number of political Web sites, 38% of them are hosted in the US, 8% are hosted in France and 8% are hosted in the Netherland.

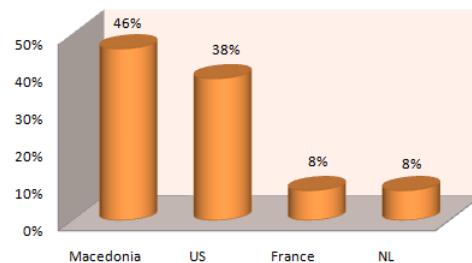


Fig.14 Hosted in Countries

Most of the websites of political parties are registered in the Republic of Macedonia 63%, while 31% are registered in the United States and 7% are registered in France.

Registered In	Percentage
Macedonia	62%
US	31%
France	7%

Table 9: Registered in countries

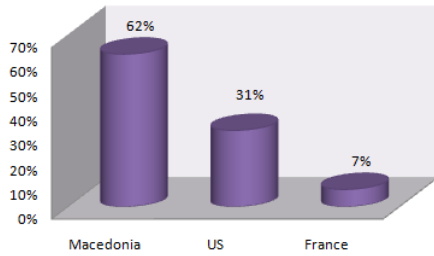


Fig.15 Registered in countries

According to the table in which the technical characteristics of the websites of political parties are presented, we can see that political parties have used free solutions. While the information for domains shows that websites of political parties often are registered and hosted in Macedonia, but there are a few exceptions, some political parties have their sites hosted in the U.S., but again Macedonia is dominant in terms of domain registration. Here it is important to mention that those domains that are not registered in Macedonia have no extension. mk.

### 6.5. Profiles on social media websites

Recently, many popular are accounts of political parties and politicians on social media, so precisely because they were included in this study. According to this fact i have collected information which politicians (party presidents) have their official account on social media (social networking sites) and the political parties use these social media to communicate with electors.

Political Party	Mandates	Facebook	Twitter	Youtube	MySpace	Hi5	Flickr	Photobucket	Wikipedia	Blog.com.mk	Other	Total	Percentage
VMRO - DPMNE	55	1	1	1	1	1	1	1	1	1	1	10	18%
Social - Democratic Union of Mandates	18	1	1	1	0	1	0	1	1	1	1	8	44%
Democratic Union for Integration	18	1	0	1	1	1	0	0	1	1	1	6	33%
Democratic Party of the Albanians	6	1	0	1	0	1	0	0	1	1	1	6	100%
Liberal - Democratic Party	4	1	0	1	1	0	0	0	1	1	1	6	150%
New Democracy	4	1	1	1	0	1	0	0	1	1	1	7	175%
New Social Democratic Party	3	1	0	0	1	0	0	0	1	1	1	5	167%
Radical Party	1	1	0	1	1	0	0	0	1	1	1	6	600%
Democratic reconstruction of Macedonia	1	1	0	1	0	1	0	0	1	1	1	6	600%
Democratic Union	1	1	0	1	0	0	0	0	1	1	1	5	500%
Democratic Party of Serbs in Macedonia	1	1	0	1	1	0	0	0	1	0	1	5	500%
Democratic Party of Turks in Macedonia	1	1	0	1	0	1	0	0	1	1	1	6	600%
Liberal Party	1	1	0	1	0	1	0	0	1	1	1	6	600%
<b>Total all party</b>	<b>113</b>	<b>13</b>	<b>9</b>	<b>12</b>	<b>6</b>	<b>7</b>	<b>1</b>	<b>2</b>	<b>13</b>	<b>12</b>	<b>13</b>	<b>113</b>	
<b>Percentage all party</b>	<b>100%</b>	<b>11%</b>	<b>8%</b>	<b>11%</b>	<b>5%</b>	<b>6%</b>	<b>1%</b>	<b>2%</b>	<b>12%</b>	<b>11%</b>	<b>12%</b>	<b>113</b>	

Political Party Leader	Mandates	Facebook	Twitter	Youtube	MySpace	Hi5	Flickr	Photobucket	Wikipedia	Blog.com.mk	Other	Total	Percentage
Nikola Gruevski	55	1	0	1	1	1	1	1	1	1	1	9	16%
Branko Crvenkovski	18	1	0	1	1	1	0	1	1	1	1	8	44%
Ali Ahmeti	18	1	0	1	1	1	0	1	1	1	1	8	44%
Mentuh Thaci	6	1	0	1	1	0	0	0	1	1	1	6	100%
Jovan Manasijevski	4	0	0	1	0	0	0	0	0	1	1	3	75%
Imer Selmani	4	1	0	1	0	1	0	0	1	1	1	6	150%
Tito Petkovski	3	1	1	1	0	1	1	0	1	1	1	8	267%
Libibisav Ivanov - Zingo	1	1	0	1	0	0	0	0	1	0	1	4	400%
Liljana Popovska	1	0	1	1	0	0	0	0	1	1	1	5	500%
Pavle Trajanov	1	0	0	1	0	0	0	0	1	1	1	4	400%
Ivan Stoilkovic	1	1	0	0	0	0	0	0	1	1	0	3	300%
Keman Hasip	1	0	0	1	0	0	0	0	0	0	1	2	200%
Borce Stojanovski	1	0	0	0	0	0	0	0	0	1	1	2	200%
<b>Total all party</b>	<b>4</b>	<b>3</b>	<b>11</b>	<b>4</b>	<b>4</b>	<b>9</b>	<b>1</b>	<b>2</b>	<b>13</b>	<b>11</b>	<b>11</b>	<b>64</b>	
<b>Percentage all party</b>	<b>52%</b>	<b>13%</b>	<b>43%</b>	<b>25%</b>	<b>46%</b>	<b>23%</b>	<b>2%</b>	<b>2%</b>	<b>20%</b>	<b>17%</b>	<b>17%</b>	<b>64</b>	

Table 10: Usage of social media by each political party and its leader

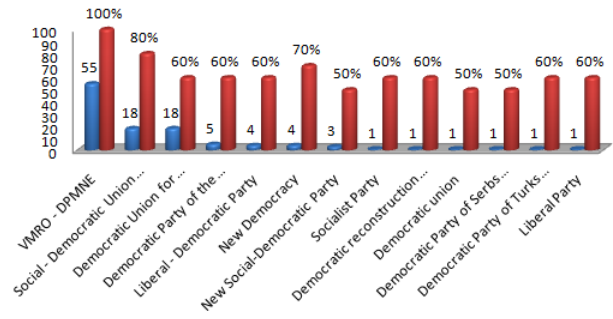


Fig.16 Usage of social media by each political party

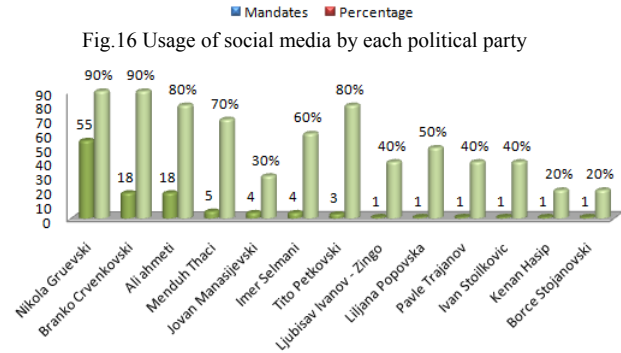


Fig.17 Usage of social media by each president of political parties

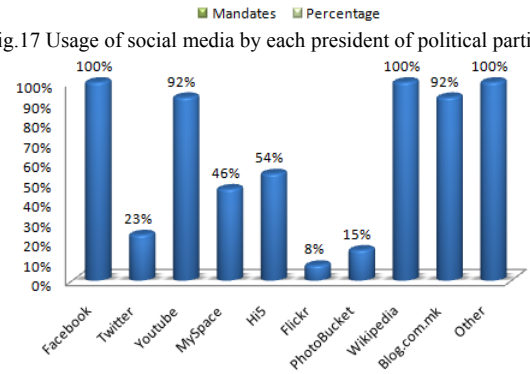


Fig.18 Usage of social media by all political parties

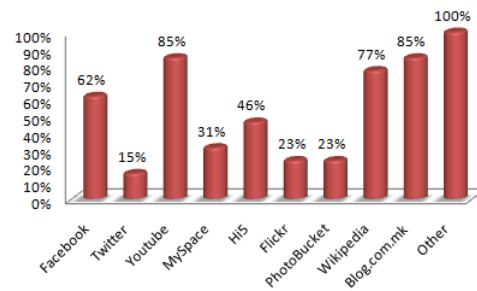


Fig.19 Usage of social media by all presidents of political parties

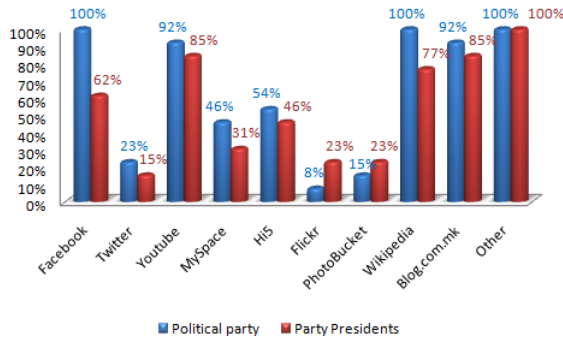


Fig.20 Usage of social media by all political parties VS all party presidents

According to the results, they show that despite the popularity of social media (SNS), political parties and politicians (presidents of the parties) have not used fully the advantages and opportunities of these tools to connect with their supporters. Most popular services are Facebook and YouTube, while platforms for sharing images were used rarely. Some political parties had implemented photo galleries on their sites.

## 7. Conclusion

The Internet first became a significant political tool in offering one-way communication for political parties with the development of political parties' websites. However, politics online is no longer as valued for its one-way communication but is now praised, and used for the opportunities it provides to conduct two way communications between political parties, their campaigns, and potential voters.

In today's political arena, websites and Internet resources, such as weblogs, social networks, podcasts and compatible video formats are being shared as a means of consuming and disseminating information via the web. As a result, websites are becoming a major if not the number one resource for political campaigns to contact supporters, volunteers, and donations. At the same time, for the consumer, or, in this case, the voter, the Internet has become a primary resource for campaign media explored via web blogging, campaign websites, news sites, social networks, video sharing and podcasts. While traditional websites are still offering significant value to the political world, technology is pushing the envelope steps further with the use of web blogging, the development of social networks, the availability of podcasts (news and opinion related), and video sharing through sites such as YouTube, which provide the general public with video clips (of up to 5+ minutes in length). Thus the issue is no longer whether politics is online but, instead, in what form and with what consequences.

Politics on the Internet has expanded beyond static two-dimensional web pages that used to serve as online billboards, flyers for a candidates position, and the traditional barriers of physical organizing. This has ushered in a new era of online consumer media and networking content that is saturated by political and campaign content. Furthermore, the phenomena of campaigns and the Internet is becoming less about what is featured on the campaign website, and instead consists more of user generated and user spread content that circulates virally on the Internet, connecting supporters from across the globe.

So far, Web 2.0 has had a weak e-ruptive effect on Macedonian party politics. On the horizontal dimension, the parties' share of activities on Web 2.0 has mainly followed what could be expected from their share of votes in 2008 parliamentary election. But deviations from the pattern indicate there are variable priorities, meaning that given a minimum of resource, parties and activists can decide to be "big in Web 2.0 politics" or decide not to. Regarding the vertical dimension of e-ruption, it appears that the national party organizations has gained more control and initiative in 2009 presidential an municipality election, the more anarchical situation of 2008 being temporary, due to sudden introduction of new technologies. Furthermore, while the number of users, viewers, members, followers and bloggers may have doubled since 2008, the party political Web 2.0 segment is still very small. This is both as a segment on Web 2.0 and as segment of voters in general.

Therefore the Web 1.5 hypothesis appears to give the best description. Furthermore, a likely next step is an even more integrated and proactive strategy, as indicated by providing guidance and cues on the party web sites, as well as setting up party specific networks or "zones" on places like Facebook. Success stories of internet politics, and especially Obama, have had a significant impact on Macedonian media. Comments like the one quoted below is quite common:

"Macedonian politicians have a lot to learn from Obama and his staff when it comes to running electoral campaigns. In particular, they should notice his priority of digital media, a part of the campaign which can be run without especially high costs."

Party strategists have also been inspired by the American experience. However, to get Macedonian voters drawn into Web politics in sufficient numbers in the first place, a more systemic approach is called for. During the American presidential campaign common "entrances" or "portals" to party politics on the main Web 2.0 sites were set up on established sites as on Facebook, YouTube and Twitter.

Some differences between the American and Macedonian party systems should also be noted. A national party in the USA and Macedonia is quite simply different entities. Population- and territorial size, as well as diversity, place different demands on local networking and autonomy, as well as effective coordination and communication between the localities. American parties also have a much looser structure, with relatively few members and dormant local branches. Macedonian parties on the other hand are still relatively strong organizations and less reliant on ad hoc networking. Thirdly, American elections are candidate-centered, in contrast to the party centered approach found in Macedonia. These differences may be reduced over time, as Macedonia – along with other European countries – is approaching a model with decoupled local branches, fewer members and more focus on individual leaders. But they are still significant enough to warrant the question whether Web 2.0 is more functional for American parties and therefore more “rational” to use for winning elections, exactly because these parties are more like network parties in the first place.

As such, it may therefore seem like a paradox that it is the SDSM and VMRO-DPMNE which have most fully embraced Web 2.0. They are one of the oldest parties and probably still have the most effective and vital party organization. However, this also means that the party has the resources and structure to effectively implement their Web 2.0 presence, provided the party leadership thinks it necessary. It is another useful media channel for communicating with members and voters.

The Internet is a unique forum for politics as it provides back and forth communication and allows for an exchange of information between users and sources. The Internet also offers its users greater access to information and the ability to express themselves in various online political arenas. In addition, individuals use the Internet as a tool to find and join groups that share their similar ideological, cultural, political and lifestyle preferences.

## 8. Reference

- [1] Marin, Alexandra, and Barry Wellman. "Social Network Analysis: An Introduction." *Computing in the Humanities and Social Sciences*. Web. 04 Apr. 2010. <<http://www.chass.utoronto.ca/~wellman/publications/newbies/newbies.pdf>>.
- [2] Oblak, Tanja. "Internet Kao Medij i Normalizacija Kibernetickog Prostora." *Hrčak Portal Znanstvenih časopisa Republike Hrvatske*. Web. 06 Apr. 2010. <[hrak.srce.hr/file/36810](http://hrak.srce.hr/file/36810)>.
- [3] "Mixing Friends with Politics: A Functional Analysis of '08 Presidential Candidates Social Networking Profiles

- Authored by Compton, Jordan." All Academic Inc. (Abstract Management, Conference Management and Research Search Engine). Web. 13 June 2010. <[http://www.allacademic.com/meta/p\\_mla\\_apa\\_research\\_citation/2/5/9/3/4/pages259348/p259348-1.php](http://www.allacademic.com/meta/p_mla_apa_research_citation/2/5/9/3/4/pages259348/p259348-1.php)>.
- [4] All Academic Inc. (Abstract Management, Conference Management and Research Search Engine). Web. 13 June 2010. <[http://www.allacademic.com/meta/p\\_mla\\_apa\\_research\\_citation/2/5/9/3/4/pages259348/p259348-1.php](http://www.allacademic.com/meta/p_mla_apa_research_citation/2/5/9/3/4/pages259348/p259348-1.php)>.
  - [5] "Alexa Internet - Website Information." Alexa the Web Information Company. Web. 20 June 2010. <<http://www.alexa.com/siteinfo>>.
  - [6] "What Is Web 2.0 - O'Reilly Media." O'Reilly Media - Technology Books, Tech Conferences, IT Courses, News. Web. 27 May 2010. <<http://oreilly.com/web2/archive/what-is-web-20.html>>.
  - [7] Web 2.0 Sites. Web. 28 May 2010. <<http://web2.ajaxprojects.com/>>.
  - [8] "Key Differences between Web1.0 and Web2.0." CiteSeerX. Web. 25 May 2010. <<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.145.3391>>.
  - [9] "Political Participation and Web 2.0." *Political Theory*. Web. 24 May 2010. <<http://arts.monash.edu.au/psi/news-and-events/apsa/refereed-papers/political-theory/allisonpoliticalparticipationandweb.pdf>>.
  - [10] "Alexa - Top Sites in Macedonia." Alexa the Web Information Company. Web. 01 June 2010. <<http://www.alexa.com/topsites/countries/MK>>.
  - [11] "What Is Social Networking? - Should You Join." *What Is Social Networking? - What Is Social Networking?* Web. 30 May 2010. <[http://www.whatissocialnetworking.com/Should\\_You\\_Join.html](http://www.whatissocialnetworking.com/Should_You_Join.html)>.
  - [12] Boyd, Danah M., and Nicole B. Ellison. "Social Network Sites: Definition, History". Web. 30 May 2010. <<http://jcmc.indiana.edu/vol13/issue1/boyd.ellison.html>>.

**S. Emruli**, received his bachelor degree from Faculty of Communication Sciences and Technologies in Tetovo SEE University (2006), MSc degree from Faculty of Organization and Informatics, Varaždin (2010). Currently works as professional IPA Advisor at Ministry of Local Self Government in Macedonia.

**M. Bača**, is currently an Associated professor, University of Zagreb, Faculty of Organization and Informatics. He is a member of various professional societies and program committee members, and he is reviewer of several international journals and conferences. He is also the head of the Biometrics centre in Varaždin, Croatia. He is author or co-author more than 70 scientific and professional papers and two books.

# A Framework for Modelling Software Requirements

Dhirendra Pandey<sup>1</sup>, Ugrasen Suman.<sup>2</sup>, A.K. Ramani<sup>2</sup>

<sup>1</sup>Member IEEE, Department of Information Technology,  
Babasaheb Bhimrao Ambedkar University,  
Lucknow-226025, India

<sup>2</sup>Schools of Computer Science & IT,  
Devi AhilyaVishwavidyalaya,  
Indore, MP, India,

## Abstract

Requirement engineering plays an important role in producing quality software products. In recent past years, some approaches of requirement framework have been designed to provide an end-to-end solution for system development life cycle. Textual requirements specifications are difficult to learn, design, understand, review, and maintain whereas pictorial modelling is widely recognized as an effective requirement analysis tool. In this paper, we will present a requirement modelling framework with the analysis of modern requirements modelling techniques. Also, we will discuss various domains of requirement engineering with the help of modelling elements such as semantic map of business concepts, lifecycles of business objects, business processes, business rules, system context diagram, use cases and their scenarios, constraints, and user interface prototypes. The proposed framework will be illustrated with the case study of inventory management system.

**Keywords:** Requirement Modelling, Inventory Control and Management System, Requirement Engineering (RE).

## 1. Introduction

Requirement Engineering (RE) is the process of collecting, analyzing and modelling software requirements in a systematic manner [1, 2, 3]. Requirement modelling is the major challenge of automotive software development [4]. One of the main problems of RE is to describe the requirements in terms of concise and manageable formal models and to integrate models to form a consistent and complete understanding of the software to be developed. Requirements modelling and analysis are the most important and difficult activities in the software development. Software development is becoming more mature by advancing development processes, methods, and tools. The famous Christ Honour and Other Served (CHAOS) has reported the statistics published by Standish Group show that still only about one third of software projects can be called successful, i.e. they reach their goals within planned budget and time [5]. Research on post-

mortem projects' analysis shows that the major problems comes when the requirements elicitation, analysis, specification, and management is not performed regularly. Deploying successful requirements process in a concrete organization is an important issue for software practitioners [6, 7]. While companies continue to use text-based documents as major means for specifying and analyzing requirements, the graphical requirements modelling are getting increasingly more attention in industry. This trend has increased after Object Management Group (OMG) standardized Unified Modelling Language (UML) [8]. As we know, that a picture is worth a thousand words. It is also applies in requirements analysis, where business people have to communicate with software developers, who do not know their domain and speak a different technical language. Additionally, UML tools support refining requirements models with design and implementation details for enabling traceability, validation, prototyping, code generation and other benefits. In large software development projects, these features are very important for evolving and managing requirement models.

There are some practical problems with UML complexity and lack of unified method or framework for requirements engineering [9]. Practitioners and scientists propose different approaches for eliciting and analyzing software requirements. The most popular tools that are used in modern requirements analysis is use cases. It was adopted by numerous companies, and described in requirements engineering textbooks [10, 11]. UML provides Use Case diagram for visualizing use case analysis artifacts. However, requirements analysis is not limited to use cases. In fact, they capture only end user-level functional requirements. A lot of research is also made in specifying business goals and processes, performing domain analysis. Although it was shown that UML might be extended and used for business modelling, the business modellers'



community was not satisfied by UML, and created a new Business Process Modelling Notation (BPMN), which has become OMG standard as well. In many cases, they also apply Integration Definition for Function Modeling (IDEF) notations [12, 13]. In domain analysis, analysts continue to apply old-style Entity Relationship (ER) notation, which was popular in database design since 70s [141]. A significant attention is paid to business goals, business rules, business object lifecycles, business roles and processes in organization, which also can be done using UML [15, 16].

Real-time and embedded system developers have also come up with a different flavour of UML – System Modelling Language (SysML). It defines requirements diagram and enables capturing various non-functional and detailed functional requirements [17]. Also, it establishes specific links between requirements and other elements. Most popular requirements text books introduce various diagrams based on both UML and other informal notations, e.g. system context diagram, and hand-drawn user interface prototypes [11, 18]. The mentioned requirements artefacts can be modelled using UML. Since UML is a general purpose modelling language with more than 100 modelling elements (UML meta classes) and without standardized method, practitioners apply it only fragmentally, and at the same time, they do not make use of its powerful capabilities to define consistent, integrated, and reusable requirements models. Various researches have already been performed to produce framework for creating UML models for MDD (Model-Driven Development) [12, 15]. This paper extends it with more focus on the details of a specific part of the framework by applying UML concepts for requirements modelling.

Most requirement documents are written in natural languages and represented in less structured and imprecise formats. Including requirement phase, artifacts created in phases of software life cycle are required to be modelled and integrated, so the traceability, consistency, and completeness can be ensured [19, 20]. The Organisation of paper as follows. We propose an effective framework for requirement modelling using some demonstrated examples, which is discussed in detail with various phases in Section 2. Future scope of this research is discusses in section 3. Finally, Section 4 describes the concluding remarks.

## 2. Requirements modelling framework

Most requirement documents are written in ambiguous natural languages which are less formal and imprecise. Without modelling the requirement documents, the

knowledge of the requirement is hard to understand [17, 18]. The lack of framework for guiding requirements models is one of the main issues. In academic community, researchers propose many detailed and focused requirements development methods [20, 21]. However, most of these methods resulting from academic research are too complex for practical application and solve just specific specialized issues. A simple and adaptable framework for requirements modelling with demonstrated examples are created using available tools on a realistic case study gives much more value for practitioners.

We have proposed requirements modelling framework using UML concepts for model-driven software development, which is shown in Figure 1. This framework consists of five major phases, namely; feasibility study, requirement collection and specification, analysis of business requirements, system requirement modelling and system design. Further, analysis of business requirements includes business conception and association, business objective life cycle, business tasks and methods and system requirement modelling incorporates actors, use cases and their scenario. The following subsections will discussed each phases of the proposed framework with the help of UML diagram and using examples.

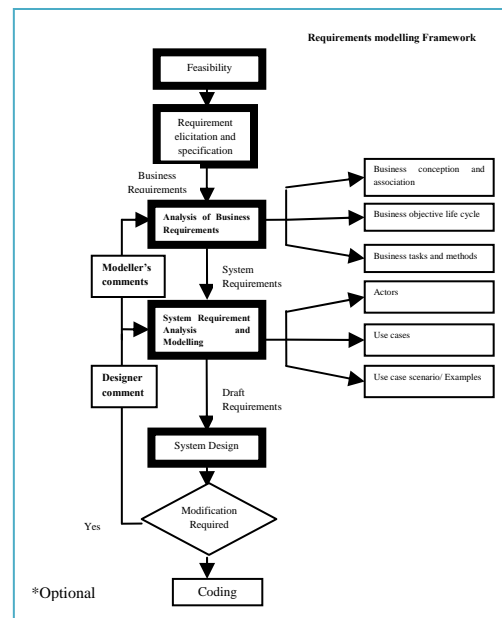


Figure 1: Requirement Modelling Process

### 2.1 Feasibility Study

Feasibility study starts when the developer faces the problem in existing system and hence recognizes a need for developing/ improving the system. It aims to



objectively and rationally uncover the strengths and weaknesses of the existing business or proposed venture, opportunities and threats as presented by the environment, the resources required to carry through, and ultimately the prospects for success. In its simplest term, the two criteria to judge feasibility are cost required and value to be attained. As such, a well-designed feasibility study should provide a historical background of the business or project, description of the product or service, accounting statements, details of the operations and management, marketing research and policies, financial data, legal requirements and tax obligations. Generally, feasibility studies precede technical development and project implementation.

## 2.2 Requirement elicitation, collection and specification

Requirement elicitation and development phase mainly focuses on examining and gathering desired requirements and objectives for the system from different viewpoints (e.g., customer, users, constraints, system's operating environment, trade, marketing and standard etc.). Requirements elicitation phase begins with identifying stakeholders of the system and collecting raw requirements from various viewpoints. Raw requirements are requirements that have not been analysed and have not yet been written down in a well-formed requirement notation. The elicitation phase aims to collect various viewpoints such as business requirements, customer requirements, user requirements, constraints, security requirements, information requirements, standards etc.

Typically, the specification of system requirements starts with observing and interviewing people [1, 2, 3]. Furthermore, user requirements are often misunderstood because the system analyst may misinterpret the user's needs. In addition to requirements gathering, standards and constraints are also play an important role in systems development. The development of requirements may be contextual. It is observed that requirement engineering is a process of collecting requirements from customer and environment in a systematic manner. The system analyst collects raw requirements and then performs detailed analysis and receives feedbacks. Thereafter, these outcomes are compared with the technicality of the system and produce the good and necessary requirements for software development [3].

Requirements requirement specification (SRS) document is produced after the successful identification of requirements. It describes the product to be delivered rather than the process of its development. Also, it includes a set of use cases that describe all the interactions

that users will have with the system/software [2]. In addition to use cases, the SRS also contains non-functional requirements. Non-functional requirements are requirements which impose constraints on the design or implementation. SRS is a comprehensive description of the intended purpose and environment for software under development. The SRS fully describes what the software will do and how it will be expected to perform. An SRS minimizes the time and effort required by developers to achieve desired goals and also minimizes the development cost. A good SRS defines how an application will interact with system hardware, other programs and users in a wide variety of real-world situations. Parameters such as operating speed, response time, availability, portability, maintainability, footprint, security and speed of recovery from adverse events are evaluated in SRS.

## 2.3 Analysis of business requirements

Many organizations already have established their procedures and methodologies for conducting business requirements analysis, which may have been optimized specifically for the business organization. However, the main activities for analysing business requirements are identifying business conception and association, determining business object life cycle, and identifying business tasks and methods. If these exist, we can use them. However, we must follow the following factors to create requirement models:

**(A) Identification of key stakeholders-** The first step toward the requirement analysis and collection is Identification of the key people who will be affected by the project. Such as, project's sponsor responsible users and clients. This may be an internal or external client. Then, identify the end users, who will use the solution, product, or service. Our project is intended to meet their needs, so we must consider their inputs.

**(B) Capture stakeholder requirements-** Another approach towards analysis of business requirement is capturing the requirement from stakeholders. In this approach, the requirement engineer requests stakeholders or groups of stakeholders for their requirements from various sources for the new product or service.

**(C) Categorize requirements-** Requirements can be classified into four categorized to make analysis easier for software design:

- **Functional requirements (FR)** – FR defines how a product/service/solution should function from the end-user's perspective. They describe the features and functions with which the end-user will interact directly.

- **Operational requirements (OR)** – OR operations that must be carried out in the background to keep the product or process functioning over a period of time.
- **Technical requirements (TCR)** – TCR defines the technical issues that must be considered to successfully implement the process or create the product.
- **Transitional requirements (TSR)** – TSRs are the steps needed to implement the new product or process smoothly. TSR indicates that how the requirements are behave as the consequence of external requirements

**(D) Interpret and record requirements-** Once we have gathered and categorized all requirements determine which requirements are achievable, and how the system or product can deliver them. The following steps should be taken to interpret the requirements:

- **Define requirements precisely** – Ensure that the requirements are not ambiguous or vague, clearly worded, sufficiently detailed, related to the business needs and listed in sufficient detail to create a working system or product design.
- **Prioritize requirements** – Although many requirements are important, some are more important than others, and budgets are usually limited. Therefore, identify which requirements are the most critical, and which are less.
- **Analyze the impact of change** – carry out an impact analysis to make sure that we understand fully the consequences our project will have for existing processes, products and people.
- **Resolve conflicting issues** – Sit down with the key stakeholders and resolve any conflicting requirements issues. We may find scenario analysis helpful in doing this, as it will allow all those involved to explore how the proposed project would work in different possible futures.
- **Analyze feasibility** – Determine reliability and easy-to-use the new product or system. A detailed analysis can help identify any major problems.

**Business conception and association:** Different methodologists have been proposed by various researchers for business conception and association techniques but still disagree on beginning of business information systems development [10]. In our proposed research, the starting point should be business concept analysis and analysis and their relationships which are shown in Figure 2. For this purpose we can apply simple organisational working model using only classes with names and without more

detailed information, associations with names and role multiplicities. Such models are discussed by business analysts and domain experts who are usually not familiar with object-oriented analysis and design.

Therefore, it is very important that all the other elements of the model, such as aggregations, compositions, generalizations, interfaces, enumerations, etc., should not be used for conceptual analysis. Keeping it simple enables even UML novices to understanding it after getting a little explanation. Additionally, we can provide textual descriptions for each of these concepts and generate printable or navigable domain vocabularies. We believe this should be the first artifact since it sets up the vocabulary, which should be used for defining other requirement model elements, cases, etc.

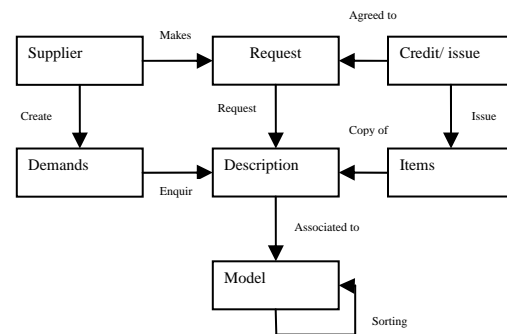


Figure 2: Analysis model for business conception and association

**Business object life cycle:** Requirements models are used when gathering requirements, and during systems analysis. Whether we consider eliciting requirements to be a separate activity, or a part of systems analysis, the importance of correct requirements must be a high priority for us. Building accurate models means that we can guarantee the correctness of our requirements. All engineering disciplines use models to develop the products they intend to build. Requirements models are used to discover and clarify the functional and data requirements for software and business systems. Additionally, the requirements models are used as specifications for the designers and builders of the system.

Organizations have business rules for managing business objects. In many cases, business rules regulate how important business objects change states and are applicable only when object is in a particular state. Requirement modelling is one of the important tools to understand these changes. The states also serve as a part of terminology, which will be used in other business and requirements models. State machine diagrams should be created only for those business concepts that have dynamic

states. Business modellers should define triggers on all transitions in state diagram.

In business modelling for transition triggers, most people use informal signals that in most cases correspond to actions of business roles. Also, time and property change triggers are used to express various states changes according to time or data based business rules. It is possible to define inner triggers that happen inside one state and doesn't fire a transition. In inventory management system, register (data store) is checking for availability of reservation for supplier. If available, supplier is assigned by a unique id to them and after that issue the item. Manufacturer notifies the overdue of product item and after one year the identified item will be notified as lost or damaged. Example of this concept is shown in Figure 3.

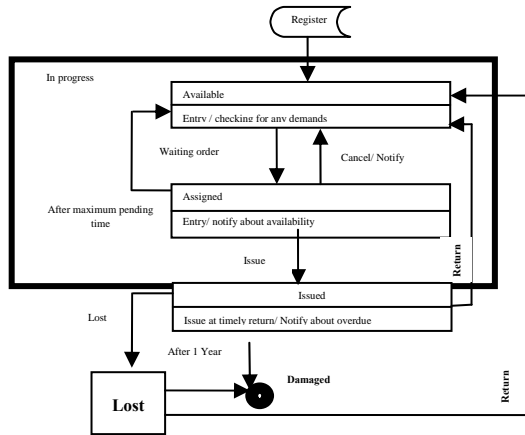


Figure 3: Business object lifecycle in Inventory Control and Management System

**Business tasks and methods:** After learning domain terminology and business rules concerning lifecycles of business objects, we can identify business tasks and methods, and associate roles to processes in which they are involved.

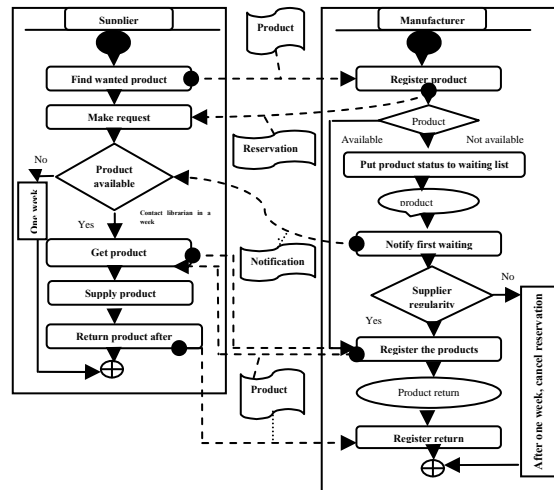


Figure 4: Inventory system process (tasks and methods)

We recommend to model business roles with actors, and business processes, if modellers need to visually separate it from system actors and use cases. The business roles association to business processes is best done within specialized use case diagram or editable relationship matrix.

In Figure 4, we are showing inventory processes with the supplier and manufacturer role perspectives. The role of supplier and manufacturer are different. Supplier starts the work with finding the wanted product at manufacturer site. Supplier makes reservation for the product; if the reservation is available he gets the item. After supplying the product he can be return the product due to damage or complaining by the customer with in prescribed date.

The first step in moving from domain analysis to requirements definition is use case analysis. We propose to do use case analysis using different steps such as identify the actors and group them into primary (main users), secondary (administration, maintenance, and support), external systems, and pseudo (e.g. time). We have defined main system use cases in a sketch use case diagram using pictorial form in figure 4.

The manufacturer registers the reservation of product, which is requested by the supplier. If the product is available, he may issue it to the supplier. If not, manufacturer put the reservation to the waiting list until the product is not available. On availability, manufacturer notify to the first waiting supplier (Supplier is too many). Otherwise he may cancel the reservation after prescribed date. The business processes are usually modelled in two forms, i.e. "as is", represents current situation, and "to be", represents target situation that should be reached after

automation or refactoring [10]. For software developers it is important to know which parts in target business processes the software system should implement or support.

## 2.4 System requirements modelling using case study

Requirement modelling is an important activity in the process of designing and managing enterprise architectures. Requirements modelling helps to understand, structure and analyse the way business requirements are related to Information Technology requirements, and vice versa, thereby facilitating the business-IT alignment. It includes actors, use cases and use case scenario. Each of these is further describe in following subsection:

**Actors:** An actor is a user or external system with which a system being modelled interacts. For example, in our inventory management system involves various types of users, including supplier, inventory management system, human resources, and manufacturer. These all users are actors. At the same time, an actor is external to a system that interacts with the system. An actor may be a human user or another system, and has some goals and responsibilities to satisfy in interacting with the system.

It is also necessary to generate actor who giving compact overview of the whole model. We have to prepare requirement specification model that incorporates the package details diagram, showing package use cases, their associations with actors and relationships between use cases including uses cases. For making good requirement modelling system engineer prepares activity diagrams visualizing scenarios of complex use cases. In model, the activities should be nested within appropriate use cases and assigned as their behaviours. And finally we describe use cases according to pre-defined templates, e.g. rational unified process use case document, actors in Figure 5.

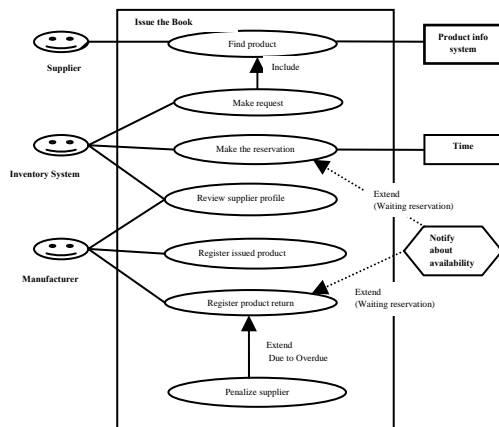


Figure 5: Issuing the product for supplier

**Use case and Use case scenario:** A use case in software engineering and systems engineering is a description of a potential series of interactions between a software module and an external agent, which lead the agent towards something useful. A use case diagram in the UML is a type of behavioral diagram defined by and created from a Use-case analysis. The purpose of use case is to present a graphical overview of the functionality provided by a system in terms of actors, their goals and any dependencies between those use cases. Also, it is useful to show what system functions are performed for which actor.

Requirement models are used to captures only functionality that the end-user needs from the system. The other requirements such as non-functional requirements or detailed functional requirements are not captured in standard requirement modelling diagrams. The simplest way is to describe those in simple textual format and include references to use cases, their scenarios, etc. Another approach is to create specific requirements modelling.

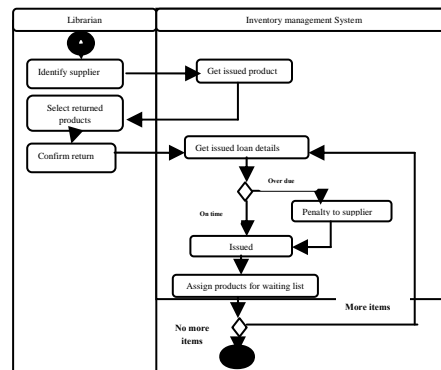


Figure 6: Register product return

For example, introduce stereotypes for each important requirement type with tags consisting requirement specific information and define types of links for tracing requirements, such as derive, satisfy, support. Another aspect on which system analyst's work in some projects is definition of data structure. It can be done using conventional requirement modelling diagrams. If necessary, object diagrams can also be used for defining samples for explanation or testing of data structure defined in class diagrams. Since the focus here is on data structure, class operations compartments can be hidden in the diagram (Figure 6).

Comparing to conceptual analysis, more elements are used here, such as attributes and association end specifications, enumerations, and generalization. Although such model is considered to be part of design, in practice quite often it is created and maintained by system analysts. For data-centric applications, it is very important to do data-flow diagrams showing information flows between different classifiers, e.g. system-context diagram indicates information flows from system to outside world entities, i.e. actors or external systems that need to be integrated.

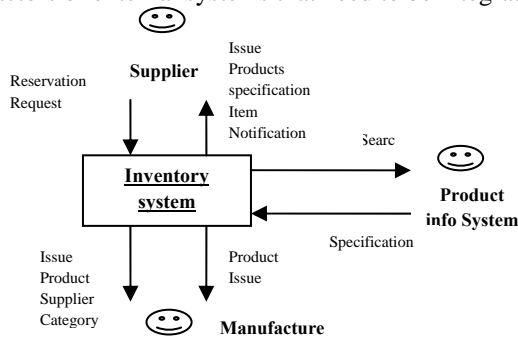


Figure 7: Information flow model

The previous requirements modelling artifact for which system analyst might be responsible is user interface prototypes. The prototype itself can theoretically be mapped to UML Composite Structure diagram. However, when focusing on separate screen prototypes, people sometimes loose the screens which can be used by each actor, and the possibilities to navigate from each screen to the other screens. For capturing this information, we can create GUI navigation map, which is shown in Figure 7. In Figure 7, we use state diagram, where each state represents a screen, in which user is at the moment, and transition triggers represent GUI events, such as mouse double-click or clicking on some button. Using this requirement model, system developers create an effective software on inventory control and management system. The user interface diagram model is shown in Figure 8.

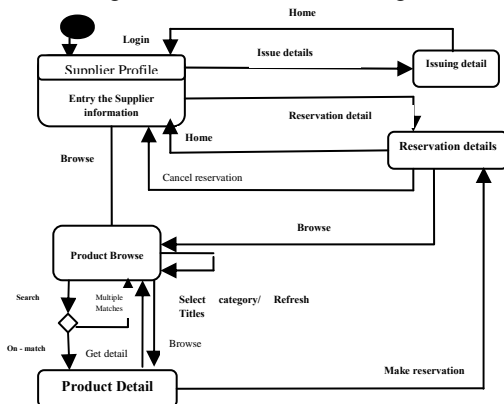


Figure 8: User interface diagram model

Finally, we emphasize that the requirements analysis work should be iterative and incremental. Also, the ordering of modelling tasks might be different based on taken approach, or some steps might be omitted.

### 2.5 System design

After the successful completion of system requirement and modelling phase, the draft (raw) requirement may be provided to the design team. Design team check the validity of these draft requirements and starts to design the system or software model. Basically, system design is the process of designing developing and implementation of the proposed system as per the requirement obtained during the analysis of existing system. The main objective of the system design is to develop the best possible design as per the requirements from users and working environment for operating the information system. It is the process of defining the architecture, components, modules, interfaces, and data for a system to satisfy specified requirements. Systems design is therefore the process of defining and developing systems to satisfy specified requirements of the user. Object-oriented analysis and design methods are becoming the most widely used methods for computer systems design. The UML has become the standard language in object-oriented analysis and design. It is widely used for modelling software systems and is increasingly used for high designing non-software systems and organizations.

After the designing of the system model, designer evaluates the efficiency of the design model. If any modification is remaining in the model, designer again checks the validity of requirements and asks for correction with comments. The process will stopped until the clear cut clarification is not received by the design team. This section is very important because according to the software engineering approach the design is the bridge the gap between requirement analysis and coding of the final software development

### 3. Discussion and future scope

The paper discusses implementation of requirement modelling for various requirements analysis purposes and mapping of conventional requirements artifacts into system elements. We have also presented some modelling aspects, which are necessary for ensuring that the requirements elements that are mapped to the same UML element can be differentiated. We can also find critics on using UML as requirements specification language, most of the issues can be solved using UML tool with rich

possibilities for modelling environment customization and extensions [18]. On the other hand, there are also suggestions to use more UML for requirements analysis and visualizations [20]. Multiple authors provide numerous papers on more detailed approaches to customizing unified modelling language for specific requirements modelling needs, such as analyzing scenarios, modelling user interface prototypes, refining requirements [21, 22]. Some researchers also suggest that UML can be specialized for early phase requirements gathering process but the proposed framework emphasizes that early phase modelling should focus on same types of artifacts with less detail.

#### 4. Conclusions

In this paper, we have discussed the major requirements artifacts described in requirements engineering literature can easily be mapped to elements of UML. Also, we have depicted a conceptual framework for requirements modelling with illustrated examples for inventory control and management system. Our future research work will focus on more detailed management for requirements modelling framework and development of different demo version for different management system.

#### References

- [1] D. Pandey, U. Suman, A. K. Ramani, Social-Organizational Participation difficulties in Requirement Engineering Process- A Study, National Conference on ETSE & IT, Gwalior Engineering College, Gwalior, 2009.
- [2] Dharendra Pandey, U. Suman, A.K. Ramani, Design and Development of Requirements Specification Documents for Making Quality Software Products, National Conference on ICIS, D.P. Vipra College, Bilaspur, 2009.
- [3] Dharendra Pandey, U. Suman, A.K. Ramani, An Effective Requirement Engineering Process Model for Software Development and Requirements Management, IEEE Xplore, 2010, Pp 287-291
- [4] M. Broy, I. Kruger, A. Pretschner and C. Salzmann. Engineering Automotive Software. Proceedings of THE IEEE. 95(2): 356-373, February 2007.
- [5] D. Rubinstein Standish Group Report: There's Less Development Chaos Today. SD Times, March 1, 2007.
- [6] J. Aranda, S. Easterbrook, G. Wilson Requirements in the wild: How small companies do it. 15th IEEE International Requirements Engineering Conference (RE 2007), pp. 39-48.
- [7] M. Panis, B. Pokrzywa, Deploying a System-wide Requirements Process within a Commercial Engineering Organization. 15th IEEE International Requirements Engineering Conference (RE 2007), pp. 295-300.
- [8] Object Management Group. Unified Modelling Language: Superstructure. Formal Specification, 15th IEEE International Requirements Engineering Conference (RE 2007), 2007.
- [9] G. Engels., R. Heckel, and S. Sauer, UML – A Universal Modelling Language? In M. Nielsen, D. Simpson (Eds.): ICATPN2000, LNCS 1825, pp. 24-38, 2000.
- [10] I. Jacobson, Object-Oriented Software Engineering. Addison Wesley Professional, 1992.
- [11] K. Wiegers. Software Requirements. 2nd edition, Microsoft Press, 2005.
- [12] Object Management Group. Business Process Modelling Notation Specification. Final Adopted Specification, version 1.0, 2006.
- [13] O. Noran. UML vs. IDEF: An Ontology-oriented Comparative Study in View of Business Modelling. Proceedings of International Conference on Enterprise Information Systems, ICEIS 2004, Porto, 2004.
- [14] P. Chen, P.-S. The entity-relationship model – toward a unified view of data. ACM Transactions on Database Systems (TODS), vol. 1 (1), 1976.
- [15] Van Lamsweerde, A. Goal-Oriented Requirements Engineering: A Guided Tour. RE'01 – International Joint Conference on Requirements Engineering, Toronto, 2001, pp.249-263.
- [16] M. Penker, H. Eriksson, E. Business Modelling With UML: Business Patterns at Work. Wiley, 2000.
- [17] Object Management Group. Systems Modelling Language. Formal Specification, version 1.0, 2007.
- [18] E. Gottesdiener, The Software Requirements Memory Jogger: A Pocket Guide to Help Software and Business Teams Develop and Manage Requirements. GOAL/QPC, 2005.
- [19] M. Glinz, Problems and Deficiencies of UML as a Requirements Specification Language. 10th International Workshop on Software Specification and Design, 2000, p.11 - 22
- [20] S. Konrad, H. Goldsby, K. Lopez, Visualizing Requirements in UML Models. International Workshop REV'06: Requirements Engineering Visualization, 2006.
- [21] H. Behrens, Requirements Analysis and Prototyping using Scenarios and Statecharts. Proceedings of ICSE 2002 Workshop: Scenarios and State Machines: Models, Algorithms, and Tools, 2002.
- [22] Da Pinheiro, P. Silva, The Unified Modelling Language for Interactive Applications. Evans A.; Kent S.; Selic B. (Eds.): UML 2000 – The Unified Modelling Language. Advancing the Standard, pp. 117-132, Springer Verlag, 2000.

**Dhirendra Pandey** is a member of IEEE and IEEE Computer Society. He is working in Babasaheb Bimrao Ambedkar University, Lucknow as Assistant Professor in the Department of Information Technology. He has received his MPhil Degree in Computer Science from Madurai Kamraj University, Madurai, Tamilnadu, India. Presently, he is perusing PhD in Computer Science from School of Computer Science & Information Technology, Devi Ahilya University, Indore (MP).

**Dr. Ugrasen Suman** has received his PhD degree from School of Computer Science & Information Technology (SCSIT), DAVV, Indore. Presently, he is a Reader in SCSIT, Devi Ahilya University, Indore (MP). Dr. Suman is engaged in executing different research project in SCSIT. He has authored more than 30 research papers.

**Professor (Dr.) A. K. Ramani** has received his ME and PhD Degree from Devi Ahilya Vishwavidyalaya, Indore (M.P.). Dr. Ramani has authored more than 100 research papers and executing several major research projects. Presently, he is the Head of the Department in SCSIT, Devi Ahilya University, Indore (MP).



# 3D Model Retrieval Based on Semantic and Shape Indexes

My Abdellah Kassimi<sup>1</sup> and Omar El beqqali<sup>2</sup>

Sidi Mohamed Ben AbdEllah University  
GRMS2I FSDM B.P 1796 Fez-Atlas, Morocco

## Abstract

The size of 3D models used on the web or stored in databases is becoming increasingly high. Then, an efficient method that allows users to find similar 3D objects for a given 3D model query has become necessary. Keywords and the geometry of a 3D model cannot meet the needs of users' retrieval because they do not include the semantic information. In this paper, a new method has been proposed to 3D models retrieval using semantic concepts combined with shape indexes. To obtain these concepts, we use the machine learning methods to label 3D models by k-means algorithm in measures and shape indexes space. Moreover, semantic concepts have been organized and represented by ontology language OWL and spatial relationships are used to disambiguate among models of similar appearance. The SPARQL query language has been used to question the information displayed in this language and to compute the similarity between two 3D models.

We interpret our results using the Princeton Shape Benchmark Database and the results show the performance of the proposed new approach to retrieval 3D models.

**Keywords:** 3D Model, 3D retrieval, measures, shape indexes, semantic, ontology.

## 1. Introduction

Recent 3D technologies scanning and 3D modeling lead to creation of 3D models stored in databases, which are used in various domains such as CAD applications, computer graphics, computer vision, games industry and medicine. Content based indexing and retrieval is considered as an important way of managing and navigating in these databases. Therefore, it become necessary to find an efficient method that allows users to find similar 3D objects for a given 3D model query which takes into account not only the shapes geometry, but also their semantics. Indeed, the use of low-level features to generate the objects descriptors can lead to big gap between low-level and high-level features. However, shape descriptors do not solve the problem of shape ambiguity because it does not consider the semantics of the model to be retrieved. 3D Model Retrieval system based on the semantic and ontology allows removing this ambiguity using combined semantic concepts and

geometrical information based on 3D shape indexes represented by concepts in ontology.

## 2. Related work

Several systems and approaches to compute similarity between 3D objects have been proposed in the literature [2] [3] [16] [18]. Most of those are based on either statistical property. Osada and al. [4] proposed the shape distribution based descriptor for extracting global geometric properties and detecting major differences between shapes. This method cannot capture detailed features. To calculate features, Volume-surface ratio, moment invariant and Fourier transform coefficients are used by Zhang and al. [5]. This approach is not efficient, but corrected in [28] using active learning. Vranic and al. [17] proposed the ray based approach, which extracts the extents from the center of mass of the object to its surface. The feature vectors constructed using this method is presented in a frequency domain by applying the spherical harmonics.

For the 3D model-semantic problem, many approaches have been proposed. The work presented in European Network of Excellence AIM@SHAPE [15] has shown the benefits of using semantic indexing based on ontology. The authors introduce knowledge management techniques in modeling the form in order to find 3D objects in terms of knowledge. In the paper [6], author explores an ontology and SWRL-based 3D model retrieval system Onto3D. It can infer 3D models semantic property by rule engine and retrieve the target models by ontology. To add semantics to geometry, Marios in [7] analyzes the 3D shape and can extract and combine knowledge and implicit information coded in the geometry of the digital content object and its sub-parties (volume, surface ...), then it allows the segmentation of 3D shapes based on semantics. The semantic description of an object based on the ontology and matching this description with the low level features such as color, texture, shape and spatial relationships [8] [9] also are used to classify and indexing images. In paper [10], authors incorporate semantics provided by multiple class labels to reduce the size of

feature vector produced by bag-of-features [11] exploiting semantics.

Various studies have also shown interest using shape indexes based indexing. For the shapes characterization and binary digital objects, Thibault [1] [12] presented a study implementing a set of values obtained by calculation of shape indexes. In this study, the author has shown that the use of shape indexes family is a robust and efficient tool in object recognition, and that flexibility and diversity shape indexes allow the creation of shape indexes for each family shapes to be studied. Rectilinearity shape index is proposed by Z.Lian in [13] to describe the extent to which a 3D mesh is rectilinear. This shape index has several desirable properties such as robustness and invariance to similarity transformation. In [14], large shape indexes are described and demonstrated (e.g. Eccentricity, Elongatedness, Circularity, Squareness, Ellipticity, Triangularity, Rectangularity, Rectilinearity, Sigmoidality, Convexity, Symmetry, Chirality). The author notes that selects the most appropriate measures depends on their suitability for particular applications. Corney and al. [27] describe the coarse filter for classifying 3D models. Several shape indexes are computed based on convex hull ratios such bounding-box aspect ratio, hull crumpliness, hull packing, hull compactness, etc.

In this paper, we suggest the implementation of two methods to retrieval a 3D object in database: the geometric method, which uses the measures and 3D shape indexes and Clustering-based Semantic to fill the gap between semantic concepts and low-level features. Motivation for using shape indexes is to extract visual concepts easily, and semantic information can be extracted using unsupervised learning method. These shape indexes, calculated from measures taken from the 3D model, are organized as semantic concepts in an ontology using OWL [19] and questioned by the SPARQL [20] query language to extract similarity between 3D models.

### 3. System overview

The proposed content-based retrieval system for 3D models consists of two processes: inline that interacts with the user and offline that the system computes descriptors for 3D models (Fig. 1). In both processes, the system extracts the measures of the model, calculating the shape indexes and extract semantic concepts.

The user can navigate in the database and sends a 3D request to the server. The system receives the query model and compares its descriptor with the descriptors of all models of class membership. This phase requires the appropriate distances to signatures, but also strategies to find semantically similar models in visual concepts [22] such as contour-shape, color or texture.

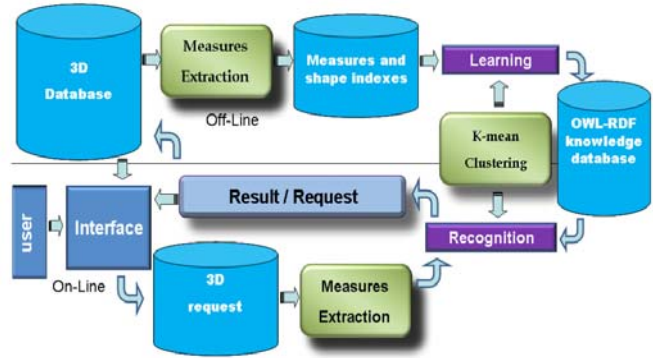


Fig. 1 Overview of the proposed system

Our 3D Database is composed of Princeton Shape Benchmark 3D models [16] that are stored in a format (\*.off) which represents the shape of 3D models by polygonal mesh, with the list of vertices  $V = \{v_1, v_2, \dots, v_N\}$  and triangular facets  $i = \{i_1, i_2, \dots, i_R\}$  defined by points  $i_r = (v_{n,1}, v_{n,2}, \dots, v_{n,k_r})$ . Where  $k = 3$  for the triangle mesh. Fig. 2 shows some examples of representations.



Fig. 2 various representations of the rabbit

As shown in this figure, there are many ways to represent a model (e.g. Point Set, Polygon Soup, Polygonal Mesh and Solid Model).

### 4. Measures and shape indexes

Shape is the most important property that allows predicting more facts about an object than color or texture. Shape index is the shape descriptor that is defined as any parameter, coefficient or combination of coefficients for providing quantitative information on the shape. Moreover, shape index must be dimensionless and has invariant to rotation and translation as property. Measure is a numerical value or set of numerical values "measured" on the shape. Shape indexes and measures definitions are detailed in [1]. Shape indexes are computed from the measures of the whole 3D model and have provided global information such as the size and the shape and are chosen for their ratio simplicity/effectiveness. The proposed method approach requires neither initial segmentation step nor the preprocessing.

#### 4.1 Measures

To compute 3D shape indexes, we directly compute 3D measures on the 3D model or transforming 2D measures. The most important 3D measures are surface area and volume. With 3D polygonal model representation, we can compute these measures [5] as follow:

$$area = \frac{1}{2} \sum_i^N |(V_{i,1} - V_{i,0}) \times (V_{i,2} - V_{i,0})| \quad (1)$$

$$Volume = \frac{1}{6} \sum_i^N (-V_{i,2}^x V_{i,1}^y V_{i,0}^z + V_{i,1}^x V_{i,2}^y V_{i,0}^z + V_{i,2}^x V_{i,0}^y V_{i,1}^z - V_{i,0}^x V_{i,2}^y V_{i,1}^z - V_{i,1}^x V_{i,0}^y V_{i,2}^z + V_{i,0}^x V_{i,1}^y V_{i,2}^z) \quad (2)$$

V is a vector containing the coordinates of the vertices of the triangle i.

These measures are used directly for calculating 3D shape indexes without transforming 2D measures. For other 3D measures, the 2D measures are used. For example, to calculate the radii, we use the distance between the centroid and a point on the surface area instead of the distance between the centroid and a point on the perimeter. There are other measures, which are dimensionless and shape indexes like a number of holes. In practice, we used the following measures: Volume, Surface area, Ferret diameter, Small and large radii, main axis and plan. In fact, the principal component analysis method is employed and three sets of main axes and planes are obtained. Ferret diameter is the longest distance from two contour points of the 3D object. These measures are used as semantic concepts in ontology and allow to define the spatial relationships. We consider that each measure is the entity.

#### 4.2 Shape indexes

From these basic measures, one can calculate the 3D shape indexes. Surface area (1) and volume (2) may be used as measures for calculating 3D shape indexes like VC (3) and AC (4), which can be considered as the basic descriptors of shape.

$$VC = \frac{V}{V(C_H)} \quad (3)$$

$$AC = \frac{A(C_H)}{A} \quad (4)$$

V and A are respectively the 3D model volume and surface area.  $C_H$  is a convex hull that is the minimum enveloping boundary.

AC and VC (called Area convexity index, Crumpliness [27] or Rectangularity and Volume convexity index) are easy to compute and are very robust with respect to noise [1]. Moreover, these shape indexes can distinguish

between shapes like angular and rounded objects [23]. Area convexity index and Volume convexity index tell us about the shape of the object, but it is difficult to identify any shape from these 3D shape indexes. Therefore, it is necessary to use a set of 3D shape indexes and combine them to retrieval the 3D model. These 3D shape indexes should be calculated very quickly and interpret the results. Basically, shape index has two types; compactness-based and boundary-based shape indexes.

Various compactness measures are used. For this reason, an early attempt to develop the compactness index is based on the values of perimeter and area. These 2D measures allow calculating the Isoperimetric shape index as follows:

$$\frac{\sqrt{4\pi S}}{P^2} \quad (5)$$

P and S are respectively the perimeter and surface of shape. This 2D shape index, defined between 0 and 1, is based on the surface to the perimeter ratio and reaches the value unity for a disk. We can also calculate the 2D circularity index shape as follows:

$$1 - \frac{\sqrt{4\pi S}}{P^2} \quad (6)$$

In 3D models, the perimeter becomes the surface area, and the surface becomes the volume. A ratio between surface area and volume is commonly used in the literature to compute compactness of 3D shapes. With this ratio an IsoSurfacic shape index can be obtained as follows:

$$I_s = 6 \frac{\sqrt{\pi V^{1/3}}}{A^{1/2}} \quad (7)$$

V and A are respectively the volume and surface area of the 3D model.

IsoSurfacic shape index is a compactness indicator which describes the form based on the surface area-to-volume ratio. Sphericity is another specific shape index for indicating compactness of a shape. It is a measure of how spherical an object is. It can be also calculated from surface area and volume 3D measures (8). The Sphericity (S) is maximum and equal to one for a sphere.

$$S = \frac{\pi^{1/3} (6V)^{2/3}}{A} \quad (8)$$

The Sphericity index shape (S) is very fast in computing. However, it is unsuited as a parameter of elongation. The latter is defined as quality of being elongated. The elongation, in this paper, is the boundary based and can be measured as the ratio of the smallest radius on the greatest radius (9) or ratio major on minor axes called Eccentricity.

$$E = \frac{R_{\min}}{R_{\max}} \quad (9)$$

The ratio of the maximum Ferret diameter and the minimum Ferret diameter is also used as the elongation parameter. We have included two aspect ratios of the bounding box for a 3D model in our system due to the

simplicity of computation and its relevance to 3D retrieval: compactness and complexity. Compactness is defined as the non-dimensional ratio of the volume squared over the cube of the surface area [27]. Complexity is defined as the surface area of the convex-hull divided by the volume of the convex-hull. There are several other shape indexes to calculate the elongation or compactness of a shape: (Isosurfacic Deficit, Morton Spread, geodesic Elongation, variance...).

Shape indexes calculated are quick to compute, easy to understand and were chosen mostly for their simplicity and are invariant to rigid motions such as translations and rotations. However, it should be noted that there are some shape indexes that are do not classify objects in the same way.

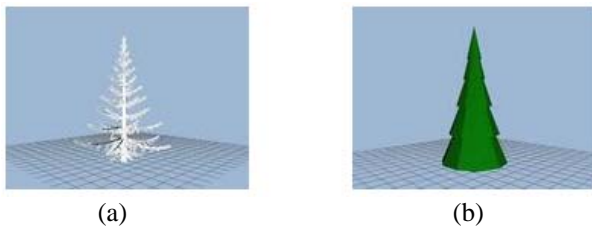


Fig. 3 Two models compared with different shape index.

The radius elongation index (9) for example, considers the model (a) in Fig. 3 and (b) similar, since they have almost the same radii, while the Volume convexity shape index (4) considers them different. Therefore, the necessity to combine several shape indexes for computing the most relevance.

## 5. Clustering-based semantic

Although the shape indexes calculated to provide global information on the 3D model and contain compactness and elongated indicators, the problems connected with 3D model retrieval are not still resolved. The first one regards the 3D shape indexes: they are insufficient to describe the 3D model in a generic 3D database; although these are relevant. Therefore, the necessity to combine several 3D shape indexes to augment our knowledge base with semantic concepts using, in our case, the ontology and spatial relationships. Second problem is caused by the semantic gap between the lower and higher level features. To reduce this ‘semantic gap’ we use machine learning methods to associate shape indexes with semantic concepts and ontology to define these semantic concepts as shown in fig. 4. In this paper, 3D shape indexes are used to represent visual concepts [22] of a 3D object.

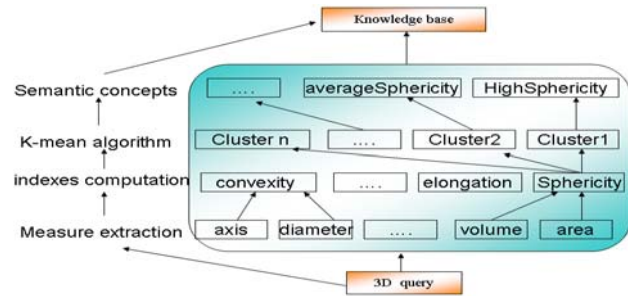


Fig. 4: Definition of semantic concepts and Knowledge base augmented and guided by a 3D Shape index ontology to describe the 3D models.

To be interpreted as visual concepts, a link must be established between computed numerical descriptors and symbolic visual concepts [8]. In our case, measures and shape indexes are clustered by a k-means algorithm into semantic clusters. The notion of similarity is based on each category of 3D shape indexes or measures like in Fig. 4. This approach is divided into the following steps: measure extraction; clustering and definition of semantic concepts. From the 3D Database, the three steps are repeated for each 3D shape index to define semantic concepts. Therefore, 3D model is described by a set of the numerical value associated with semantic concepts. We should create a database describing all models by the semantic concepts guided by a 3D Shape indexes ontology and relations among entities. The ontology defines a database structure as containing of a set of concepts that can describe qualitatively the visual semantic concepts and should allow similarity searches.

## 6. Ontology

Ontology is a set of concepts and useful relations to describe a domain, and thus makes more explicit the implicit semantics of models. One advantage of shape indexes is its flexibility to create other shape indexes for each model to be indexed in a domain-specific. In this paper, ontology is employed to allow the user to query a generic 3D collection, where no domain-specific knowledge can be employed, using the 3D model as query. The Ontology has been used to organize semantic concepts that are defined by the k-mean algorithm (e.g. Sphericity, elongation, convexity...). It includes other concepts such as semantic entities (e.g. lines, points, surface, and plan), a set of spatial relations and some axioms (transitivity, reflexivity, symmetry). The proposed ontology is represented in Ontology Language OWL [19], is the W3C recommended standard for ontology that precise formal semantics. As shown in Fig. 5, the OWL is structured into two parts: The first part contains shape index concepts and regroups the descriptors into classes



according to their characteristic properties: The topological descriptors and geometric descriptors (Fig.6). The second part contains the concepts spatial or entities together in primitive geometric: point, line, surface, Plan...

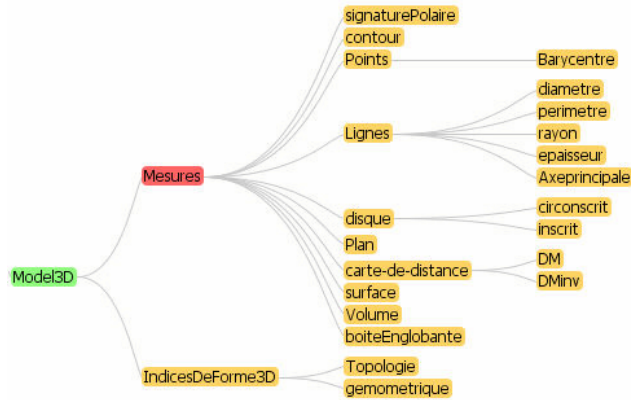


Fig. 5 The structure of our ontology

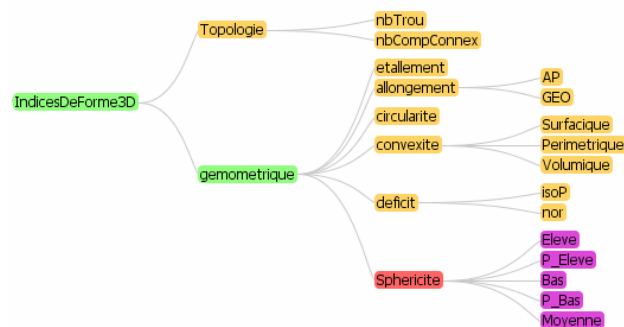


Fig. 6 The partial hierarchy of domain concepts of geometry

The structure of the ontology is represented in OWL as follows:

```
<owl:Class rdf:about="http://www.exemple/ontologie#Mesures">
  <rdfs:subClassOf>
  <owl:Class
rdf:about="http://www.exemple/ontologie#Modèle3D"/>
  </rdfs:subClassOf>
  </owl:Class>
  <owl:Class
rdf:about="http://www.exemple/ontologie#IndicesDeForme3D">
  <rdfs:subClassOf>
  <owl:Class
rdf:about="http://www.exemple/ontologie#Modèle3D"/>
  </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:about="http://www.exemple/ontologie#Points">
  <rdfs:subClassOf>
  <owl:Class rdf:about="http://www.exemple/ontologie#Mesures"/>
  </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:about="http://www.exemple/ontologie#Lignes">
  <rdfs:subClassOf>
  <owl:Class rdf:about="http://www.exemple/ontologie#Mesures"/>
```

```
</rdfs:subClassOf>
</owl:Class>
...
<owl:Restriction>
  <owl:maxCardinality
rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
  >1</owl:maxCardinality>
  <owl:onProperty>
  <owl:DatatypeProperty
rdf:about="http://www.exemple/ontologie#hasURL"/>
  ...
```

Ontology contains the concepts and their relations and facilitates the inference the spatial relation. The implicit rules are defined using OWL properties such as similarity owl: SameAs.

```
<RDF:Description rdf:about="#sphericity">
  <owl:sameAs rdf:resource="#circularity"/>
</Rdf:Description>
```

We can define other explicit rules to infer spatial relationships based on other relationships. For example, the position "leftCenter" has a unique meaning when associated with some information.

## 7. Spatial relationships

Shape indexes calculated are globally characterized the shape. Without segmenting the model, we calculated the local characteristics using spatial relationships that are usually defined according to the location of the measure in the 3D model. In our method, spatial relationships are defined by measures or entities that can increase the quality of detection and recognition of the model content and can disambiguate among models of similar appearance including for example the meaning of orientation and respect the distances. Therefore, other concepts are added to the 3D shape indexes to describe position, distances and orientation of an entity in the 3D model. There are various entities that need spatial relationships to describe 3D model to represent correctly the 3D models content. In this paper, the following relationships are described (Fig. 7):

- Metric (distance, area...)
- Orientation (near of, left of ...)
- Topology (Inclusion, adjacent ...).

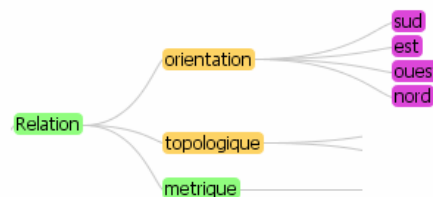


Fig. 7: Partial hierarchy of relationships.

The notion of position, distance and orientation in spatial relations are dependent on the notion of the frame of reference. The object centroid is used as the frame of

reference to compute measures and to respect proprieties: Rotation and translation. Then, the method does not require preprocessing for these properties. The bounding box centroid is used as the frame of reference to describe concept of position, distance and orientation. Therefore, to calculate the position "centered", we should calculate the Euclidean distance between the center of the 3D model and bounding box centroid. Entities such as the 3D model centroid, lines (e.g. radii, diameter and axes), plan and its minimum bounding box are used to calculate distances in order to provide spatial information. The distances can be computed from a point to point, line to line, point to line, point to plan and line to plan. In practice, we used the following distances: Distance between radii, Distance between radii and Diameter, Distance between two centers: 3D model centroid and bounding box centroid and A3, D1, D3, D4 introduced in [4].

To describe the distance relationship between two 3D models, the following distances are usually used: very near, near, far, far away. However, such distance relationships single are not sufficient to represent the 3D model content ignoring the topological and directional relationships. To get an idea about the overall direction of the entities in the 3D model, main axes can be used. In fact, the main axes of the 3D model can be calculated, employing the principal component analysis method, and the value of its direction is given by the angle with the axes of the bounding box. The example is the following relationship: RightOf; LeftOf; Above; Below...

We are also interested in topological relationships among entities that are related to how objects interconnect. In this paper, we adopt the topological relationships as shown in table 1. The RCC-8 [24] [25] relations can be used for taking into account spatial relations. RCC (Region Connection Calculus) is a logic-based formalism to symbolically represent and reason with topological properties of objects [14]. Topological reasoning can be implemented based on Pellet engine [21].

Table 1: Topological relations implemented in our system

Point-Point	Point-Line	Line-Line	Line-Plan
Overlap	On	Cross	Contained
Adjacent	Adjacent	Not Cross	Adjacent

Based on the spatial relationships and their properties, we build the ontology using the web ontology language (OWL).

## 8. Method for Classification Database

Each model of database, in our content based indexing and retrieval system for 3D models, is represented by two descriptors considered signatures of the 3D model: semantic concept and 3D shape indexes. To increase the identification rate and decrease the time to search for items, we have developed and implemented a classification by applying the k-Means algorithm in the 3D shape index space. K-Means is an efficient classification approach and very easy. Each model of database is clustered by the K-Means algorithm using the Euclidean distance as a similarity measure. Classification based on 3D shape indexes allows a global classification of models and it can detect major differences between shapes. Fig. 8 shows some classes of objects.



Fig. 8: 3D models of some Clusters

3D models are classified into clusters regardless of their spatial positions and according to the similarity of their 3D shape index.

## 9. SPARQL engine and similarity

Based on the semantic concepts and the 3D shape indexes introduced, the similar 3D model retrieval will be conducted. To this end, query by concept and numeric value is proposed to evaluate the similarity between two 3D models as has been shown in Fig. 9.



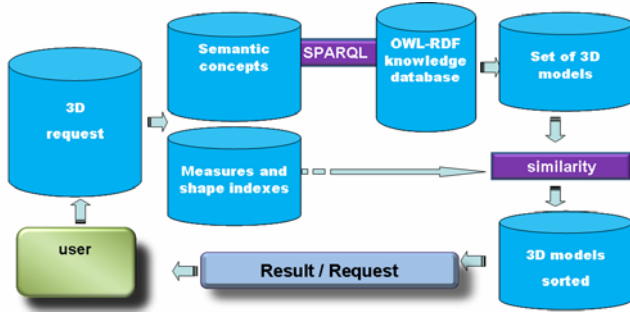


Fig. 9: The evaluation of similarity by semantic and numeric query.

SPARQL, a rich query language for OWL-DL ontology, is used to query the knowledge contained in the OWL ontology for the extraction of implicit and explicit semantics that are included in the model OWL. For example, the query: "Show all 3D Models URLs of a given cluster with a high sphericity and variance" is written in language SPARQL:

```
String jungle=jenaTools.findBasicNameSpace(ONT_MODELE);
String prologI = "PREFIX jungle: <"+jungle+">";
String qr=prologI + NL+"select* where " +
"/"+
"?3Dmodel jungle:hasCluster ?hasCluster FILTER (?hasCluster =
"+Cluster+")." +
"?3Dmodel jungle:hasSphericity '"+sphericity+"'" +
"?3Dmodel jungle:hasVarianceSurfacique '"+variance+"'" +
"?3Dmodel jungle:hasURL ?hasURL " +
"}";
```

SPARQL admits the use of numeric values to compute the similarity on the retrieved models that are semantically similar. The query can be easily adapted to obtain the distance between any pair of 3D models. Therefore, the similarity between two models is measured through the use of distance between their 3D shape indexes. To define the distance between two points, different metrics could be implemented. The most famous and used metric is the Euclidean distance or as it is called "Manhattan" which is just a special case of Minkowski measure:

$$L_p = \left\{ \sum_{i=1}^n |Z_i - x_i|^p \right\}^{1/p} \quad (10)$$

Depending on the parameter p: if p = 1 the distance is "city block" or Manhattan and when p = 2 is Euclidean distance.

In our system, the Euclidean distance is used to measure the similarity between 3D shape indexes. But, the latter does not have the same importance in the recognition process. Therefore, to provide the best results, it is necessary to combine several 3D shape indexes to compute the most relevant ones. A simple approach for the

combination of these 3D shape indexes is to calculate the weighted sum of the distances. The following formula which is used to determine the degree of similarity S between two 3D models has been implemented to calculate the distances:

$$S = \sum_{i=1}^n W_i L_p(SI)_i, \sum_{i=1}^n W_i = 1 \quad (11)$$

Where  $W_i > 0$  ( $i = 1, 2 \dots n$ ), are the weights of 3D shape index (SI)<sub>i</sub> and n number of shape indexes.

Weights are calculated and normalized during learning by k-mean algorithm using precision, recall and F-measure that allow the comparison of the performances of 3D shape indexes. Therefore, for each 3D shape index, we compute the average recall (aR) and precision (aP) on the entire 3D shape index:

$$aR = \sum_{i=1}^n \frac{r(SI)}{n(SI)}, aP = \sum_{i=1}^n \frac{r(SI)}{r(SI) + w(SI)} \quad (12)$$

"n(SI)" is the number of models labeled by SI. "r(SI)" is the number of models initially labeled by SI and the system which has returned with the same SI. "w(IF)" number of the unlabeled model by the SI and found by the system with the same SI. F-measure F is the weighted harmonic mean of precision and recall. The formula of F-measure is as follows:

$$F = 2 \frac{aRaP}{aR + aP} \quad (13)$$

When using average recall (aR) and precision (aP), it is important to specify the number of shape indexes for the finding of at least one model.

## 10. Experimental Results

Java language has been used to develop our content-based retrieval systems for 3D models. The tests are performed on the Princeton Shape Benchmark Database which contains 1814 objects that are given by triangular meshes and classified by semantic aspect. The concepts of ontology have been created by learning phase whose development has been realized with OWL and the Java programming tool (Jena). The library Jena contains inference engine customizable and offers the possibility of including an external reasoner. The display of the ontology is done with the API (Application Programming Interface) OWL2Prefuse.

Our programs are compiled under the windows platform, using 1.4 GHz, Core 2 Duo machine with 1 GB memory. The average time used to compute all shape indexes is 0.6 seconds for a model, using the Princeton Shape Benchmark Database.

As has been shown in Fig. 1, in the online process, the user submits a query model selected from the 3D

collection. During this process, shape indexes are computed, and we can directly retrieve models as will be shown in Fig. 11 (a) and Fig. 12 (c) using our descriptor or Area Volume Ratio Descriptor [5] that is not efficient.

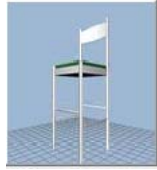
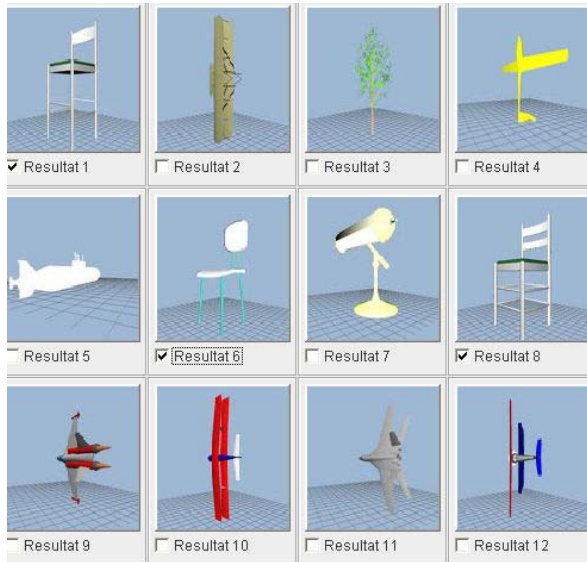


Fig. 10: query model



(a)

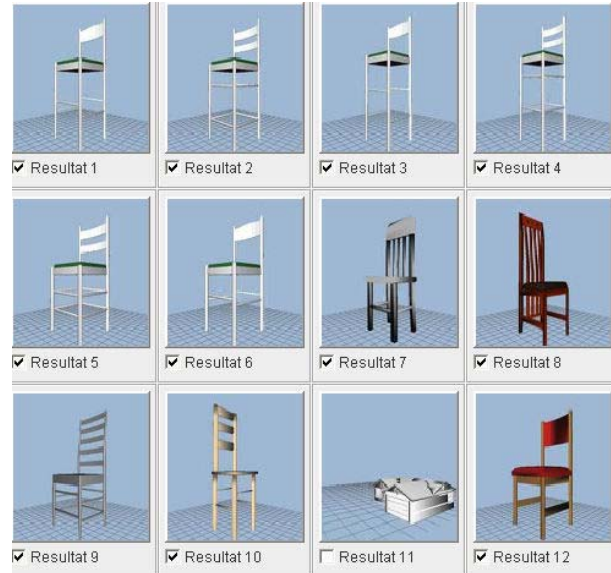


(b)

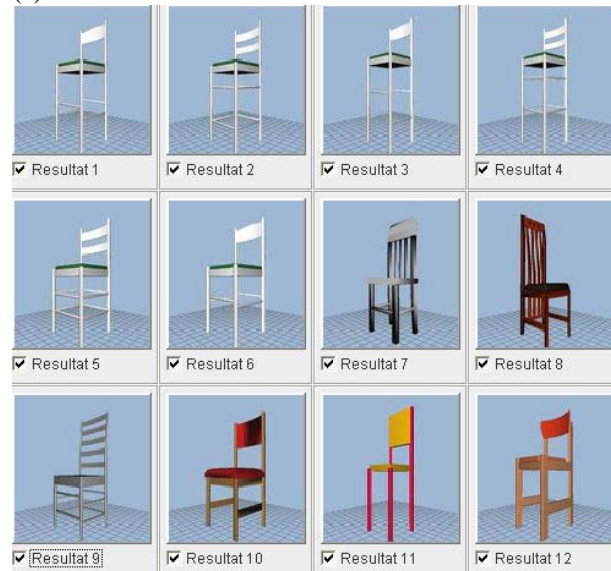
Fig. 11: (a) Models found by Area Volume Ratio descriptor without introducing the semantic descriptor. (b): Models found with Area Volume Ratio descriptor introducing our semantic descriptor.

In order to retrieve 3D models by introducing the semantic descriptor Fig. 11 (b) and Fig. 12 (d), the query is labeled before the search happens with a semantic concept by associating 3D shape low-level features with high-level semantic of the models.

The 12 most similar models are extracting and returning to user by 2D images. To visualize the 3D models in the 3D space, the user clicks the button or image.



(c)



(d)

Fig. 12: (c) Models found with our descriptor without introducing semantic descriptors. (d) Models found with our descriptor introducing the semantic descriptor.

For the evaluation of the performance of our system based on shape indexes and semantic concepts descriptors, we

used also the Recall and Precision. In this sense, we have compared our descriptor to the descriptor based on the volume area ratio proposed by Vranic and al. [5] that is implemented in our system. The volume area ratio is incorporated in our system as a shape index, and according to the evaluation the relevance, it is insufficient alone to characterize a 3D model and is not a performing descriptor, but it is used here to evaluate the performance implemented semantic and ontology.

Two methods are used for each descriptor: content-based and semantic-based retrieval. Fig. 13 shows the Precision-Recall diagram for each one of the two methods.

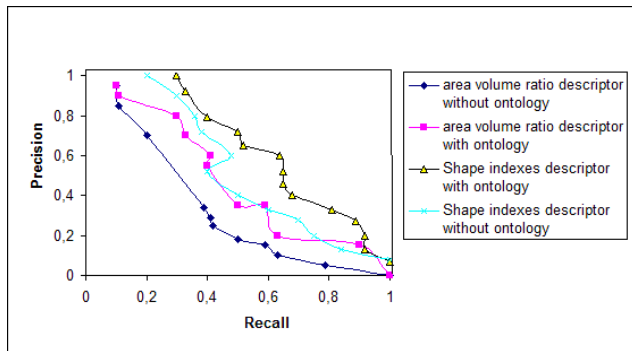


Fig. 13: The precision-recall curves of 3D model retrieval with and without ontology using different descriptors.

Fig. 13 shows that our proposed semantic descriptor performs well, and the descriptor Area Volume Ratio with ontology is compared to our descriptor without ontology justifies the use of ontology and semantic based retrieval as a most efficient method.

The developed classification based on shape indexes reduces the similarity gap, and the retrieval method by introducing the semantic descriptor is considered as more efficient than the one based solely on the shape indexes or Area Volume Ratio. This performance is linked to the combination of shape indexes and semantic concepts structured in ontology.

## 11. Conclusion

A new method for 3D models retrieval has been introduced in this article. The method combines semantic concepts and 3D shape indexes which are structured in ontology. The new approach is tested with a large 3D database using the developed search engine, which allows us to show the relevance of our method. The results are promising and show the interest of our approach.

To complete our work, it is interesting to improve our system by another method of robust classification based on semantics. For the very soon future, the shape index will be enriched with textures and color indexes.

## References

- [1] G. Thibault, Fertil B., Sequeira J., Mari J-L. "Shape and texture indexes. Application to cell nuclei classification" *MajecSTIC*, vol. 15, no 2 (117p.) pages 73-97 2010.
- [2] P. Min, J. A. Halderman, M. Kazhdan, and T. Funkhouser, "Early experiences with a 3D model search engine," in *Proceedings of the 8th International Conference on 3D Web Technology*, pp. 7–18, Saint Malo, France, March 2003.
- [3] Y.-B. Yang, H. Lin, and Q. Zhu, "Content-based 3D model retrieval: a survey", *Chinese Journal of Computers*, vol. 27, no. 10, pp. 1297–1310, 2004.
- [4] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin, "Shape distributions", *ACM Transactions on Graphics*, vol. 21, no. 4, pp. 807–832, 2002.
- [5] Cha Zhang and Tsuhan Chen. "Efficient feature extraction for 2D/3D objects in mesh representation". In *IEEE International Conference on Image Processing (ICIP 2001)*, pages 935-938, Thessaloniki, Greece, October 2001.
- [6] Xin-ying Wang, Tian-yang Lv, Sheng-sheng Wang, Zhengxuan Wang: "An Ontology and SWRL Based 3D Model Retrieval System". *AIRS* : 335-344. 2008.
- [7] Marios Pitikakis, Chiara Catalano: "A semantic-based framework for managing, searching and retrieving 3D resources" in *Conference Program of the VSMM 2009*.
- [8] N. Maillot and M. Thonnat. "A weakly supervised approach for semantic image indexing and retrieval". In *International Conference on Image and Video Retrieval (CIVR)*, volume 3568 of *Lecture Notes in Computer Science*, pages 629-638. Springer-Verlag Berlin Heidelberg, 2005.
- [9] N. Maillot, M. Thonnat, A. Boucher, "Towards ontology based cognitive vision", in: J. L. Crowley, J. H. Piater, M. Vincze, L. Paletta (Eds.), *Computer Vision Systems*, Third International Conference, ICVS, Vol. 2626 of *Lecture Notes in Computer Science*, Springer, 2003.
- [10] Ryutarou Ohbuchi, Masaki Tezuka, Takahiko Furuya, Takashi Oyobe, "Squeezing Bag-of-Features for Scalable and Semantic 3D Model Retrieval", *Proc. 8 th International Workshop on Content-Based Multimedia Indexing (CBMI) 2010*, Grenoble, France. 23-25 June 2010,
- [11] R. Ohbuchi, K. Osada, T. Furuya, T. Banno, "Salient local visual features for shape-based 3D model retrieval", *Proc. SMI '08*, 93-102, (2008).
- [12] Thibault G., Devic C., Horn J.-F., Fertil B., Sequeira J., Mari J.-L., "Classification of cell nuclei using shape and texture indexes", *International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG)*, Plzen, Czech Republic, p. 25-28, February, 2008.
- [13] Lian, Zhouhui and Rosin, Paul L. and Sun, Xianfang "A rectilinearity measurement for 3d meshes". In: *Proceeding of the 1st ACM international conference on Multimedia information retrieval*, October 30 - 31, 2008.
- [14] P. L. Rosin. "Computing global shape measures". *Handbook of Pattern Recognition and Computer Vision*, 3rd edition, page 177:196, 2005.
- [15] Network of excellence AIM@SHAPE, <http://www.aimatshape.net> (2004).
- [16] P. Shilane, P.Min, M. Kazhdan, et T. Funkhouser. "The princeton shape benchmark". In *Shape Modelling International*, June 2004.
- [17] Vranic D. and Saupe D., "3D Model Retrieval with Spherical Harmonics and Moments," *DAGM 2001*, B. Radig and S. Florczyk, Eds., Munich, Germany, pp. 392-397, 2001.

- [18] B. Bustos, D. Keim, D. Saupe, T. Schreck, "Content-based 3d object retrieval". IEEE Computer Graphics and Applications, 27(4): p. 22-27. 2007
- [19] OWL web ontology language guide. <http://www.w3.org/tr/2004/rec-owl-guide-20040210/>, w3C Recommendation (Feb 2004).
- [20] Prud'hommeaux, E., Seaborne, A.: Sparql query language for rdf. W3C Working Draft <http://www.w3.org/TR/rdf-sparql-query/> (2006)
- [21] <http://pellet.owldl.com>
- [22] Boder, M.: "Mpeg-7 visual shape descriptors". IEEE Transactions on Circuits and Systems For Video Technology 716-71911 (2001)
- [23] N. Marwan, P. Saporin, J. Kurths, W. Gowin: "3D measures of complexity for the assessment of complex trabecular bone structures, Proceedings of the International meeting "Complexity in the living: a problem-oriented approach", Rome, 2004Rapporti ISTISAN, 05/20, 53-58 (2005).
- [24] D. A. Randell, Z. Cui et A. G. Cohn. "A Spatial Logic based on Regions and Connection". In : 3rd International Conference on Principles of Knowledge Representation and Reasoning. pp. 165-176. Morgan Kaufmann Publishers, 1992.
- [25] Cohn AG, Hazarika SM "Qualitative Spatial Representation and Reasoning: An Overview". Fundamenta In-formaticae 46(1-2):1629. 2001
- [26] AKSOY, S., HARALICK, R., CHEIKH, F., AND GABBOUJ, M. 2000. "A weighted distance approach to relevance feedback". In International Conference on Pattern Recognition, 4812-4815. 2000
- [27] Corney JC, Rea HJ, Clark DER, Pritchard J, MacLeod RA, Breaks ML. "Coarse filters for shape matching". IEEE Computer Graphics and Applications;22(3):65-74. 2002
- [28] C. Zhang and T. Chen "Indexing and retrieval of 3 D models aided by active learning", Proc. Int. Multimedia Conf., vol. 9, p.615 , 2001

**My abdellah Kassimi** is a PhD student at Sidi Med Ben AbdEllah University (GRMS21 group) in Morocco. He received his DESS in Computer Science from the University of Sidi Md Ben AbdEllah in 2007. His current research interests are 3D indexing and retrieval, 3D shape indexes, semantic and ontology.

**Omar El Beqqali** is currently Professor at Sidi Med Ben AbdEllah University. He is holding a Master in Computer Sciences and a PhD respectively from INSA-Lyon and Claude Bernard University in France. He is leading the 'GRMS21' research group since 2005 (Information Systems engineering and modeling) of USMBA and the Research-Training PhD Unit 'SM31'. His main interests include Supply Chain field, distributed databases and Pervasive information Systems. He also participated to MED-IST project meetings. O. El Beqqali was visiting professor at UCB-Lyon1 University, INSA-Lyon, Lyon2 University and UIC (University of Illinois of Chicago). He is also an editorial board member of the International Journal of Product Lifecycle Management (IJPLM).



# A Thought Structure for Complex Systems Modeling Based on Modern Cognitive Perspectives

Kamal Mirzaie<sup>1</sup>, Mehdi N. Fesharaki<sup>2</sup> and Amir Daneshgar<sup>3</sup>

<sup>1</sup> Department of Computer Engineering, Science and Research Branch, Islamic Azad University  
Tehran, Iran

<sup>2</sup> Department of Computer Engineering, Science and Research Branch, Islamic Azad University  
Tehran, Iran

<sup>3</sup> Department of Mathematical Sciences, Sharif University of Technology  
Tehran, Iran

## Abstract

One of the important challenges for complex systems modeling is finding an appropriate thought structure for designing and implementing a suitable simulation software. In this paper, we have proposed a suitable worldview for complex systems modeling according to Capra's conceptual framework, which is based on modern cognitive theories. With this worldview, the important and fundamental concepts for complex systems modeling are determined. Adding more details to the model that depends on the field of problem, we can simulate a complex system. Also using Popper's Three Worlds, the position of this simulation has been described. Following this thought structure, each simulation designer of complex systems can take advantage of modern cognitive theories in modeling.

**Keywords:** *Thought Structure, Complex Systems, Cognitive, Agent Based Modeling (ABM).*

## 1. Introduction

Appropriate modeling of complex systems is one of the fields of research today [1][2][3][4][5][6][7]. Researchers in this field are trying to extract appropriate concepts, provide frameworks and computational methods and mechanisms in order to create simulation models to describe the behavior of complex systems [1][2][3][4]. A complex system is made of interconnected components and as a result of the interactions between these components, the emergent behavior would appear [3][8][9]. Although we may describe the interactions among components with simple rules, but as the number of system components rises, the number of interactions between components will increase too.

Living systems such as cells, organizations, society and the earth in which there is the concept of life are all examples of complex systems [10]. These systems can be biological or social [10][11][12]. In living systems as complex systems, there are interactions between components for survival and evolution. One common approach in modeling living systems is Complex Adaptive Systems theory (CAS) [3][4]. In this theory, the living system is a complex system that adapts to its surrounding environment throughout its life for survival and evolution. Adaptation means how a system responds to the changing environment and adapts to it [1][2][3][4].

The modeling of complex systems usually leads to a simulation software, with which researchers can simulate and test their models and theories [8][12][14][15]. In addition, simulation software is a suitable alternative and in some cases the only possible way to test the models and theories [8][15]. Simulation approach both reduces costs and also enables researchers to study their models and theories with various parameters, aspects, and iterations [11][17]. A model is the foundation of simulation software which describes the main concepts, components, and processes as formal relationships [14][16][17]. The closer a model to reality, the better it will be. However, good modeling does not necessarily include more details, rather it means choosing and including features, components, and concepts that has a greater influence on reality [8][14].

A common terminology in complex systems modeling and simulation is agent [8][11][18]. Agent is an entity that can represent a cell, a human, or any living organisms in a complex systems modeling. Modeling based on the agent concept leads to Agent Based Modeling (ABM) [11]. With the advent of CAS theory and its wide applications,

researchers found out that models use CAS and MAS (Multi-Agent System) to model nonlinear dynamic interactions that have been missing in the previous linear models [8]. However, it is suitable to utilize a thought structure that makes modeling and simulation of complex systems more accurate and produces a high quality software simulation.

In this paper, first the necessity of a suitable complex systems modeling worldview is explained and then it is illustrated by Capra's conceptual framework. Then a thought structure for complex systems modeling with regard to Popper's Three Worlds is proposed. The first world is about complex systems worldview, the second world is about individual and social awareness and finally the third world is an artifact that is a methodology for simulator development.

## 2. Complex Systems Modeling Worldview

In complex systems, global behavior emerges from high number of interactions between components [3][4]. As the number of interactions is very high, the emergent behavior appears. Therefore, for understanding and modeling of complex systems, a special worldview is required. This worldview is the base of some methodologies such as CommonKADS and it precedes theory [19].

Overall, the methods that have been used for systems modeling during the past decades can be divided into two main approaches:

- 1) Model-oriented approach: It is based on methods of traditional system thinking. Worldview of this approach is based on reductionism. Reductionism is breaking a problem into smaller ones, solving each one separately and then combining the answers to get the solution of the main problem. In other words, for understanding the main system, we divide it into sub-systems and they can be further divided into smaller systems until we get to the systems that are knowable.
- 2) Data-oriented approach: The main idea of this approach is that complex systems cannot be understood with reductionism worldview. Therefore, as the behavior of the system is from bottom to top, for understanding it we need a new holistic worldview. In this worldview, emergent behavior becomes meaningful. It is according to this worldview that complex systems theories, cognitive theories, and other theories based on the new holistic thinking are used in complex systems modeling.

### 2.1 Capra's Conceptual Framework as a Worldview

Capra's conceptual framework is based on new interpretations and definitions of cognition. Cognition is the process of knowing in life; knowing how and what capabilities are used for survival. With this definition, the smallest living organisms, such as cells are cognitive phenomena using cognition for survival in life. Defining cognition based on biological view enables us to use the cognitive concepts in a wide range to explain the behavior of living organisms. We can find network patterns everywhere, from the smallest cognitive living organisms such as cells to organizations and human societies. Thus, network is a common pattern for life [10].

One of the cognitive theories based on the biological view is Santiago theory [10]. According to this theory, cognition is synonymous with the process of life. The organizing activities of living systems at all levels of life are cognitive. These activities include interactions among living organisms such as plants, animals, or human beings and their environment. Thus life and cognition are inseparable, as though mental activity is immanent in matter. Santiago's cognitive theory expands the cognitive concept in a way that it involves the entire process of life including perception, emotion, and behavior. In this theory, cognition is not just for human beings with a brain and a nervous system, rather it can be for each living organism, from cells to social organizations [10].

Capra has presented a unique framework for understanding the biological and social phenomena in four perspectives. Three out of these four perspectives is about life and the fourth one is meaning. (Fig. 1)

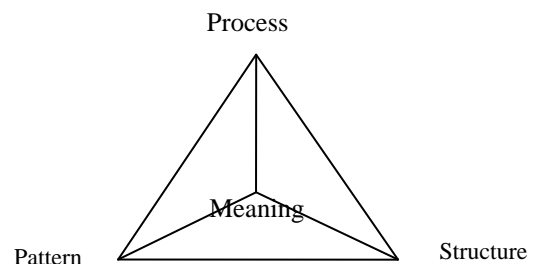


Fig. 1 Four perspectives of Capra's Conceptual Framework.

The first perspective of Capra's conceptual framework is pattern that includes various relations among system components. The organization pattern of a living system defines the relation types among the system components which determines the basic features of the system. Structure, the second perspective, is defined as the



material embodiment of system pattern. The Structure of a living organism evolves in interaction with its environment. The third perspective is life process which integrates the pattern and the structure perspectives. For example, the study of living systems from these three perspectives includes the study of form (patterns of organizations), matter (or material structure), and process. From the perspective of form, the pattern of organization is a self-generating network. From the perspective of matter, the material structure of a living system is a dissipative one, that is, an open system that operates far from equilibrium. And from the process perspective, living systems are cognitive systems in which the process of cognition is closely linked to self-generating network [10].

When we try to extend new understanding of cognition to the social life, we immediately encounter many misleading phenomena - rules of behavior, values, goals, strategies, intentions, designs and power relations - that often do not have a role in non-human world, but they are essential for human social life. For expanding life to the social domain, meaning perspective is added to three other ones. Thus, we can understand social phenomena from four perspectives: pattern, structure, process, and meaning. Culture, for instance, has created and preserved a network (pattern) of communication (process) with embedded meaning. Material embodiment of culture includes art and literary masterpieces (structure) that transfer meaning from one generation to another.

As there is the concept of life and evolution in the living systems such as cells, organizations, and societies, there are all examples of a complex system. So, Capra's conceptual framework can be used as a worldview to understand complex systems.

## 2.2 Complex Systems Modeling in Capra's Conceptual Framework

According to Capra's conceptual framework, any complex phenomena can be discussed and studied in four perspectives. In order to close these four perspectives to the terminology of complex systems modeling, we replace "pattern" with "network" and "structure" with "agent".

Pattern perspective is the relationship between components, thus network is a good terminology. Structure is a set of features that evolves during life. These features together make the agent concept. Therefore, Capra's conceptual framework is redefined in four perspectives: network, agent, process, and meaning (Fig. 2).

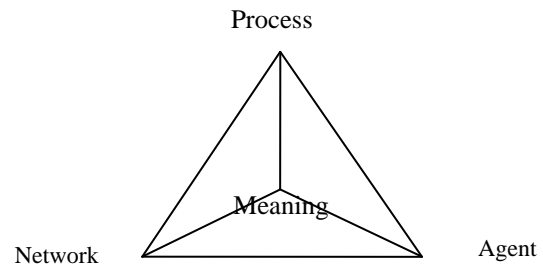


Fig. 2 Redefinition of Capra's Conceptual Framework for Complex System Modeling

## 3. Extending the Thought Structure Using Popper's Three Worlds

According to Popper's Three Worlds, the first world is the physical world which is related to the worldview. The second world is the subjective realm in which theories and concepts are formed. And the third world is the objective one which is the realm of artifacts and objective knowledge [20].

In the previous section, the worldview of complex systems modeling in Capra's conceptual framework has been described. Now, we define individual and social awareness as the second world in Popper's Three Worlds. According to Fig. 3, individual and social awareness are both affected and affect agent, network, and process in the first world. Individual awareness refers to what knowledge each agent has and what it has learned from its environment and also from other agents. In other words, individual awareness is a memory that every agent has from its surrounding environment and this memory evolves during the life of the complex system. Hence, individual awareness is a mental model and every agent makes decision based on situation awareness. In a complex systems modeling, a set of agents are related to each other in order to achieve certain goals; therefore, in addition to individual awareness, social awareness is formed. Social awareness is the knowledge that a set of agents create together. It is a collective memory that is created by agents interacting with each other. The collective memory is a shared mental model that appears as shared situation awareness and can be used for coordination in social environments [21].

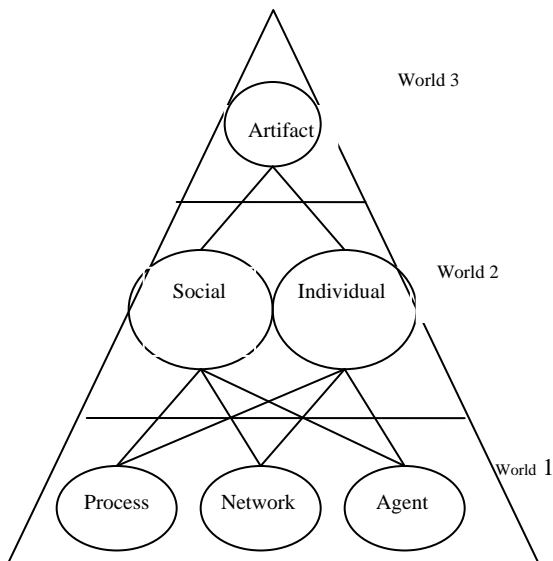


Fig. 3 General Thought Structure for Complex System Modeling Using Popper's Three Worlds

But, how individual and social awareness are created in a complex system? Although it is difficult to answer this question clearly, agent, network, and process influence the creation of individual and social awareness (Fig. 3). To observe this influence in the formation of awareness, modeling and simulation are suitable approaches. In other words, the first world that views the complex system from the perspectives of agent, network, and process can be developed and examined as a simulation software.

### 3.1 Layers of Simulation Software

Before we develop a simulator software, an architectural design is required which is based on some theories. In other words, the simulator development is based on the theories that explain a given phenomenon. Overall, design and implementation of the simulation software can be described in three layers (Fig. 4):

- 1 - Theoretical basics
- 2 - Software architecture
- 3 - Computational models

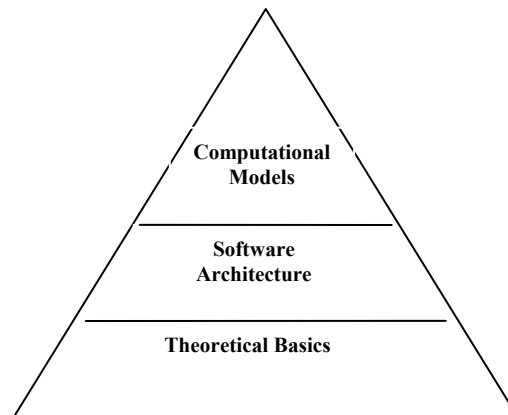


Fig. 4 Three Layers of Simulation for Complex Systems.

**Theoretical Basics Layer:** Theories refer to the philosophy of problem-solving method of simulation software. Therefore, theoretical basics are the base of simulation pyramid (Fig. 4). In simulation software design and implementation, for example, complex systems theory, graph theory, queue theory, and network theory can be used. That which theories should be used in the design and implementation is determined by answering this question: What theories justify modeling and simulation for a simulation developer? In other words, which theories are consistent with a given simulation problem?

Given to the adopted worldview, and after examining the given problem, we select appropriate theories for modeling and simulation in order to have theoretical basics.

**Software Architecture Layer:** Having determined the theoretical basics, software architecture is defined. It is based on theories and consists of software components, and relationships between them. Software architecture is an overall design that defines building blocks, relationships between blocks, and entities within each block [22]. There are two main approaches for modeling entities: object-oriented and agent-oriented. As the agent concept has more capabilities than the object concept, it models the entity better and is often used in the simulation software development. This approach is called Agent Based Modeling (ABM).

**Computational Models Layer:** Computational models express software components and relationships between them in the architecture layer in the form of mathematical and computational relations. In other words, computational models provide computational mechanisms that a software developer uses to generate executable

codes. Therefore, this layer is the provider of a formal language for simulator software. Computational models are chosen based on theoretical basics and software architecture (Fig 4). For example, we can use soft computing such as genetic algorithms, neural networks, and fuzzy computations in this layer.

### 3.2 Development Methodology as an Artifact

Two worlds out of Popper's Three Worlds for complex systems modeling have been described so far. The first world is the complex systems modeling worldview that we redefined in three perspectives of agent, network, and process, based on Capra's conceptual framework. The second world is individual and social awareness of agents that is essential for their coordination.

We call the third world of Popper's Three Worlds artifact (Fig. 3). This world is objective knowledge and is falsifiable, that is, it is true as long as we cannot prove its falseness. Artifact is a methodology in our proposed thought structure. This methodology determines what concepts, components, and methods should be used for complex systems modeling. In other words, it illustrates and confirms the effect of meaning perspective (the fourth perspective of Capra's Conceptual Framework) in the form of some general principles. In a way, meaning is the interpretation of simulation results. We can interpret the results of simulation according to a given meaning. Overall, this methodology determines general principles for software architecture. For example, what principles and structures should be used for network design? What features are more important for agent design and definition? What kinds of processes are suitable for modeling a given complex system?

The principles obtained from the results of modeling and simulation can be used in the design of products and real applications. That is, these principles are used in the design of agent, network, and process in order to create a given meaning. They can be reviewed and revised after being used in real applications.

## 4. Conclusions

Complex systems modeling is one of the challenges and necessities of today's researchers which demands a suitable thought structure. Many researchers consider a living system as a complex system that adapts to its surrounding environment for survival and evolution. Consequently, cognitive theories and thought frameworks suggested for describing living systems can be utilized for understanding complex systems. Capra's Conceptual

Framework is based on modern cognitive theories; therefore, we have used its modified version as the proposed thought structure worldview. This thought structure is based on Popper's Three Worlds. The first world is the complex systems modeling worldview that we have redefined in three perspectives of agent, network, and process. The second world is individual and social awareness that concerns with individual and shared situation awareness. The third world is an artifact that explains methodology for complex systems modeling. In other words, the artifact determines general principles and approaches for the software architecture.

## References

- [1] C. Gros, *Complex and Adaptive Dynamical Systems: A Primer*, Springer-Verlag Berlin Heidelberg, 2008.
- [2] A. Yang, and Y. Shan, *Intelligent Complex Adaptive Systems*, IGI Publishing, 2008.
- [3] J. H. Miller, and S. E. Page, *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*, Princeton University Press, 2007.
- [4] J. Clymer, *Simulation Based Engineering of Complex Systems*, Wiley-Interscience, 2009.
- [5] C. F. Kurtz and D. J. Snowden, "The New Dynamics of Strategy: Sense-making in a Complex and Complicated World", *IBM Systems Journal*, Vol. 42, No. 3, 2003, pp. 462-483.
- [6] C. A. Aumann, "A Methodology for Developing Simulation Models of Complex Systems", *Ecological Modelling*, Vol.202, No. 3-4, 2007, pp. 385-396.
- [7] M. A. Janssen, and W. J. M. Martens, "Modeling Malaria as a Complex Adaptive System", *Artificial Life*, Vol. 3, No. 3, 1997, pp. 213-236.
- [8] A. Yang, "A Networked Multi-Agent Combat Model: Emergence Explained", Ph.D. thesis, University of New South Wales, Australian Defence Force Academy, 2006.
- [9] C. Joslyn, and L. Rocha, "Towards Semiotic Agent-Based Models of Socio-Technical Organizations", *AI, Simulation and Planning in High Autonomy Systems (AIS 2000) Conference*, Tucson, Arizona, 2000, pp. 70-79.
- [10] F. Capra, *The Hidden Connections: Integrating the Biological, Cognitive, And Social Dimensions of Life Into A Science of Sustainability*, Doubleday, 2002.
- [11] N. Gilbert, and K. G. Troitzsch, *Simulation for the Social Scientist*, Open University Press, McGraw-Hill Education, Second Edition, 2005.
- [12] N. Cannata, F. Corradini, E. Merelli, A. Omicini, and A. Ricci, "An Agent-oriented Conceptual Framework for Biological Systems Simulation", *Transaction on Computation System Biology* Vol. 3, 2005, pp.105-122.
- [13] A. Ilachinski, *Artificial War: Multiagent-Based Simulation of Combat*, Singapore, World Scientific Publishing Company, 2004.
- [14] M.A. Niazi, and A. Hussain, "A Novel Agent-Based Simulation Framework for Sensing in Complex Adaptive Environments", *IEEE Sensors Journal*, Vol. 11, No.2, 2010, p.p. 404-412.

- [15] K. Sprague, and P. Dobias, "Behaviour in Simulated Combat: Adaptation and Response to Complex Systems Factors", Defence R&D Canada, Centre for Operational Research and Analysis, DRDC CORA TM 2008-044, November 2008.
- [16] M. A. Niazi, and A. Hussain, "Agent based Tools for Modeling and Simulation of Self-Organization in Peer-to-Peer, Ad-Hoc and other Complex Networks", IEEE Communications Magazine, Vol.47 No.3, 2009, pp. 163–173.
- [17] A. M. Law, and M. G. McComas, "How to Build Valid and Credible Simulation Models", Winter Simulation Conference Miami, FL, 2001, pp. 22-29.
- [18] R. Allan, "Survey of Agent Based Modelling and Simulation Tools", Computational Science and Engineering Department, STFC Daresbury Laboratory, Warrington WA4 4AD, June 3, 2009.
- [19] G. Schreiber, H. Akkermans, A. Anjewierden, R. Hoog, N. Shadbolt, W. V. Velde, and B. Wielinga, Knowledge Engineering and Management: The CommonKADS Methodology, MIT Press, 2000.
- [20] K. R. Popper, The Logic of Scientific Discovery, New York, NY: Routledge, 1992.
- [21] A. Fetanat, and M. F. Naghian, "A Trust Model in Sensemaking Process", International Journal of Computational Cognition, Vol.8, No.2, 2010, pp. 1-3.
- [22] R. S. Pressman, Software Engineering: A Practitioner's Approach, Seventh Edition, McGraw-Hill, 2010.

**Kamal Mirzaie** is a PhD student in Computer Engineering at Science and Research Branch, Islamic Azad University. He has MSc in Computer Engineering from Isfahan University and BS in Computer Engineering from Iran University of Science and Technology (IUST). His research interest is focused on complex system modeling based on cognitive theories.

**Mehdi N. Fesharaki** is an Associate Professor in the Computer Engineering Department, Science and Research Branch, Islamic Azad University. He has BS and MSc in electrical engineering from Sharif University of Technology, and a PhD in Computer Engineering from NSW University of Australia.

**Amir Daneshgar** is a Professor in the Department of Mathematical Sciences, Sharif University of Technology.

# Identification of Priestley-Taylor transpiration Parameters used in TSEB model by Genetic Algorithm

Abdelhaq Moudia<sup>1</sup> and Nouredine Alaa<sup>2</sup>

<sup>1</sup> Direction Régionale Météorologique Nord, Marocmeteo, DMN  
Rabat, Morocco

<sup>2</sup> Department of Applied Mathematics and Informatics, University of Cadi Ayad, Faculty of Science and Techniques  
Marrakech, Morocco

## Abstract

The accuracy degree of extracted canopy latent heat from canopy net radiation is depending extremely to the proposed Priestley-Taylor approximation. This extracting canopy latent heat is an initial approximation to compute iteratively partitioned energy components to soil and vegetation using in Two Source Energy Balance (TSEB) Model. This approximation is using a Priestley-Taylor coefficient ( $\alpha_p$ ) and fractional of Leaf Area Index ( $f_g$ ) that is green. The standard values are 1.26 and 1 for respectively ( $\alpha_p$ ) and ( $f_g$ ). This study is focused to identify these two transpiration parameters ( $\alpha_p$ ) and ( $f_g$ ) by Genetic Algorithm method to accurately predict patterns of turbulent energy fluxes by TSEB Model (Norman et al. 1995), over irrigated olive orchard in semi-arid area (Marrakech, Morocco). The ( $\alpha_p$ ) and ( $f_g$ ) are depending on local climatic characteristics and data measurements accuracy for different periods of the year 2003. In summer 2003, the GA gives optimal values for ( $\alpha_p=0.93$ ) and ( $f_g=0.61$ ). Ten runs of GA computing have been applied to guaranty stability of the optimization process. In fact, the simulation of latent heat becomes improved as presented as below, since comparison to ground measurements shows acceptable representativeness in summer 2003 with enhancement of TSEB Model performance assuming correlation to (0.45), bias is to (+15 W.m<sup>-2</sup>), and the root mean square have been improved to (63 W.m<sup>-2</sup>). Thus, the results obtained here show the most important support of Genetic Algorithm through the calibration and optimization processes.

**Keywords:** Genetic algorithm, Optimization, Fitness function, Cost function, TSEB Model.

## 1. Introduction

Many methods have been used to estimate canopy evapotranspiration from regions using standard climate data. Priestley-Taylor approximation suggest one of these based on physical argument about processes in the whole of turbulent planetary boundary layer, and their arguments

were concerned the relative sizes of advective and radiant energy inputs to land areas of local size (Priestley-Taylor, 1972; McNaughton et al 1991).

They were forced to proceed empirically, and asked whether it was still a principal component of evaporation from a wet region. They looked that a value of coefficient ( $\alpha_p=1.26$ ) was found to fit data from several sources especially for wet regions. The TSEB Model uses either this formula adding another coefficient ( $f_g=1$ ) which is a fractional of Leaf Area Index that is green (Norman et al, 1995; Kustas et al 1999). Several studies are also proposed values of ( $\alpha_p$ ) and ( $f_g$ ) ranging respectively from 0.5 to 3 and 0 up to 1 (Castellvi et al, 2001; Kustas et Norman et al, 1999a, Agam et al. 2010). In this study, for a semi-arid areas, we suggest to use stochastic method as Genetic algorithms (GAs) to identify Priestley-Taylor transpiration Parameters over olive irrigated area (in wet and dry conditions). GAs approach are used for solving parameters estimation for its independency to problem types, such as non linear, multimodal and/or non-differentiable functions (Holland, J. H, 1975; Goldberg, David E, 1989). GAs are a way of addressing hard search and optimization problems which provides a good solution although it requires large execution time.

In section 2 we present study area and data collection, while section 3 describe the

Priestley-Taylor approximation of transpiration used in TSEB Model. The section 4 highlights GAs theoretical bases and implementation. In section 5 we show results but conclusion is presented in section 6.



## 2. Study area and data collection

### 2.1 Site description

The study site was located in the 275 hectare Agdal olive (*Olea europaea* L.) orchard in the southern side of Marrakech City, Morocco (31,601 N; 07,974 W). It is characterized by low and irregular rainfall (annual average of about 240 mm, but 263.4 mm has been collected in 2003). The climate is typically Mediterranean semi arid; precipitation falls mainly during winter and spring, from November to April. The atmosphere is very dry with an average humidity of 56% and the evaporative demand is very high (1600mm per year), greatly exceeding the annual rainfall. The orchard was periodically surface irrigated through level basin flood irrigation, with water supplies of about 100 mm every each irrigation event. We have approximately 3 irrigation events during summer 2003. Each tree was occupied over 45 m<sup>2</sup>, and bordered by small earthen levy (about 30 cm) retained irrigation water (Williams et al, 2004). Plant spacing was about (6.5x6.5 m); the trees had an average leaf area index (LAI) of 3. Mean tree height was 6 m and ground cover was 55% (Ezzahar, 2007).

### 2.2 Measurements

Measurements were acquired at a sampling frequency of 20 Hz and passed through a low-pass filter to compute 30-min flux averages. Intensive data were collected in Agdal site. Vertical fluxes of heat and water vapor at 9.2 m height were registered on twelve month of 2003 and are measured by an Eddy-Covariance (EC) system (Ezzahar et al, 2007). Finally, the resulting dataset of sensible and latent heat fluxes were available for the 2003 growing seasons, with missing data for few days due to power supply troubles. Almost 6247 hourly observations, during daytime, everyday along the year 2003 without any exclusion related to season or climatic conditions, were used to run and evaluate TSEB model output.

A 3D sonic anemometer (CSAT3, Campbell Scientific, Logan, UT) measured the fluctuations in the wind velocity components and temperature. An open-path infrared gas analyzer (LI7500, LiCor, Inc., Lincoln, NE) measured concentrations of water vapour. The wind speed and concentration measurements were made at 20 Hz on CR23X dataloggers (Campbell Scientific, Logan, UT) and on-site portable computers to enable the storage of large raw data files. Air temperature and humidity were measured at 8.8 and 3.7 m heights on the tower with Vaisala HMP45C probes. Total shortwave irradiance was measured at 9.25 m height with a BF2 Delta T radiometer. Net radiation was measured with a Kipp and Zonen CNR1 net radiometer placed over the olive canopy at 8 m height.

Soil temperature was recorded at 5 cm depth at two locations approximately 30 m from the tower. Three heat flux plates continuously monitored changes in soil heat storage at the tower site. In addition, five point measurements of soil moisture variables were located throughout the site. Each point contained a pair of steel rods for time domain reflectometry (TDR) measurements at 40, 30, 20, 10 and 5 cm depths to estimate volumetric water content. Olive transpiration was measured by sap flow method following the procedure of Williams et al., 2003. Soil evaporation was computed as the difference between evapotranspiration measured by eddy correlation system and transpiration measured by sap flow method.

## 3. Priestley-Taylor transpiration in TSEB Model

The Priestley-Taylor equation is only an initial approximation of canopy latent heat simulated by TSEB Model. TSEB is based on energy balance closure using surface radiometric temperature, vegetation parameters and climatic data. TSEB outputs surface turbulent fluxes, and temperatures of canopy and soil. The version implemented in this study basically follows what is described in appendix A as the “parallel resistance network”. As such, the model implemented is described in detail in (Norman et al. 1995, Kustas and Norman 1999). The canopy latent heat  $LE_c$  is given by Priestly-Taylor approximation (Priestly-Taylor. 1972).

$$LE_c = R_{nc} \alpha_p f_g \frac{\Delta}{\Delta + \Gamma} \quad (1)$$

where  $\alpha_p$  is the Priestly-Taylor constant, which is initially set to 1.26 (Priestley-Taylor, 1972; Norman et al 1995; Agam et al 2010),  $f_g$  is the fraction of the LAI that is green,  $\Delta$  is the slope of saturation vapour pressure versus temperature curve,  $\Gamma$  is the psychrometer constant (e.g: 0.066 kPa C<sup>-1</sup>). If no information is available on  $f_g$ , then it is assumed to be near unity.

## 4. Genetic algorithms method

### 4.1 Overview

Genetic Algorithms (GAs) are an optimization algorithms based on techniques derived from the genetic and the Darwin's theory of evolution in selection, crossover, mutation, generation, parent, children, etc (Goldberg 1989; Holland 1975). As a considerable development in the computing systems, GAs has shown a significant



improvement by using stochastic and mathematic methods which has been applied into many domains such as ecologies, biology and even economy, in order to experiment it for understanding natural systems, and modelling it to optimize (or at least improve) the performance of the system.

#### 4.2 GAs theoretical bases and implementation

Genetic algorithms have been used to solve difficult problems with objective functions that do not possess some properties such as continuity, differentiability, satisfaction of the Lipschitz Condition, etc (Michalewicz 1994; Goldberg 1989; Holland 1975).

GAs search extremum of function defined in space data. These algorithms maintain and manipulate a family, or population, of solutions and implement a “survival of fittest” strategy in their search for better solutions. GAs have shown their advantages in dealing with the highly non-linear search spaces that result from noisy and multimodal functions.

The genetic algorithm works as follows:

- Initialization of parent population randomly
- Evaluation (fitness function)
- Selection
- Recombination of possible solutions (Crossover and Mutation)
- Evaluate child and go to step 3 until termination criteria satisfies.

##### 4.2.1 Solution representation

The chromosome (individual) chosen to represent a solution is a vector coded of floating number representing

$$K = \langle \alpha p, fg \rangle \quad (2)$$

The ranges of a parameters are a and b. The  $\alpha p$  is the Priestly-Taylor constant, and fg is the fraction of the LAI that is green. The real-valued representation moves the problem closer to the problem representation which offers higher precision with more consistent results across replications (Michalewicz 1992).

##### 4.2.2 Initialization, Termination and Evaluation

The most common method providing an initial population is to randomly generate solutions for the entire population such as:

$$K = a + (b - a) * \text{rand}(2, N) \quad (3)$$

where N is the dimension of population, such that each element of array contains a possible value of parameters; and  $\text{rand}(2, N)$  returns a pseudorandom vector value are drawn from a uniform distribution on the unit interval.

The GA moves from generation to generation selecting and reproducing parents until a termination criterion is met. The most frequently used stopping criterion is a specified maximum number of generations.

Fitness in biological sense is a quality value which is a measure of the reproductive efficiency of chromosomes (Goldberg, 1989). In genetic algorithm, individuals are evaluated with it fitness function which is a measure of goodness to be selected.

The evaluation is calculated at each TSEB run through the fitness function  $\Phi(K)$  which is equal to

$$\Phi(K) = \left[ \frac{1}{1 + \frac{1}{2} \int_0^T [LE_{sim}(t, K) - LE_{obs}(t)]^2} \right] \quad (4)$$

where (t) is the instant of observed latent heat  $LE_{obs}(t)$  and  $LE_{sim}(t, K)$  is the simulated latent heat.

The cost function to minimize is represented by a practical evaluation of  $\mathfrak{Z}(K)$  where

$$\mathfrak{Z}(K) = \frac{1}{2} \int_0^T [LE_{sim}(t, K) - LE_{obs}(t)]^2 \quad (5)$$

where T is the time period.

##### 4.2.3 Genetic Operators

Genetic algorithm uses some operators to create children forming next new generation by parents selected from the current population. The algorithm usually selects a group of individuals that have better fitness values as parents.

The genetic operators are as follows:

- Selection: Reproduction (or selection) is usually the first operation applied on a population to breed a new generation. Individual solutions are selected through probability that individual  $(K_i)_{1 \leq i \leq N}$  is selected from the  $i$ th line of matrix, to be a member of the next generation at each experiment is given by

$$\text{Prob}(K_i \text{ is selected}) = \frac{\Phi(K_i)}{\sum_{i=1}^N \Phi(K_i)} \quad (6)$$

The process is also called roulette wheel parent selection. This selection step is then a spin of the wheel, which in the long run tends to eliminate the least fit population

members. The population will be represented by a slice that is directly proportional to the member's fitness.

- **Crossover:** A crossover operator is used to recombine pairs of parents to get better children which generate a second generation of solutions. In the case of individual probability is less than 0.5, the son child chromosome will be an average of two times value of father with one value of mother, and vice versa for the daughter child, but if individual probability is great or equal to 0.5, the son and daughter chromosome will stay respectively like father and mother.

- **Mutation:** Mutation is an operator that introduces diversity in the population to avoid homogeneous generation due to repeated use of reproduction and crossover operators. Mutation proceeds to Gaussian perturbation with deviation equal to 0.5 and probability mutation equal to 0.0001. Mutation adds simply new information in a random way to the genetic search process.

#### 4.2.4 Implementation of GAs to TSEB Model

Possible solutions to a problem are evaluated and ordered according to its adaptation (i.e: fitness function). From generation (k) to new one (k+1), then other chromosome populations are produced after selecting candidates as 'parents' and applying mutation or crossover operators which combine chromosome of two parents to produce two children. The new set of candidates is then evaluated, and this cycle continues until an adequate solution is found (figure.1). In all experiments, GA experimental parameters are as follows: the population size is 10, the crossover rate is 0.5, the mutation rate is 0.0001 and we generate population until the 10th generation. The observations used in TSEB Model are taken each 30 minutes. In this optimization we want to minimize the cost function, then we proceed the minimization to find a vector  $K_{opt}$  as follows:

$$\mathfrak{S}(K_{opt}) = \inf \mathfrak{S}(K) \quad (7)$$

where  $K = \langle \alpha, \beta, \gamma \rangle$  is the vector of parameters to be controlled, and  $\mathfrak{S}(K)$  is the cost function.

The state variable is the simulated latent heat  $LE_{sim}(t, K)$  evolving in the time during summer 2003 between DOY=152 to DOY=243. The cost function is computed by comparing simulated  $LE_{sim}$  and observed latent heat  $LE_{obs}$  during the all period T. The two unknown parameters controlling the Priestley-Taylor transpiration used in TSEB Model are estimated by optimization of the cost function with the evolution strategies algorithm as follow:

-**START:** Create random population of 10 chromosomes  $K = \langle \alpha, \beta, \gamma \rangle$  between 0.5 to 2 for  $\alpha$ , and 0.1 to 1 for  $\beta, \gamma$ ,

-**Run TSEB:** Calculate the simulated latent heat  $LE_{sim}(t, K)$ , the bias to measured latent heat  $LE_{obs}(t)$  and the function cost  $\mathfrak{S}(K)$ ,

-**FITNESS:** Evaluate the fitness function  $\Phi(K)$  of each chromosome in the population,

-**NEW POPULATION:**

\* **SELECTION** : Based on  $\Phi(K)$

\* **RECOMBINATION:** Cross-over chromosomes

\* **MUTATION** : Mutate chromosomes

\* **ACCEPTATION** : Reject or accept new one

-**REPLACE** : Replace old with new population as the new generation

-**TEST** : Test problem criterion to indicate the best solution  $K = \langle \alpha, \beta, \gamma \rangle$  minimizing the cost function  $\mathfrak{S}(K)$ , else to turn over to the next generation

**LOOP** : Continue step 2– 6 until criterion is satisfied.

## 5. Results

Different number of generations (not shown) with ten individuals population have been experimented in order to

optimize values of  $K_{opt} = \langle \alpha, \beta, \gamma \rangle$  and to carry out stability test to GA with showing performance to Priestley-Taylor formulation. The founded parameters by GA are changing with reproduction in generations. The GA start generally with a randomly values of parameters

in the beginning of minimized cost function  $[\mathfrak{S}(K)]$ , but in the absence of stopping criterion to the most minimizing cost function, the GA change choice to selected individuals who decrease Latent heat error to reach its minimum. The GA continues to generate elite chromosomes for computing predicted surface fluxes until stability of Latent heat error (fig.2). The stability error phase is characterized by a little changing in reproductive individual's adaptation. The convergence will be reached during generation when the best individual is founded to the medium one (fig.1). The estimation of Priestley-Taylor formulation has been improved then the TSEB Model performance will come acceptable with best parameters

giving by 10 generations. We proceed in the following to experiment 10 runs of GA to show best parameters changing and test stability reproduction procedure with 10 individuals' population and 10 generations. During error stabilization error process, the 10 runs of GA shows (table.1) changing in parameters value, since  $\alpha_p$  is ranging between 0.72 to 1.00 and  $f_g$  vary from 0.26 to 0.79. These optimized values for  $\alpha_p$  and  $f_g$  are less than the standard value ( $\alpha_p=1.26$  and  $f_g=1$  for wet conditions), then we can considered them for semi arid area. Optimized values for  $f_g$  are conforming to irrigated area explaining conditions supporting soil and canopy transpiration. GA gives sometimes optimal parameters corresponding to minimum error before reaching its stabilization, but GA continue computing process since there is no stopping criterion for this case to reduce calculation time. The mean parameters value optimized in 10 previous runs of  $\alpha_p$  and  $f_g$  (table.1) are respectively 0.93 and 0.61. Now let us see the influence of these optimal mean values to TSEB Model. Figures 3 and 4 present the comparison of measured and predicted daily latent heat before and after optimization process. These figures show an improvement of latent heat representativeness. The correlation becomes from (0.43) to (0.45), the bias is reduced from (+240 W.m<sup>-2</sup>) to (+15 W.m<sup>-2</sup>), and the root mean square have been improved from (251 W.m<sup>-2</sup>) to (63 W.m<sup>-2</sup>). Furthermore the measured and predicted latent heat evolve both in the same direction expect during irrigation event, because soil is submerged by traditional irrigation system water.

## 6. Conclusion

In this comparison of cases studied here, we observe that GA stability is essential to optimize parameters . The results obtained don't change significantly from each 10 runs, then the optimal vector is  $K_{opt} = \langle 0.93, 0.61 \rangle$  . We have tried to show that genetic algorithm is a powerful method to optimize parameters of Priestley-Taylor approximation of canopy transpiration. Instead of standard values of  $\alpha_p = 1.26$  and  $f_g = 1$  for wet regions, which depend on climatic and soil characteristic, GA gives an optimal values as  $\langle \alpha_p=0.93, f_g=0.61 \rangle$  for semi-arid area. Stability optimization is essential, furthermore the GA can be identifying another minimum of optimal parameters in the beginning of computation, but the computation continue since there is no stopping criterion other than the final generation.

The results show an improvement of canopy transpiration then also enhance the TSEB Model performance, since correlation, bias and root mean square error become

respectively equal 0.45, +15 W.m<sup>-2</sup>, and 63 W.m<sup>-2</sup>. Thus, the results obtained in this study show the most important support of Genetic Algorithm in the calibration and optimization processes. This GAs optimization could replace measures terrain and long experiments since it improve results mostly by making use of fitness function and genetic operators such as selection, crossover and mutation. However, the set of canopy transpiration was improved.

## Appendix A

### TSEB Equations

Soil and vegetation temperature contribute to the radiometric surface temperature in proportion to the fraction of the radiometer view that is occupied by each component along with the component temperature. In particular, assuming that the observed radiometric temperature, (Trad) is the combination of soil and canopy temperatures, the TSEB model adds the following relationship (Becker and Li, 1990) to the set of (Eqs 12 and 13):

$$\text{Trad}(\theta) = [f(\theta) \cdot T_c^4 + (1-f(\theta)) \cdot T_s^4]^{1/4} \quad (\text{A.1})$$

where  $T_c$  and  $T_s$  are vegetation and soil surface temperatures, and  $f(\theta)$  is the vegetation directional fractional cover (Campbell and Norman, 1998).

$$f(\theta) = 1 - \exp(-0.5 \text{LAI} / \cos(\theta)) \quad (\text{A.2})$$

The simple fractional cover ( $f_c$ ) is as follows:

$$f_c = 1 - \exp(-0.5 \text{LAI}) \quad (\text{A.3})$$

LAI is the leaf area index, and the fraction of LAI that is green ( $f_g$ ) is required as an input and may be obtained from knowledge of the phenology of the vegetation.

The total net radiation  $R_n$  (Wm<sup>-2</sup>) is

$$R_n = H + LE + G \quad (\text{A.4})$$

where  $H$  (Wm<sup>-2</sup>) is the sensible heat flux,  $LE$  (Wm<sup>-2</sup>) is the latent heat, and  $G$  (Wm<sup>-2</sup>) is the soil heat flux. The estimation of total net radiation,  $R_n$  can be obtained by computing the net available energy considering the rate lost by surface reflection in the short wave (0.3/2.5μm) and emitted in the long wave (6/100μm):

$$R_n = (1 - \alpha_s) \cdot SW + \epsilon_s \cdot LW - \epsilon_s \cdot \sigma \cdot \text{Trad}^4 \quad (\text{A.5})$$

where SW ( $Wm^{-2}$ ) is the global incoming solar radiation, LW ( $Wm^{-2}$ ) is the terrestrial infrared radiation,  $\alpha_s$  is the surface albedo,  $\epsilon_s$  is the surface emissivity,  $\sigma$  is the Stefan-Boltzmann constant,  $T_{rad}$  ( $^{\circ}K$ ) is the radiometric surface temperature.

The estimation of soil net radiation,  $R_{ns}$  can be obtained by

$$R_{ns} = R_n \exp(-K_s LAI / \sqrt{2 \cdot \cos(\theta)}) \quad (A.6)$$

where  $k_s$  is a constant ranging between 0.4 to 0.6 and  $\theta$  is the zenithal solar angle.

The  $R_{nc}$  is the canopy net radiation as

$$R_{nc} = R_n - R_{ns} \quad (A.7)$$

where  $R_n$  is obtained using (A.4-5) and  $\theta$  is the solar zenith angle. The soil heat flux,  $G$  ( $Wm^{-2}$ ) can be expressed as a constant fraction  $c_g$  ( $\approx 0.35$ ) of the net radiation at the soil surface by

$$G = c_g R_{ns} \quad (A.8)$$

The constant of  $c_g$  ( $\approx 0.35$ ) is midway between its likely limits of 0.2 and 0.5 (Choudhury et al 1987). The canopy latent heat  $LE_c$  is given by Priestly-Taylor approximation (Priestly-Taylor, 1972).

$$LE_c = R_{nc} \cdot c_p \cdot f_g \cdot \frac{\Delta}{\Delta + \Gamma} \quad (A.9)$$

where  $c_p$  is the Priestly-Taylor constant, which is initially set to 1.26 (Norman et al 1995; Agam et al 2010),  $f_g$  is the fraction of the LAI that is green,  $\Delta$  is the slope of saturation vapor pressure versus temperature curve,  $\Gamma$  is the psychrometer constant (e.g: 0.066 kPa  $C^{-1}$ ). If no information is available on  $f_g$ , then it is assumed to be near unity. As will become apparent later (A.9) is only an initial approximation of canopy latent heat.

If in any case  $LE_c \leq 0$ , then  $LE_c$  is set to zero (i.e: no condensation under daytime convective conditions)

The sum of the contribution of the soil and canopy net radiation, total latent and sensible heat is according to the following equations

$$R_{ns} = H_s + LE_s + G \quad (A.10)$$

$$R_{nc} = H_c + LE_c \quad (A.11)$$

$$LE_t = LE_c + LE_s \quad (A.12)$$

where the subscript s and c designs soil and canopy.

The TSEB model considers also the contributions from the soil and canopy separately and it uses a few additional parameters to solve for the total sensible heat  $H_t$  which is the sum of the contribution of the soil  $H_s$  and of the canopy  $H_c$  according to the following equations

$$H_t = H_s + H_c \quad (A.13)$$

$$H_c = \rho C_p \left[ \frac{T_c - T_a}{R_a} \right] \quad (A.14)$$

$$H_s = \rho C_p \left[ \frac{T_s - T_a}{R_s + R_a} \right] \quad (A.15)$$

Where  $\rho$  ( $Kg.m^{-3}$ ) is the air density,  $C_p$  is the specific heat of air ( $JKg^{-1} K^{-1}$ ),  $T_a$  ( $^{\circ}K$ ) is the air temperature at certain reference height, which satisfies the bulk resistance formulation for sensible heat transport (Kustas et al, 2007).  $R_a$  ( $sm^{-1}$ ) is the aerodynamic resistance to heat transport across the temperature difference that can be evaluated by the following equation (Brutsaert, 1982):

$$R_a = \frac{\ln \left[ \frac{(z - d_0) - \Psi_H}{z_0} \right]}{k U_*} \quad (A.16)$$

Where  $z_0$  is the height of air wind measurements,  $U_*$  is the wind friction velocity,  $d_0$  (m) is the displacement height,  $Z_0, H$  is a roughness parameter (m) that can be evaluated as function of the canopy height (Shuttleworth and Wallace, 1985),  $k$  is the von Karman's constant ( $\approx 0.4$ ),  $\Psi_H$  is the diabatic correction factor for heat is computed (Paulson, 1970):

$$\Psi_H = 2. \ln \left[ \frac{1 + \theta_h^2}{z} \right] \quad (A.17)$$

where  $\theta$  is a universal function for heat defined by: (Brutsaert, 1982; Paulson, 1970)

$$\theta_h = (1 - 16 \cdot \xi)^{1/4} \quad (A.18)$$

The term  $\xi$  is dimensionless variable relating observation height  $Z$ , to Monin-Obukhov stability  $L_{mo}$ .

$L_{mo}$  is approximately the height at which aerodynamic shear, or mechanical, energy is equal to buoyancy energy (i.e: convection caused by an air density gradient). It is determined from

$$Lmo = -\rho \frac{u_*^3}{k g \left( \frac{H}{c_p T_a} + 0.61 \frac{LE}{\lambda} \right)} \quad (A.19)$$

Where  $\rho$  ( $\text{Kg m}^{-3}$ ) is the air density,  $C_p$  is the specific heat of air ( $\text{JKg}^{-1} \text{K}^{-1}$ ),  $T_a$  ( $^{\circ}\text{K}$ ) is the air temperature at certain reference height,  $H$  is a sensible heat flux,  $LE$  is a latent heat flux, and  $\lambda$  is the latent heat.

Friction velocity is a measure of shear stress at the surface, and can be found from the logarithmic wind profile relationship:

$$U_* = \frac{k U_a}{\ln \left[ \frac{z_M - d_g}{z_{0,M}} - \Psi_M \right]} \quad (A.20)$$

Where  $U_a$  is the wind speed and  $\Psi_M$  is the diabatic correction for momentum.

The  $R_s$  ( $\text{sm}^{-1}$ ) is the soil resistance to the heat transfer (Goudriaan, 1977; Norman et al 1995; Sauer et al 1995; Kustas et al, 1999), between the soil surface and a height representing the canopy, and then a reasonable simplified equation is:

$$R_s = \frac{1}{a' + b' U_s} \quad (A.21)$$

Where  $a' = 0.004$  ( $\text{ms}^{-1}$ ),  $b' = 0.012$  and  $U_s$  is the wind speed in ( $\text{ms}^{-1}$ ) at a height above the soil surface where the effect of the soil surface roughness is minimal; typically 0.05 to 0.2 m. These coefficients depend on turbulent length scale in the canopy, soil surface roughness and turbulence intensity in the canopy and are discussed by (Sauer et al. 1995). If soil temperature is great than air temperature the constant  $a'$  becomes  $a' = c \cdot (T_s - T_c)^{1/3}$  with  $c = 0.004$ .

$U_s$  is the wind speed just above the soil surface as described by (Goudriaan 1977):

$$U_s = U_c \cdot \exp \left[ -a \left( 1 - \frac{0.95}{h_c} \right) \right] \quad (A.22)$$

Where the factor ( $a$ ) is given by (Goudriaan 1977) as

$$a = 0.28 \cdot P^{2/3} \cdot h_c^{-1/3} \cdot s^{-2/3} \quad (A.23)$$

The mean leaf size ( $s$ ) is given by four times the leaf area divided by the perimeter.

$U_c$  is the wind speed at the top of the canopy, given by:

$$U_c = U_a \frac{\ln \left( \frac{h_c - d}{z_M} \right)}{\ln \left( \frac{z_M - d}{z_{0,M}} \right) - \Psi_M} \quad (A.24)$$

Where  $U_a$  is the wind speed above the canopy at height  $Z_u$  and the stability correction at the top of the canopy is assumed negligible due to roughness sublayer effects (Garratt, 1980; Cellier et al, 1992).

### TSEB implementation and algorithm

The TSEB model is run with the use of ground thermal remote sensing and meteorological data of Agdal site during 2003. Some model constant parameters are supposed invariable along time such as the Priestly-Taylor constant  $\alpha_p$ , albedo, emissivity, leaf area index (LAI), the fraction of the LAI that is green ( $fg$ ), leaf size ( $s$ ), the vegetation height and a constant fraction ( $cg$ ) of the net radiation at the soil surface. These considerations are certainly some consequences on model results according to seasons. The Priestly-Taylor constant  $\alpha_p$  is fixed to 1.26 (McNaughton and Spriggs 1987). The albedo, value of 0.11 is an annual averaged measured with CNR1, and a surface emissivity of 0.98, the leaf area index (LAI) is equal to 3 (Ezzahar et al, 2007). The fraction of LAI ( $fg$ ) that is green is fixed to 90% of vegetation (i.e: 10% of vegetation could be considered no active). The mean leaf size ( $s$ ), is given by four times the leaf area divided by the perimeter ( $s=0.01$ ). The average height of the olive trees is 6 meters. The fraction of the net radiation at the soil surface is fixed to  $cg=0.35$ .

Sensible and latent heat flux components for soil and vegetation are computed by TSEB, only in the atmospheric surface layer instability. Note that the storage of heat within the canopy and energy for photosynthesis are considered negligible for the instantaneous measurements. The total computed heat flux components are then from equations (A.5-8).

The canopy heat fluxes are solved by first estimating the canopy latent heat flux from the Priestly-Taylor relation (A.9), which provides an initial estimation of the canopy fluxes, and can be overridden if vegetation is under stress (Norman et al., 1995). Outside the positive latent heat situation, two cases of stress occur, when the computed value for canopy ( $LE_c$ ) or soil ( $LE_s$ ) latent heat become negative which are an unrealistic conditions.

In the first case, the normal evaluation procedure is overridden by setting ( $LE_c$ ) to zero and the remaining flux components are balanced by (A. 1-10-11-13-15). But in the second case, ( $LE_s$ ) is recomputed by using specific soil Bowen Ratio determined by  $\beta = H_s/LE_s$  and flux components are next balanced by (A.1-10-11-13-15).



In order to solve (A.15) additional computations are needed to determine soil temperature, and the resistance terms  $R_{ah}$  and  $R_s$  but as will become apparent, they must be solved iteratively. Soil temperature is determined from two equations: one to relate the observed radiometric temperature to the soil and vegetation canopy temperature, and another to determine the vegetation canopy temperature. The composite temperature is related to soil and canopy temperatures by (A.1). The resistance components are determined from (A.16), for  $R_{ah}$  and the following equation (Sauer et al., 1995) for  $R_s$  (A.18).

To complete the solution of the soil heat flux components, the ground stock heat flux can be computed as a fraction of net radiation at the soil surface (A.8).

Applying energy balance for the two source flux components resolves the surface fluxes, which cannot be reached directly because of the interdependence between atmospheric stability corrections, near surface wind speeds, and surface resistances (A.16-17). In these equations, the

stability correction factors  $\Psi_M$  and  $\Psi_H$  depend upon the surface energy flux components  $H$  and  $LE$  via the Monin-Obukhov roughness length  $L_{mo}$ .

TSEB computation for solving the surface energy balance by ten primary unknowns and ten associated equations (Table.1), needs an iterative solution process by setting a large negative value to  $L_{mo}$  (i.e: in highly unstable atmospheric conditions). This permits an initial set of stability correction factors  $\Psi_M$  and  $\Psi_H$  to be computed. Computed iteration is repeated until  $L_{mo}$  converges.

## Acknowledgments

This work is considered within the framework of research between the University of Cadi Ayad Gueliz, Marrakech, Morocco, and the Department of National Service of Meteorology, Morocco (DMN, Morocco). The first author is very grateful for encouragement to all his family especially to Mrs F. Bent Ahmed his mother, Mrs K. Aglou his wife and Mr Mahjoub .Mouida his brother and his sister Khadija Mouida

Finally the authors gratefully acknowledge evaluation and judgments by reviewers, and the editor.

## References

- [1] Agam et al, "Application of the Priestley-Taylor Approach in Two Source Surface Energy Balance Model", Am Meteo Soc, Journal of Hydrometeorology, Volume 11, 2010, pp. 185-198.
- [2] Becker. F, and Li. Z.L. "Temperature independent spectral indices in thermal infrared bands," Remote Sensing of Environment, vol. 32, 1990, pp. 17-33.
- [3] Brutsaert. W, "Evaporation Into The Atmosphere," D. Reidel, Dordrecht. 1982.

- [4] Campbell. G. S, and Norman. J. M, "An Introduction to Environmental Biophysics" (2nd ed.). New York: Springer-Verlag., 1998.
- [5] Castellvi. F, Stockle. C.O, Perez. P.J, Ibanez, M, "Comparison of methods for applying the Priestley-Taylor equation at a regional scale" Hydrol. Process. 15, 2001. pp. 1609-1620.
- [6] Cellier et al, "Flux-gradient relationships above tall plant canopies," Agric. For. Meteorol. 58, 1992. Pp. 93-117.
- [7] Choudhury. B.J, Idso. S.B, and Reginato. R.J, "Analysis of an empirical model for soil heat flux under a growing wheat crop for estimating evaporation by an infrared-temperature based energy balance equation," Agric. For. Meteorol., 39, 1987. pp. 283-297.
- [8] Ezzahar. J, "Spatialisation des flux d'énergie et de masse à l'interface Biosphère-Atmosphère dans les régions semi-arides en utilisant la méthode de scintillation", Ph.D. Thesis University Cadi Ayyad, Marrakech, Morocco, 2007.
- [9] Garratt et al, "Momentum, heat and water vapor transfer to and from natural and artificial surfaces,". Q. J. R. Meteorol. Soc., 99, 1973. pp.680-687.
- [10] Goldberg et al. "A Comparative Analysis of Selection Schemes Used in Genetic Algorithms", Foundations of Genetic Algorithms, G.Rawlins, ed. Morgan-Kaufmann. Pp 69-93
- [11] Goudriaan. J, "Crop Micrometeorology: A Simulation Study", Center for Agricultural Publications and Documentation, Wageningen. 1977.
- [12] Holland. J, "Adaptation In Natural and Artificial Systems" University of Michigan Press. 1975.
- [13] Kustas. W.P, Norman. J.M, "Evaluation of soil and vegetation heat flux predictions using a simple two-source model with radiometric temperatures for partial canopy cover", Agric. For. Meteorol. 94, 1999a, pp. 75-94.
- [14] Kustas. W. P, & Norman. J. M, "A two-source energy balance approach using directional radiometric temperature observations for sparse canopy covered surfaces", Agronomy Journal, 92, 2000. Pp. 847-854.
- [15] Kustas et al, "Utility of radiometric-aerodynamic temperature relations for heat flux estimation", Bound.-Lay. Meteorol., 122, 2007. pp.167-187,
- [16] McNaughton. K. G, and T. W. Spriggs, "An evaluation of the Priestley and Taylor equation and the complimentary relationship using results from a mixed-layer model of the convective boundary layer", T. A. Black, D. L, 1987. pp. 89-104
- [17] McNaughton. K. G, & Jarvis. P. G, "Effects of spatial scale on stomatal control of transpiration", Agricultural and Forest Meteorology, 54, 1991. pp. 269-301.
- [18] Michalewicz. Z, "Genetic Algorithms and Data Structures", Evolutionary Programs, Springer-Verlag, AI Series, New York. 1992.
- [19] Norman et al, "Source approach for estimating soil and vegetation energy fluxes in observations of directional radiometric surface temperature", Agricultural and Forest Meteorology 77, 1995. pp. 263-293.
- [20] Priestley. C. H. B, & Taylor. R. J, "On the assessment of surface heat flux and evaporation using large-scale parameters", Monthly Weather Review, 100, 1972. pp. 81-92.

- [21] Paulson. C.A, "The mathematical representation of wind speed and temperature profiles in the unstable atmospheric surface layer", J. Appl. Meteorol, 9, 1970. pp. 857-861.
- [22] Sauer et al, " Measurement of heat and vapor transfer at the soil surface beneath a maize canopy using source plates", Agric. For. Meteorol., 75, 1995. pp. 161-189.
- [23] Shuttleworth. W.J, and Wallace. J.S, "Evaporation from sparse canopies-an energy combination theory", Q. J. R. Meteorol. Sot., 111, 1985. pp. 839-855.

**First Author** is an engineer in meteorology since 1986 to 2004, Chief Engineer in meteorology 2004-2011, and Chief Operating Meteorological Service 2000-2011, current research is about estimation of fire forest risk using water stress mapping and meteorological data.

**Second Author** received his Master of Science and his Ph.D. degrees from the University of Nancy I France respectively in 1986 and 1989. In 2006, he received the HDR in Applied Mathematics from the University of Cadi Ayyad, Morocco. He is currently Professor of modeling and scientific computing at the Faculty of Sciences and Technology of Marrakech. His research is geared towards non-linear mathematical models and their analysis and digital processing applications.

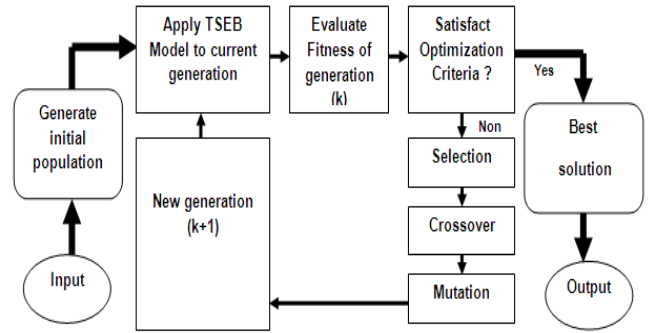


Fig.1: Iterative Procedure of a Genetic Algorithm to TSEB Model

Fig. 1 Iterative procedure of a Genetic Algorithm to TSEB Model

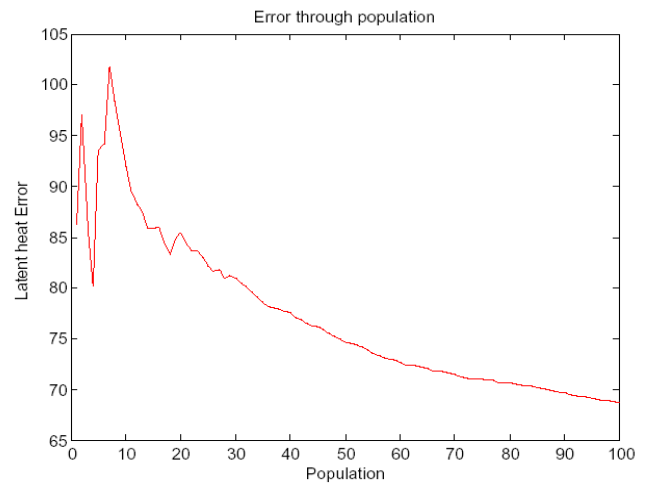


Fig. 2 Error evolution during genetic algorithm with 10 generations and 10 individual's population

## Figures

Table 1: Results of Ten Runs genetic algorithm

Runs	Error Stabilization		
	ap	fg	$\sigma(K)$
1	0.75	0.65	66.4
2	0.72	0.59	71.1
3	1.9	0.26	67.0
4	1.00	0.78	85.2
5	0.82	0.71	78.2
6	0.72	0.49	77.1
7	0.94	0.61	84.7
8	0.76	0.79	69.9
9	0.95	0.55	66.4
10	0.78	0.73	85.1

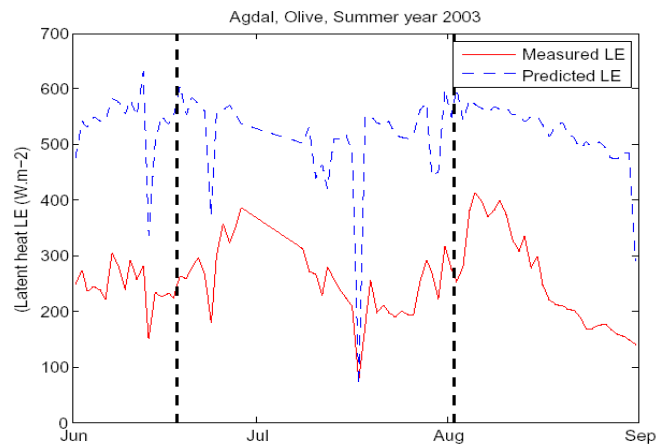


Fig. 3 Comparison between predicted and measured latent heat before optimization with Standard values of  $K=\langle ap=1.26, fg=1 \rangle$

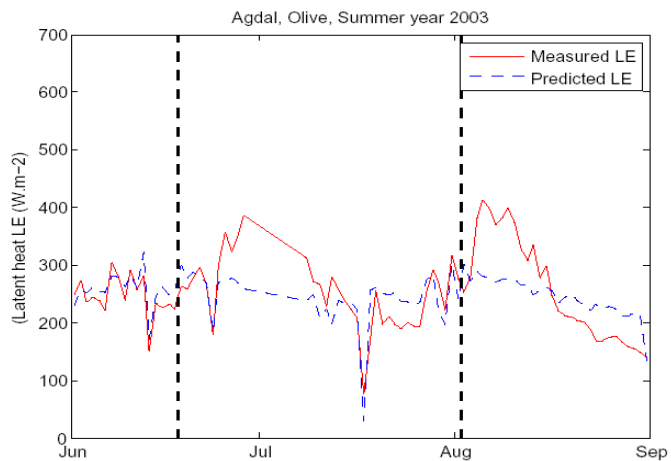


Fig. 4 Comparison between predicted and measured latent heat after optimization with optimal values of  $K = \langle \alpha_p = 0.93, f_g = 0.61 \rangle$

# An Approach to Cost Effective Regression Testing in Black-Box Testing Environment

Prof. A. Ananda Rao <sup>1</sup> and Kiran Kumar J <sup>2</sup>

<sup>1</sup> Prof. of CSE and Principal, JNTUA  
Anantapur, Andhra Pradesh, India

<sup>2</sup> Software Engineer, IBM  
India

## Abstract

Regression testing is an expensive and frequently executed maintenance activity used to revalidate the modified software. As the regression testing is a frequently executed activity in the software maintenance phase, it occupies a large portion of the software maintenance budget. Any reduction in the cost of regression testing would help to reduce the software maintenance cost. The current research is focused on finding the ways to reduce the regression testing cost. In this paper, an approach to test suite reduction for regression testing in black box environment has been proposed. This type of approach has not been used earlier. The reduced regression test suite has the same bug finding capability and covers the same functionality as the original regression test suite. The proposed approach is applied on four real-time case studies. It is found that the reduction in cost of regression testing for each regression testing cycle is ranging between 19.35 and 32.10 percent. Since regression testing is done more frequently in software maintenance phase, the overall software maintenance cost can be reduced considerably by applying the proposed approach.

**Keywords:** *Software maintenance cost, ETL DB Component, reduced test suite, reduced regression test suite, test case design, regression testing cost reduction.*

## 1. Introduction

The estimated cost of software maintenance activities occupies as much as two-thirds of the total cost of software production [18]. Regression testing is a critical part of the software maintenance that is performed on the modified software to ensure that the modifications do not adversely affect the unchanged portion of the software. As regression testing is performed frequently in software maintenance, it accounts for a large portion of the maintenance costs [9, 10, 11]. Regression testing is “selective retesting of a system or component to verify that modifications have not caused unintended effects and that

the system or component still complies with its specified requirements.” [1].

Numerous techniques have been proposed to deal with the regression testing costs. Regression test selection techniques select a subset of existing test case set for execution, depending on criteria such as changes made to the software. Test suite reduction techniques reduce the test suite permanently by identifying and removing redundant tests. Test case prioritization techniques retain the complete test suite, but change the order of test cases prior to execution, attempting to find the defects earlier during the testing. During software maintenance phase, testing teams need to run regression test case set on many intermediate builds, to ensure that the bug fixes or enhancements made to the software do not adversely affect unchanged portions of the software. In this paper, an approach to reduce the total number of regression test cases in black box environment without affecting the defect coverage and functionality coverage of software is proposed. This reduction in the regression test suite size will reduce the effort and time required by the testing teams to execute the regression test suite.

Most of the existing approaches consider test suite which contain, test cases to test the functionality, boundary values, stress, and performance of the software. Any reduction in this test suite size will reduce the testing time, effort, and cost. Many of the test cases in this test suite belong to the functionality and boundary values of the software. The proposed approach is applied on the original test suite to derive the reduced test suite. This reduced test suite covers the same functionality of the software as the original test suite. A regression test selection method is applied on this reduced test suite, to get the reduced regression test suite. This reduced regression test suite covers the same defect coverage and functionality as the

original regression test suite. In this proposed approach, it is shown that the two aspects of testing, that is testing for functionality and testing for boundary values can be tested with reduced test suite as these two aspects can be tested together simultaneously in most of the situations. The situations where these two aspects can be tested simultaneously, is also shown with help of the case-studies. In this paper, testing simultaneously means, a single test case can cover both the above mentioned aspects for a particular situation. The proposed approach is applied on four real-time case studies and also estimated the reduction in cost of regression testing using a cost estimation model. It is found that the reduction in cost per one regression testing cycle is ranging between 19.35 and 32.10 percent. Since regression testing is more frequently done activity in software maintenance phase, the overall regression testing cost can be reduced considerably by applying the proposed approach.

The rest of the paper is organized as follows: Section II reviews the various regression testing techniques and summarizes related work. Section III describes the proposed approach to cost effective regression testing for black-box testing environment. Section IV describes the Empirical studies and results of the proposed approach. Section V concludes and discusses future work.

## 2. Related Work

Researchers, practitioners and academicians proposed various techniques on test suite reduction, test case prioritization, and regression test selection for improving the cost effectiveness of the regression testing.

Rothermel and Harrold presented a technique for regression test selection. Their algorithms construct control flow graphs for a procedure or program and its modified version and use these graphs to select tests that execute changed code from the original test suite [9]. James A. Jones and Mary Jean Harrold proposed new algorithms for test suite reduction and prioritization [2]. Saifur-Rehman Khan, Aamer Nadeem proposed a novel test case reduction technique called TestFilter that uses the statement-coverage criterion for reduction of test cases [3]. T. Y. Chen and M. F. Lau presented dividing strategies for the optimization of of a test suite [4]. M. J. Harrold et al presented a technique to select a representative set of test cases from a test suite that provides the same coverage as the entire test suite [5]. This selection is performed by identifying, and then eliminating, the redundant and obsolete test cases in the test suite. This technique is illustrated using data flow testing methodology. A recent study by Wong, Horgan, London, and Mathur [6],

examines the costs and benefits of test suite minimization. Rothermel et al [7] described several techniques for using test execution information to prioritize test cases for regression testing, including: techniques that order test cases based on their total coverage of code components, techniques that order test cases based on their coverage of code components not previously covered, and techniques that order test cases based on their estimated ability to reveal faults in the code components that they cover.

Most of the techniques described in the above papers assume that source code of the software is available to the testing engineer at the time of testing. But in most of the organizations the testing is done in black box environment and the source code of the software is not available to the testing engineers. In this paper, an approach to reduce cost of software regression testing in black box environment, without affecting the functionality coverage, is presented.

## 3. The Proposed Approach

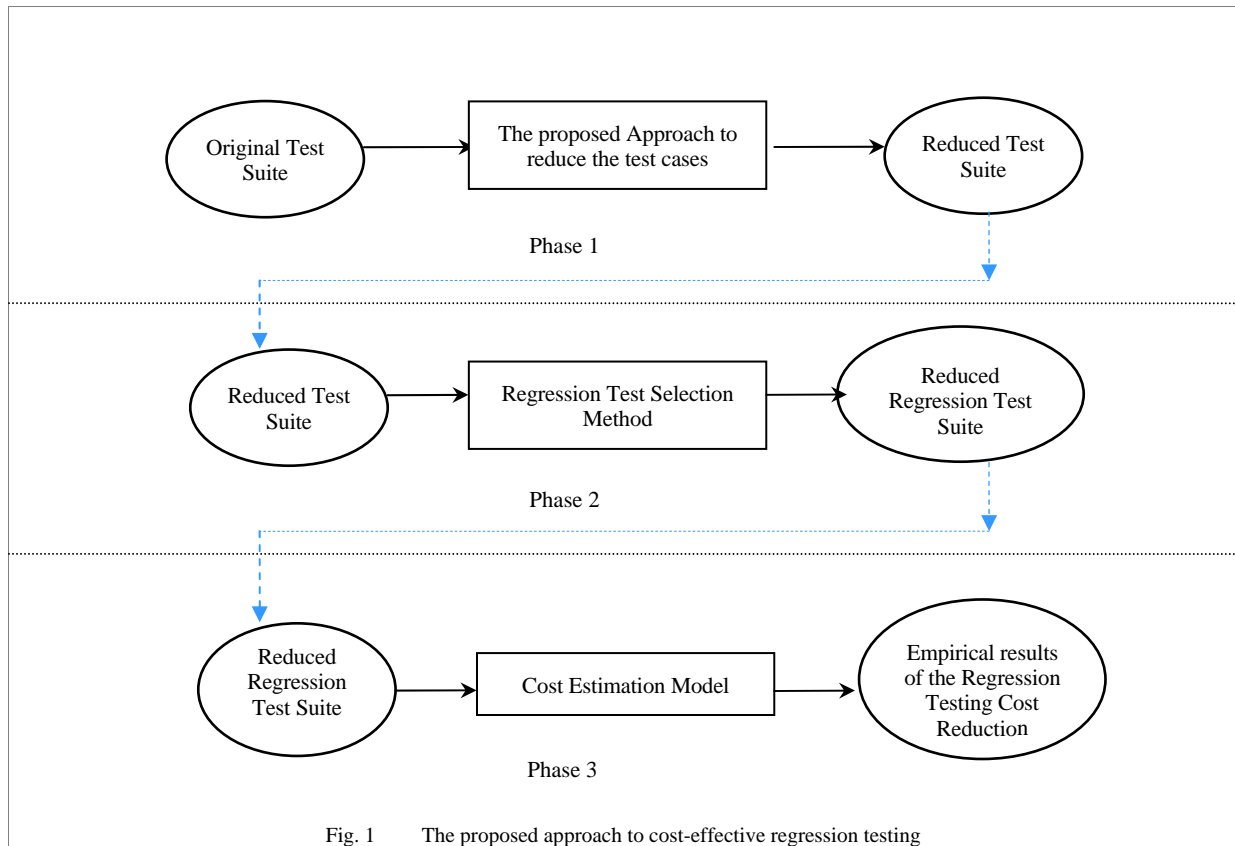
The estimated cost of software maintenance exceeds 70% of total software costs [16], and large portion of this maintenance expense is devoted to regression testing. Regression testing is a frequently executed activity, so reducing the cost of regression testing would help in reducing cost of the software maintenance.

The proposed approach is shown in three phases (Fig.1). In Phase 1 (Fig. 1), the “Reduced Test Suite” is derived by applying the proposed approach on the Original test suite. Phase 1 of the approach is already proposed by the authors in [17], and in Phase 2 (Fig. 1), the “Reduced Regression Test Suite” is derived by applying a regression test selection method on the “Reduced Test Suite” that is derived in the Phase 1. In Phase 3, a testing cost-estimation model is applied on the reduced regression test suite and empirically calculated the regression testing cost reduction by the proposed approach.

### *Phase 1: Deriving the “Reduced Test Suite”*

A large number of test cases are derived by applying various testing techniques to test complete functionality of a software product. This test suite contains test cases to test functionality, boundary values, stress, and performance of the software product. Majority of these test cases will be test cases that test the functionality and boundary values. The Phase 1 of the proposed approach is focused on reducing test cases considering test cases that test functionality and boundary values.





The Phase 1 (Fig.1) of the approach contains the following four steps:

1. View the two aspects that is functionality and boundary value testing together
2. Identify the situation(s) (considering functionality and boundary values) which can be tested in single test case(s) so as to design minimal test cases
3. Proving logically that the single test case(s) in-fact covering both the aspects.
4. Applying above three steps to case studies and validating

By applying the above mentioned approach we get the “Reduced Test Suite” that covers the same functionality of the software as the original test suite. This is validated in the case studies.

**Phase 2: Deriving the “Reduced Regression Test Suite”**

Regression testing process involves selecting a subset of the test cases from the original test suite, and if necessary creates some new test cases to test the modified software.

Let  $P$  is the original software product,  $P'$  is the modified software product and  $T$  is the set test cases to test  $P$ . A typical regression testing on modified software proceeds as follows:

- A. Select  $T' \subseteq T$ , a set of test cases to execute on the modified software product  $P'$ .
- B. Test  $P'$  with  $T'$ , to verify modified software product’s correctness with respect to  $T'$ .
- C. If necessary, create  $T''$ , a set of new test cases to test  $P'$ .
- D. Test  $P'$  with new tests  $T''$ , to verify  $P'$  correctness with respect to  $T''$ .

In Phase 1 (Fig 1), the “Reduced Test Suite” is derived. In Phase 2 (Fig 1), the “Reduced Regression Test Suite” is derived by applying the regression test selection method shown in the Figure 2. This regression test select ion method contains the following 3 steps:

1. Select a subset of test cases from the reduced test suite (derived in Phase1) which covers the major functionality of the product.
2. Select test cases that cover the scenarios to test the bug fixes included in the regression build

3. Create new test cases, to test the (if any) new enhancements included in the regression build.

In step1 of this approach, we are selecting subset of test cases from the reduced test suite. So, this selected subset will also contain the less number of tests as compared to the subset selected from the original test suite. This reduced regression test suite covers the same functionality as the original regression test suite that is derived without applying our approach.

The reduced regression test suite derived using this approach is empirically evaluated in the ‘case studies’ section of the paper.

**Phase 3: Regression Testing Cost Estimation**

In Phase 3 of the proposed approach we calculate the estimated reduction in regression testing achieved by using the proposed approach. The authors proposed an approach to cost estimation in black-box testing environment in [19]. Using this approach the regression testing in black-box environment involves the following major activities.

- Environment setup for testing ( $T_{env}$ )

- Verification of the fixed bugs which were reported in the previous testing cycle ( $T_{bv}$ )
- Test Suite execution ( $T_e$ )
- Test Report Generation ( $T_{rg}$ )
- Test Report Analysis ( $T_{ra}$ )
- Reporting the Bugs ( $T_{br}$ )

As the above mentioned actives are performed on each and every build, they occupies major portion of the overall regression testing time. The time required to complete regression testing on one intermediate or regression build is calculated using the following equation.

$$Eib = T_{env} + ((Nt \times T_e) / 60) + T_{rg} + T_{ra} + T_{bv} + T_{br} \tag{1}$$

where, the ‘ $T_e$ ’ indicates the average time required to execute a single test case and the ‘ $Nt$ ’ is the total number of the test cases executed for that particular regression testing cycle.

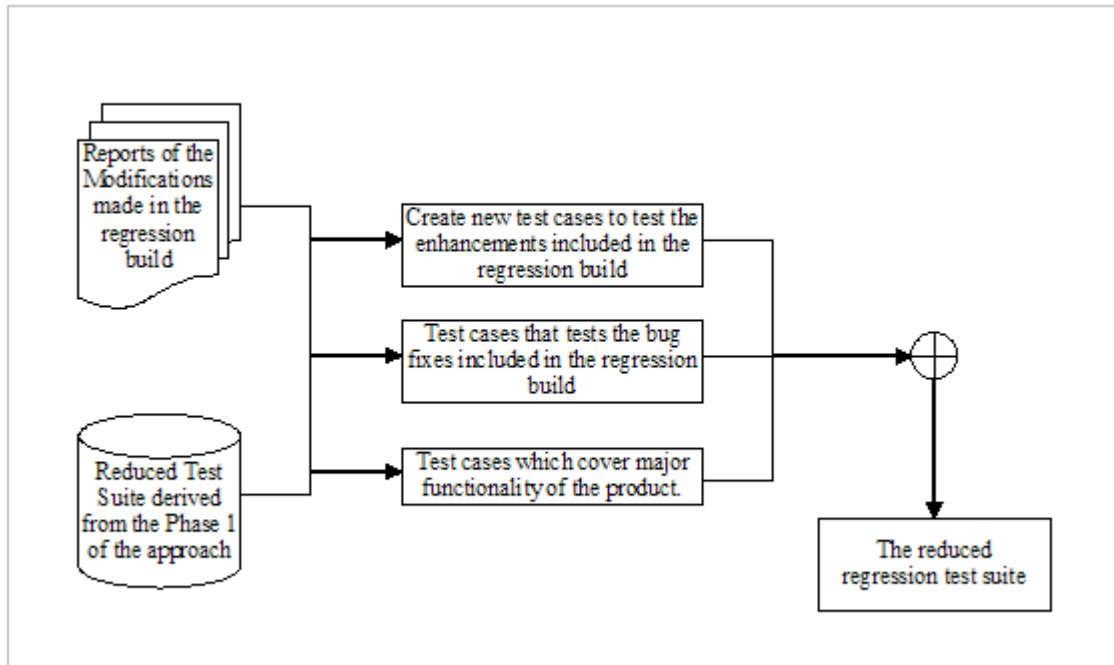


Fig. 2 The regression test case selection

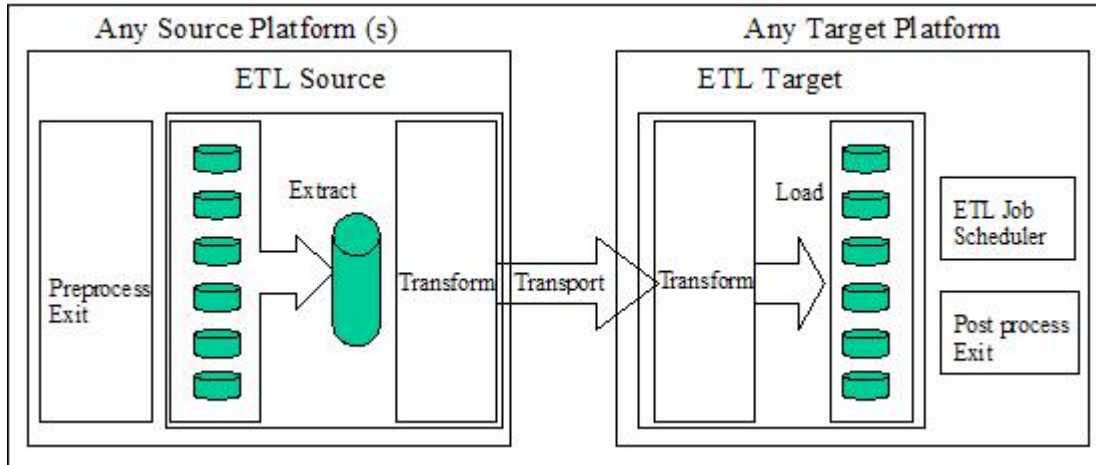


Fig. 3 The ETL process

The equation (1) gives the estimated effort required to test one regression build, in man-hours. The estimated the regression testing cost on a regression build can be calculated using the following equation.

$$C_{total} = Se \times E_{total} \quad (2)$$

where, 'Se' is the average salary paid to a testing engineer per man-hour.

The salary paid to the employee per man-hour mainly depends on the organization and geography of the employee. So, the estimated regression testing cost for the product can be calculated based on these factors and using equation (2).

The following section describes the empirical validation of the proposed approach.

#### 4. Empirical Studies and Results

The proposed approach is applied on four real-time ETL tool (Data ware housing tool) components: DB2 ETL DB Component, Sybase ETL DB Component, Teradata ETL DB Component and MySQL ETL DB Component. Concepts explained in Fig. 3 and Fig. 4, are generic and applicable to all the above four case studies. In Fig. 3, ETL, which stands for "extract, transform and load", is the set of functions combined into one tool or solution that enables companies to "extract" data from numerous databases, applications and systems, "transform" it as appropriate, and "load" it into another databases, a data mart or a data warehouse for analysis, or send it along to another operational system to support a business process.

The phase 1 of the approach is applied to the case studies as given below:

#### Phase 1: Deriving the "Reduced Test Suite"

The test suite that tests the complete functionality of an ETL tool include: Functional test cases ( $T_f$ ), Boundary Value test cases ( $T_b$ ), Stress test cases ( $T_s$ ), Performance test cases ( $T_p$ ) and other test cases ( $T_o$ ) like negative test cases. So the Total Number of test cases ( $T_n$ ) are:

$$T_n = T_f + T_b + T_s + T_p + T_o$$

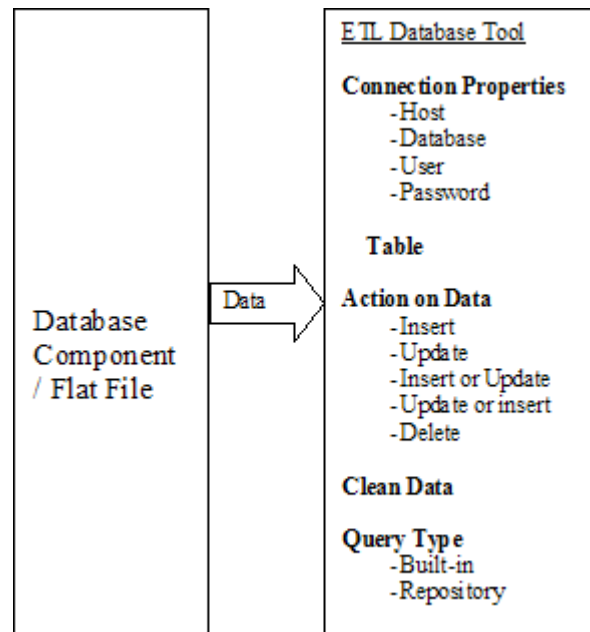


Fig. 4 The ETL Database Component write process

TABLE1. FUNCTIONAL TEST CASES BEFORE APPLYING THE PROPOSED APPROACH OF PHASE 1

Test Case ID	Description	Preconditions	Expected Result	Test Status	Comments
TCf1	Test on writing the data to the target table with Action on data = Insert		The job should add new rows to the target table and stop if duplicate rows are found.		
TCf2	Test on writing the data to the target table with Action on data = Update		The job should make changes to existing rows in the target table with the input data.		
TCf3	Test on writing the data to the target table with Action on data = Insert or Update		The job should add new rows to the target table first and then update existing rows.		
TCf4	Test on writing the data to the target table with Action on data =Update or Insert		The job should update existing rows first and then add new rows to the target table.		
TCf5	Test on writing the data to the target table with Action on data =Delete		The job should remove rows from the target table corresponding to the input data.		

The Fig. 4 shows some attributes of a generalized ETL Database Component write process. In this write process, the source could be an ETL DB Component or a flat file and the target is a ETL DB Component.

In the write process, the target ETL DB Component reads data from the source component, connects to the respective database using the connection properties specified and writes that data in to the target table.

The test case design using the phase 1 of proposed approach, for DB2 ETL DB Component is described in section A.

#### A. DB2 ETL DB Component Test Case Design

The Fig. 5 shows the metadata of the table 'sampletable' used in the DB2 ETL DB Component case study. This is a DB2 table that contains 5 columns. The col1 is integer type, col2 is character type, col3 is varchar type, col4 is decimal type and col5 is date type.

The Table 1 shows some sample Functional test cases for the DB2 ETL DB Component write process. Each of these test cases tests a single functionality or scenario of the DB2 ETL DB Component to ensure the particular attribute or function is working properly.

Column name	Datatype schema	Data type name	Column Length	Scale	Nulls
COL1	SYS	INTEGER	4	0	No
COL2	SYS	CHARACTER	9	0	Yes
COL3	SYS	VARCHAR	9	0	Yes
COL4	SYS	DECIMAL	12	3	Yes
COL5	SYS	DATE	4	0	Yes

Fig. 5 Metadata of the sample table

The Table 2 shows some sample Boundary Value test cases for the DB2 ETL DB Component write process. Each of these test cases tests a single column or data type to ensure the boundary values of that data type are written properly to the target table.

The test case design for DB2 ETL DB Component using the proposed approach of phase 1 is described in the following four sub sections (A.1 – A.4).

TABLE 2. BOUNDARY VALUE TEST CASES BEFORE APPLYING THE PROPOSED APPROACH OF PHASE 1

Test Case ID	Description	Preconditions	Expected Result	Test Status	Comments
TCb1	Test on writing the data to col1 with INTEGER data type boundary values		The job should read the INTEGER data type boundary values from input data and write to the target table successfully.		
TCb2	Test on writing the data to col2 with CHAR data type boundary values		The job should read the CHAR data type boundary values from input data and write to the target table successfully.		
TCb3	Test on writing the data to col3 with VARCHAR data type boundary values.		The job should read the VARCHAR data type boundary values from input data and write to the target table successfully.		
TCb4	Test on writing the data to col4 with DOUBLE data type boundary values		The job should read the DOUBLE data type boundary values from input data and write to the target table successfully.		

<b>TCb5</b>	Test on writing the data to col5 with DATE data type boundary values		The job should read the DATE data type boundary values from input data and write to the target table successfully.		
-------------	--	--	--	--	--

**A.1. View the two aspects together (Step 1)**

Many software testing techniques are required to test functionality of a software product completely. A large number of test cases are generated by applying the various testing techniques. These test cases include: functional test cases (Tf), Boundary Value test cases (Tb) , Stress test cases (Ts), Performance test cases (Tp) and other test cases (To) like negative test cases.

$$T_n = T_f + T_b + T_s + T_p + T_o.$$

Most of the test cases in this test suite belong to test cases that test the functionality and boundary values of the product. The proposed approach in Phase1 is focused to reduce test cases considering test cases that test functionality and boundary values.

**A.2. Identifying the situations that can be tested in a single test case and designing minimized test case set ( Step 2)**

The test case TCf1 tests the functionality of the DB2 ETL DB Component when the attribute ‘Action on Data’ is set to ‘Insert’ and the test case TCb1 tests the INTEGER data type boundary value that is written to the target DB2 table. Both of these test cases TCf1 and TCb1 are testing the two aspects i.e. functionality and boundary values of the DB2 ETL DB Component.

By using the proposed approach in phase1 these two test cases could be viewed together and tested in a single test case. For example, the test cases TCf1 and TCb1 are viewed together and designed a single test case TCm1 (Table 3) that covers the both aspects. The minimized test case set designed using the proposed approach in phase 1 is shown in the Table 3.

**A.3. Providing logically that the single test case in fact covers both the aspects (Step 3)**

Each test case in the minimized test case set described in Table 3 will test the functionality of the DB2 ETL DB Component to ensure that the particular attribute is working properly and also tests the boundary values for various columns in the target table to ensure that the boundary values of that column data type are written properly. For example, the TCm1 in the minimized test case set tests whether the DB2 ETL DB Component is working properly when the attribute ‘Action on Data’ is set to ‘Insert’ and also tests whether the INTEGER data type boundary value is written to the target table properly which were tested by the test cases TCf1 and TCb1.

In similar way, the remaining test cases in the minimized test case set {TCm1 – TCm5} described in Table 3 will test the both aspects, functionality and the boundary values of DB2 ETL DB Component which have been tested by the test cases {TCf1-TCf5 and TCb1-TCb5}.

**A.4. Applying the above three steps to case studies and validating (step 4)**

If the number of boundary value test cases that are viewed together with functional test cases, the number of test cases reduced is  $T_{br}$ . Then, after applying the phase 1 of the proposed approach, the total number of test cases is minimized to:

$$T_{min} = T_n - T_{br}$$

And, the percentage of test case reduction ( $T_{red} \%$ ) is:

$$T_{red} \% = ((T_n - T_{min}) / T_n) * 100$$

TABLE 3. THE MINIMIZED TEST CASE SET DESIGNED USING THE PROPOSED APPROACH IN PHASE 1

Test Case ID	Description	Preconditions	Expected Result	Test Status	Comments
<b>TCm1</b>	Test on writing the data to the target table with Action on data = Insert and col1 contains INTEGER data type boundary values		The job should read the input data, add new rows to the target table successfully and stop if duplicate rows are found.		
<b>TCm2</b>	Test on writing the data to the target table with Action on data = Update and col2 contains CHAR data type boundary values		The job should read the input data and make changes to existing rows in the target table with the input data		
<b>TCm3</b>	Test on writing the data to the target table with Action on data = Insert or Update and col3 contains VARCHAR data type boundary values		The job should read the input data, add new rows to the target table first and then update existing rows		
<b>TCm4</b>	Test on writing the data to the target table with Action on data = Update or Insert and col4 contains DOUBLE data type boundary values		The job should read the input data, update existing rows first and then add new rows to the target table		



<b>TCm5</b>	Test on writing the data to the target table with Action on data = Delete and col5 contains DATE data type boundary values		The job should read the input data and remove rows from the target table corresponding to the input data		
-------------	--	--	--	--	--

TABLE 4. REDUCED REGRESSION SUITE

ETL DB Component	Original Test Suite (T <sub>n</sub> )	Reduced Test Suite – Phase 1 (T <sub>min</sub> )	Original Regression Suite (T <sub>R</sub> )	Reduced Regression Suite- Phase 2 (T <sub>Rmin</sub> )
<b>DB2 ETL DB Component</b>	3563	2609 (26.7 %)	1846	1304
<b>Sybase ETL DB Component</b>	2968	2079 (29.98 %)	1497	1034
<b>Teradata ETL DB Component</b>	4234	2798 (33.91 %)	2534	1624
<b>MySQL ETL DB Component</b>	3657	2484 (32.07 %)	1668	1166

In similar way, the proposed approach is also applied on Sybase ETL DB Component, Teradata ETL DB Component and MySQL ETL DB Component. The second column of Table 4 describes the total number of test cases (T<sub>n</sub>) before applying phase 1 of the proposed approach, the third column describes the total number of test cases in the minimized test case suite (T<sub>min</sub>) after applying the phase 1 of the proposed and the percentage of test case reduction (T<sub>red</sub> %), given in parenthesis.

After applying the proposed approach in phase 1, the total number of test cases for DB2 ETL DB Component, Sybase ETL DB Component, Teradata ETL DB Component and MySQL ETL DB Component test cases are reduced by 34 %,27 %,30 % and 32 % respectively. The results indicate that the number of test case reduction is ranging between 27 to 34 percent (Table 4, 3rd column). Hence the Phase 1 of the proposed approach is validated through case studies.

### Phase 2: Deriving the “Reduced Regression Test Suite”

Regression testing is a critical part of the software maintenance that is performed on the modified software to ensure that the modifications do not adversely affect the unchanged portion of the software.

Using the proposed approach for regression test selection, we have selected a subset of test cases from the reduced test suite (derived in Phase1) which covers the major functionality of the product, selected test cases that cover the scenarios to test the bug fixes included in the regression build, and created new test cases, to test the (if any) new enhancements included in the regression build. This derived “Reduced Regression Test Suite” covers the same functionality of the software product as the regression suite that is derived from the original test suite (without reduction).

The phase 2 of the approach is applied on four case studies and the results are recorded in Table 4. The fourth column in table 4 describes the number of regression test cases (T<sub>R</sub>) that are derived by applying the proposed regression test selection method on the original test suite (i.e before applying the Phase1 of the proposed approach). The fifth column in Table 4 describes the “Reduced Regression Test Suite” (T<sub>Rmin</sub>) which is derived by applying the proposed regression test selection method on the “Reduced Test Suite” derived in Phase1.

This reduction is independent of the regression test selection method that is used to select the regression test cases. If the number of test cases in the original test suite is reduced, then subsequently the number of regression test cases also reduced.

### Phase 3: Regression Testing Cost Estimation

The table 5 presents the required average effort for each of the testing activities in black-box testing, based the historical data derived from analyzing 40 completed software projects [19].

TABLE 5. AVERAGE TIME REQUIRED FOR TESTING ACTIVITIES

Testing activity	Avg. Estimated effort
Environment setup for testing	3 Hrs
Verification of the fixed bugs	20 min / bug
Test Suite execution	1.2 min / test case
Test Report Generation	9 min
Test Report Analysis	20 min

Reporting the Bugs	18 min / bug
--------------------	--------------

The estimated effort required to complete the testing on one regression build calculated using the equation (1) is:

For original regression test suite:

$$E_{ib} = 3 + ((1864 \times 1.2) + 9 + 20 + 4 \times 20 + 4 \times 18) / 60 = 43.29 \text{ Hrs}$$

For reduced regression test suite:

$$E_{ib} = 3 + ((1304 \times 1.2) + 9 + 20 + 4 \times 20 + 4 \times 18) / 60 = 32.09 \text{ Hrs}$$

According to C. Jones [18] the average salary paid to a software engineer is \$100 per hour. The total estimated cost for testing the complete product before it gets released to the customer is calculated using the equation (2):

For original regression test Suite:

$$C_{total} = 100 \times 43.29 = 4329 \$$$

For reduced regression test Suite:

$$C_{total} = 100 \times 32.09 = 3209 \$$$

So, the estimated regression testing cost of the 'DB2 ETL DB Component' using the original regression suite is 4329 \$, and the estimated regression testing cost of the 'DB2 ETL DB Component' using the reduced regression suite is 3209 \$. In Table 6, the 4<sup>th</sup> column indicates the estimated regression testing cost using the original regression test suite, and the 5<sup>th</sup> columns indicates the estimated regression testing cost using the reduced regression test suite. For the remaining three projects the regression testing costs are estimated using the proposed approach and the final results are given in the table 6.

The average salary paid to a software engineer varies based on the organization and the geography location. As we have estimated the exact amount of effort required, the project manager could easily estimate the exact testing cost using equation (2), by substituting average salary paid to the employee in their organization.

The regression testing cost reduced by applying the proposed approach is:

$$C_{Rred} = C_R - C_{Rmin}$$

The percentage of reduction in regression testing cost is:  $C_{Rred} \% = ((C_R - C_{Rmin}) / C_R) * 100$

The regression testing cost reduced for 'DB2 ETL DB Component' calculated using the above equation is:

$$C_{Rred} \% = ((4329 - 3209) / 4329) * 100 = 25.87 \%$$

The percentage of reduction in regression testing cost ( $C_{Rred} \%$ ) by using the proposed approach, on one regression testing cycle, for various projects calculate using the above equations are shown in the 6<sup>th</sup> column of the Table 6.

The regression testing needs to be performed on many intermediate software builds of the product during the software maintenance phase.

Let  $B_n \{n=1,2,3,\dots,12\}$  is the number of builds for a particular month on which the regression testing needs to done.

$$\text{Then the total number of builds per year is } \sum_{n=1}^{12} B_n,$$

and the average number of builds per month

$$\text{is } \frac{1}{12} \times \left( \sum_{n=1}^{12} B_n \right).$$

So, the regression testing cost reduced per month is

$$(C_{Rred} \% ) \times \frac{1}{12} \times \left( \sum_{n=1}^{12} B_n \right), \text{ and}$$

$$\text{per year is } (C_{Rred} \% ) \left( \sum_{n=1}^{12} B_n \right).$$

TABLE 6. ESTIMATED REGRESSION TESTING COST REDUCTION

ETL DB Component	Original Regression Suite ( $T_R$ )	Reduced Regression Suite- Phase 2 ( $T_{Rmin}$ )	Estimated Cost to test the original Regression suite ( $T_R$ )	Estimated Cost to test the Reduced Regression Suite ( $T_{Rmin}$ )	Percentage of reduced Regression testing cost ( $T_{Rmin}$ )
DB2 ETL DB Component	1846	1304	4329	3209	25.87 %
Sybase ETL DB Component	1497	1034	3595	2669	25.75 %
Teradata ETL DB	2534	1624	5669	3849	32.10 %

<b>Component</b>					
<b>MySQL ETL DB Component</b>	1668	1166	3637	2933	19.35 %

By applying the proposed approach,  $C_{red}$  percent regression testing cost is reduced for a ETL DB Component. These case studies show that, the proposed approach saves a substantial amount of regression testing time and effort. The cost of the regression testing for DB2 ETL DB Component, Sybase ETL DB Component, Teradata ETL DB Component and MySQL ETL DB Component is reduced by 25.87 %, 25.75 %, 32.10 % and 19.35 % respectively. The results indicate that by applying the proposed approach, the reduction in cost of regression testing is ranging between 19.35 to 32.10 percent (Table 6, 6<sup>th</sup> column).

## 5. Conclusions and Future work

The proposed approach reduces the number of regression test cases in black box environment, independent of the regression test selection methods that are available. The effort required to apply this approach is a one-time effort, but it reduces the effort and time required for all the remaining regression testing cycles of the software.

The proposed approach is applied on four real-time ETL Tools (Data ware housing tools) that are used by many customers all over the world. The tested ETL tool components are DB2 ETL DB Component, Sybase ETL DB Component, Teradata ETL DB Component and MySQL ETL DB Component. It is found from the case studies that the cost of regression testing can be reduced by applying the proposed method and the reduction in regression testing cost is ranging between 19.35 and 32.10 percent. Hence, by using the proposed approach the regression testing cost can be reduced considerably.

As part of the future work, we are planning to propose an enhanced regression test selection method in black-box environment which further reduces the regression testing cost.

## References

[1] IEEE Std 610.12-1990, IEEE Standard Glossary of Software Engineering Terminology.  
 [2] James A. Jones and Mary Jean Harrold, "Test-Suite Reduction and Prioritization for Modified Condition/Decision Coverage", IEEE Transactions on Software Engineering, Vol. 29, Issue. 3, March 2003.

[3] Saif-ur-Rehman Khan Nadeem, A.Awais, "TestFilter: A Statement-Coverage Based Test Case Reduction Technique", IEEE Multitopic Conference, page(s): 275 - 280, 23-24 Dec. 2006.  
 [4] T. Y. Chen and M. F. Lau, "Dividing strategies for the optimization of a test suite", Information Processing Letters, 60(3):135-141, Mar. 1996.  
 [5] M. J. Harrold, R. Gupta, and M. L. Soffa, "A methodology for controlling the size of a test suite", ACM Transactions on Softw.Eng. and Meth., 2(3):270-285, July 1993.  
 [6] W. E.Wong, J. R. Horgan, S. London, and A. P.Mathur, "Effect of test set minimization on fault detection effectiveness", 17th international conference on Software engineering, pages 41 - 50, 1995.  
 [7] G. Rothermel, R.H. Untch, C. Chu, and M.J. Harrold, "Prioritizing Test Cases for Regression Testing," IEEE Trans. Software Eng., vol.27, no. 10, pp. 929-948, Oct. 2001.  
 [8] H. K. N. Leung and L. White, "A cost model to compare regression test strategies", In Proc. Conf. Softw. Maint., pages 201-208, Oct. 1991.  
 [9] G. Rothermel and M. J. Harrold, "A safe, efficient regression test selection technique", ACM Transactions on Software Engineering Meth.,6(2):173-210, April 1997.  
 [10] B. Beizer. Software Testing Techniques. VanNostrand Reinhold, New York, NY, 1990.  
 [11] H. K. N. Leung and L. White. "Insights into regression testing", In Conf. Softw. Maint., pages 60-69, October 1989.  
 [12] M. Jean Harrold, Rajiv Gupta, Mary Lou Soffa, "A methodology for controlling the size of a test suite, ACM Transactions on Software Engineering and Methodology, Volume 2, Issue 3, 1993.  
 [13] Zhenyu Chen, Baowen Xu, Xiaofang Zhang, Changhai Nie, "A novel approach for test suite reduction based on requirement relation contraction", Proceedings of the 2008 ACM symposium on Applied computing, Pages 390-394, 2008  
 [14] S. Parsa, A. Khalilian and Y. Fazlalizadeh, "A New Algorithm to Test Suite Reduction Based on Cluster Analysis", iccsit, pp.189-193, 2009 2nd IEEE International Conference on Computer Science and Information Technology, 2009..  
 [15] Pravin M. Kamde, V. D. Nandavadekar, R. G. Pawar, "Value of Test Cases in Software Testing", International Conference on Management of Innovation and Technology, IEEE, 2006.  
 [16] G. Rothermel, M.J. Harrold, J. Ostria, and C. Hong, "An Empirical Study of the Effects of Minimization on the Fault Detection Capabilities of Test Suites", Proc. Int'l Conf. Software Maintenance, PP. 34-43, Nov. 1998.  
 [17] Kiran Kumar J, A. Anada Rao, M. Gopi Chand, K. Narender Reddy, "An Approach to test case Design for cost effective Software Testing", IMECS-IAENG-2009.  
 [18] S. Schach, Software Engineering. Boston: Aksen Assoc., 1992.  
 [19] Kiran Kumar J and Prof. A. Ananda Rao, "An Approach to Software Testing Cost Estimation in Black-Box

- Environment", International Journal of Electrical, Electronics and Computer Systems, April 2011.
- [20] H. Agrawal, J. Horgan, E. Krauser, and S. London, "Incremental Regression Testing," Proc. Conf. Software Maintenance, pp. 348–357, Sept. 1993.
- [21] T. Ball, "On the Limit of Control Flow Analysis for Regression Test Selection," Proc. Int'l Symp. Software Testing and Analysis, ISSTA, Mar. 1998.
- [22] S. Bates and S. Horwitz, "Incremental Program Testing Using Program Dependence Graphs," Proc. 20th ACM Symp. Principles of Programming Languages, Jan. 1993.
- [23] P. Benedusi, A. Cimitile, and U. De Carlini, "Post-Maintenance Testing Based on Path Change Analysis," Proc. Conf. Software Maintenance, pp. 352–361, Oct. 1988.
- [24] D. Binkely, "Semantics Guided Regression Test Cost Reduction," IEEE Trans. Software Eng., vol. 23, no. 8, Aug. 1997.
- [25] Y.F. Chen, D.S. Rosenblum, and K.P. Vo, "TestTube: A System for Selective Regression Testing," Proc. 16th Int'l Conf. Software Eng., pp. 211–222, May 1994.
- [26] K.F. Fischer, "A Test Case Selection Method for the Validation of Software Maintenance Modification," Proc. COMPSAC'77, pp.421–426, Nov. 1977.
- [27] K.F. Fischer, F. Raji, and A. Chruscicki, "A Methodology for Retesting Modified Software," Proc. Nat'l Telecommunications Conf., pp. 1–6, Nov. 1981.
- [28] R. Gupta, M.J. Harrold, and M.L. Soffa, "An Approach to Regression Testing Using Slicing," Proc. Conf. Software Maintenance, pp.299–308, Nov. 1992.
- [29] M.J. Harrold and M.L. Soffa, "An Incremental Approach to Unit Testing During Maintenance," Proc. Conf. Software Maintenance, pp. 362–367, Oct. 1988.
- [30] M.J. Harrold and M.L. Soffa, "An Incremental Data Flow Testing Tool," Proc. Sixth Int'l Conf. Testing Computer Software, May 1989.
- [31] J. Hartmann and D.J. Robson, "RETEXT—Development of a Selective Revalidation Prototype Environment for Use in Software Maintenance," Proc. 23rd Hawaii Int'l Conf. System Sciences, pp. 92–101, Jan. 1990.
- [32] J. Hartmann and D.J. Robson, "Techniques for Selective Revalidation" IEEE Software, vol. 16, no. 1, pp. 31–38, Jan. 1990.
- [33] J. Laski and W. Szermer, "Identification of Program Modifications and Its Applications in Software Maintenance," Proc. Conf. Software Maintenance, pp. 282–290, Nov. 1992.
- [34] J.A.N. Lee and X. He, "A Methodology for Test Selection," J. Systems and Software, vol. 13, no. 1, pp. 177–185, Sept. 1990.
- [35] H.K.N. Leung and L. White, "Insights into Regression Testing," Proc. Conf. Software Maintenance, pp. 60–69, Oct. 1989.
- [36] H.K.N. Leung and L. White, "Insights into Testing and Regression Testing Global Variables," J. Software Maintenance, vol. 2, pp. 209–222, Dec. 1990.
- [37] H.K.N. Leung and L.J. White, "A Study of Integration Testing and Software Regression at the Integration Level," Proc. Conf. Software Maintenance, pp. 290–300, Nov. 1990.
- [38] T.J. Ostrand and E.J. Weyuker, "Using Dataflow Analysis for Regression Testing," Proc. Sixth Ann. Pacific Northwest Software Quality Conf., pp. 233–247, Sept. 1988.
- [39] B. Eherlund and B. Korel, "Modification Oriented Software Testing," Conf. Proc.: Quality Week, pp. 1–17, 1991.
- [40] B. Sherlund and B. Korel, "Logical Modification Oriented Software Testing," Proc. 12th Int'l Conf. Testing Computer Software, June 1995.
- [41] A.B. Taha, S.M. Thebaut, and S.S. Liu, "An Approach to Software Fault Localization and Revalidation Based on Incremental Data Flow Analysis," Proc. 13th Ann. Int'l Computer Software and Applications Conf., pp. 527–534, Sept. 1989.
- [42] F. Vokolos and P. Frankl, "Pythia: A Regression Test Selection Tool Based on Textual Differencing," Proc. Third Int'l Conf Reliability, Quality, and Safety of Software Intensive Systems, ENCRESS'97, May 1997.
- [43] L.J. White and H.K.N. Leung, "A Firewall Concept for Both Control-Flow and Data-Flow in Regression Integration Testing," Proc. Conf. Software Maintenance, pp. 262–270, Nov. 1992.
- [44] L.J. White, V. Narayanswamy, T. Friedman, M. Kirschenbaum, P. Piwowarski, and M. Oha, "Test Manager, A Regression Testing Tool," Proc. Conf. Software Maintenance, pp. 338–347, Sept. 1993.
- [45] S.S. Yau and Z. Kishimoto, "A Method for Revalidating Modified Programs in the Maintenance Phase," COMPSAC'87: Proc. 11th Ann. Int'l Computer Software and Applications Conf., pp. 272–277, Oct. 1987.

**Prof. Ananda Rao Akepogu** received B.Sc. (M.P.C) degree from Silver Jubilee Govt. College, SV Univer-sity, Andhra Pradesh, India. He received B.Tech. degree in Computer Science & Engineering and M.Tech. degree in A.I & Robotics from University of Hyderabad, Andhra Pradesh, India. He received Ph.D. from Indian Institute of Technology, Madras, India. He is Professor of Computer Science & Engineering and Principal of JNTU College of Engineering, Anantapur, India. Prof. Ananda Rao published more than fifty research papers in international journals, conferences and authored three books. His main research interest includes software engineering and data mining.

**Kiran Kumar J** is pursuing Ph.D. in Computer Science & Engineering from JNTUA, Anantapur, India and he received his M.Tech. in Computer Science & Engineering from the same university. He received B.E. degree in Computer Science & Engineering from Amaravati University, India. He has received the "Teradata Certified Master" certification from the Teradata. Currently he is working for IBM India Software Labs in the area of Software Testing since 2005. His main research interests include software engineering and Software Testing. He is a member of IEEE, ACM and IAENG.

# Normalized Distance Measure: A Measure for Evaluating MLIR Merging Mechanisms

Chetana Sidige<sup>1</sup>, Sujatha Pothula<sup>1</sup>, Raju Korra<sup>1</sup>, Madarapu Naresh Kumar<sup>1</sup>, Mukesh Kumar<sup>1</sup>

<sup>1</sup> Department of Computer Science, Pondicherry University  
Puducherry, 605014, India.

## Abstract

The Multilingual Information Retrieval System (MLIR) retrieves relevant information from multiple languages in response to a user query in a single source language. Effectiveness of any information retrieval system and Multilingual Information Retrieval System is measured using traditional metrics like Mean Average Precision (MAP), Average Distance Measure (ADM). Distributed MLIR system requires merging mechanism to obtain result from different languages. The ADM metric cannot differentiate effectiveness of the merging mechanisms. In first phase we propose a new metric Normalized Distance Measure (NDM) for measuring the effectiveness of an MLIR system. We present the characteristic differences between NDM, ADM and NDPM metrics. In the second phase shows how effectiveness of merging techniques can be observed by using Normalized Distance Measure (NDM). In first phase of experiments we show that NDM metric gives credits to MLIR systems that retrieve highly relevant multilingual documents. In the second phase of the experiments it is proved that NDM metric can show the effectiveness of merging techniques that cannot be shown by ADM metric.

**Keywords:** Average Distance Measure (ADM), Normalized Distance Measure (NDPM), Merging mechanisms, Multilingual Information Retrieval (MLIR).

## 1. Introduction

The Information Retrieval identifies the relevant documents in a document collection to an explicitly stated query. The goal of an IR system is to collect documents that are relevant to a query. Information retrieval uses retrieval models to get the similarity between the query and documents in form of score. Retrieval models are like binary retrieval model, vector space model, and probabilistic model.

Cross-language information retrieval (CLIR) search a set of documents written in one language for a query in another language. The retrieval models are performed between the translated query and each document. There are three main approaches to translation in CLIR: Machine translation, bilingual machine-readable dictionary, Parallel or comparable corpora-based methods.

Irrelevant documents are retrieved by information retrieval model when translations are performed with unnecessary terms. Thus translation disambiguation is desirable, so that relevant terms are selected from a set of translations. Sophisticated methods are explored in CLIR for maintain translation disambiguation part-of-speech (POS) tags, parallel corpus, co-occurrence statistics in the target corpus, the query expansion techniques. Problem called language barrier issues raised in CLIR systems [2].

Due to the internet explosion and the existence of several multicultural communities, users are facing multilingualism. User searches in multilingual document collection for a query expressed in a single language kind of systems are termed as MLIR system. First, the incoming question is translated into target languages and second, integrates information obtained from different languages into one single ranked list. Obtaining rank list in MLIR is more complicated than simple bilingual CLIR. The weight assigned to each document (RSV) is calculated not only according to the relevance of the document and the IR model used, but also the rest of monolingual corpus to which the document belongs is a determining factor.

Two types of multilingual information retrieval methods are query translation and document translation. As document translation causes more complications than query translation, our proposal is applying query translation. Centralized MLIR and distributed MLIR are two type architectures. Our proposed metric is applied on distributed MLIR. Distributed MLIR architecture has problems called merging the result lists. Merging techniques are like raw score, round robin. Performance of MLIR system differs due to merging methods. To measure the MLIR performance correctly we need to consider the MLIR features like translation (language barrier), merging methods. Our new metric is based on the concept of ADM metric. The drawbacks of the ADM metric are overcome in the proposed formula.



In this paper, Section 2 explains the related work of the proposed metric and merging methods. Section 3 explains the proposed metric in two phases. First phase explains newly proposed metric and second phase explains how the proposed metric is applied for merging methods of MLIR. Section 4 explains the experimental results and section 5 states conclusion.

## 2. Related work

There are two types of translation methods in MLIR - query translation and document translation [2]. Document translation can retrieve more accurate documents than query translation because the translation of long documents may be more accurate in preserving the semantic meaning than the translation of short queries. Query translation is a general and easy search strategy.

There are two architectures in MLIR [12]. In centralized architecture consists of a single document collection containing document collections and a huge index file. It needs one retrieving phase. Advantage of centralized architecture is it avoids merging problem. Problem with centralized architecture is the weights of index terms are over weighting. Thus, centralized architecture prefers small document collection. In distributed architecture, different language documents are indexed in different indexes and retrieved separately. Several ranked document lists are generated by each retrieving phase. Obtaining a ranked list that contains documents in different languages from several text collections is critical; this problem is solved by merging strategies. In any architecture problem called language translation issues are raised.

In a distributed architecture, it is necessary to obtain a single ranked document list by merging the individual ranked lists that are in different languages. This issue is known as merging strategy problem or collection fusion problem. Merging problem in MLIR is more complicated than the merging problem in monolingual environments because of the language barrier in different languages.

Following are some of the merging strategies.

**Round-robin merging strategy:** This approach is based on the idea that document scores are not comparable across the collections, each collection has approximately the same number of relevant documents and the distribution of relevant documents is similar across the result lists [11]. The documents are interleaved according to ranking obtained for each document.

**Raw score merging strategy:** This approach is based on the assumption that scores across different collections are comparable. Raw score sorts all results by their original similarity scores and then selects the top ranked documents. This method tends to work well when same methods are used to search documents [11].

**Normalized score merging:** This approach is based on the assumption that merging result lists are produced by diverse search engines. A simplest normalizing approach is to divide each score by the maximum score of the topic on the current list. After adjusting scores, all results are sorted by the normalized score [10], [11]. Another method is to divide difference between the score and maximum score by difference between maximum score and minimum score. This type of merging favours the scores which are near the best score of the topic on the list. This approach maps the scores of different result lists into the same range, from 0 to 1, and makes the scores more comparable. But it has a problem. If the maximum score is much higher than the second one in a result list, the normalized-score of document at rank 2 would be low even if its original score is high.

System evaluation is measured by calculating gap between system and user relevance. Due to Lack of control variables measuring the user centered approach is becoming difficult. The motivation of our proposal is performance measurement can be examined by the agreement or disagreement between the user and the system rankings.

New metric NDM is generated by considering the features of below IR metrics.

**Discount Cumulated Gain (DCG):** As rank gets increased the importance of document gets decreased.

**Normalized Distance-based Performance Measure (NDPM):** NDPM gives performance of MLIR system by comparing the order of ranking of two documents [1] [5]. NDPM is based on a preference relation  $>$  on a finite set of documents  $D$  is a weak order.

**Average Distance Measure (ADM):** [3] ADM measures the average distance between UREs (user relevance estimation) (the actual relevances of documents) and SREs (system relevance estimation) (their estimates by the IRS) [2]. Drawback of ADM metric is low ranked documents are given equal importance high ranked documents [3][1]. Problem with precision and recall is, they are highly sensitive to the thresholds. Instead of changing the relevance, retrieval values suddenly, there should be a continuous varying of relevance and retrieval.

### 3. Proposed metric

Normalized Distance Measure (NDM) is a new metric designed mainly for evaluating MLIR system. MLIR system has to access more information in an easier and faster way than monolingual systems. Distributed MLIR system has three steps translation, retrieval and merging. NDM considers ranking as a suitable measurement, because continuous rank performance measurement is better than non continuous groping and also the document score of one language cannot be compared to another language. Normalized Distance Measure measures the difference between the user's estimated ranked list and final MLIR ranked list. The NDM value ranges from 0 to 1. Final rank list of MLIR represented as  $R_{MLIR}$ . The ranked list obtained from user is represented as  $R_{USER}$ .

$$NDM = 1 - \frac{\sum_{i=0}^m \left| \frac{R_{MLIR(i)} - R_{USER(i)}}{R_{MLIR(i)} + \alpha} \right|}{\sum_{i=0}^m \left| \frac{R_{Threshold(i)} - R_{USER(i)}}{R_{Threshold(i)} + \alpha} \right|} \quad (1)$$

Where  $i = \{0, 1, 2, \dots, m\}$  where  $m$  is total number of documents.

In (1) equation, the term  $R_{MLIR(i)} + \alpha$  is total penalty calculated. ' $\alpha$ ' is included in (1) equation the penalty when an relevant document is not retrieved or when non relevant document is retrieved. Penalty  $R_{MLIR}$  measures the precision

Six cases are as follows.

Case (a):  $R_{MLIR(i)} = R_{USER(i)}$

Case (b):  $R_{MLIR(i)} > R_{USER(i)}$

Case (c):  $R_{MLIR(i)} < R_{USER(i)}$

Case (d):  $R_{MLIR(i)} = 0, R_{USER(i)} = 0$

Case (e):  $R_{MLIR(i)} \neq 0, R_{USER(i)} = 0$

Case (f):  $R_{MLIR(i)} = 0, R_{USER(i)} \neq 0$

First three cases consider a document as relevant by both MLIR system and USER. Last three cases a document is considered as not relevant by either MLIR system or by USER. In case (a), (d) difference between rankings is 0 as both ranks are same. In case (c), (f) difference between rankings is positive. This is represented on left bottom of the diagonal in table 1. In case (b), (e) difference between

rankings is negative. This is represented on top right of the diagonal in table 1.

Table 1. Calculation of Distance Between MLIR and USER Rank Systems In All Six Possibilities

	$R_{MLIR(i)}$	0	1	2	3	4	5
$R_{USER(i)}$	$R_{MLIR(i)} + \alpha$ $R_{USER(i)} + \alpha$	1	2	3	4	5	6
0	1	0	0.5	0.67	0.75	0.8	0.83
1	2	1	0	0.33	0.5	0.6	0.67
2	3	2	0.5	0	0.25	0.4	0.5
3	4	3	1	0.33	0	0.25	0.33
4	5	4	1.5	0.67	0.25	0	0.17
5	6	5	2	1	0.5	0.25	0

We can estimate the good MLIR System by using the user estimated values but estimating a worst MLIR is not possible because worseness of MLIR system increases as the irrelevant documents are increased. Thus we are using threshold MLIR as a least bad case MLIR system. The denominator measures the difference between the resulted ranked lists and threshold MLIR system. The numerator measures the difference between the MLIR ranked list and ranked list estimated by user.

Table 2 shows the different characteristics of ADM, NDPM and NDM. In Table 2, the characteristic called "document score" is not needed for user. User is concerned only about ordering and ranking of the document list. NDM gives different importance for first and last documents. other characteristics shows the reasons, why NDM metric is performing better than other metrics.

Table 2: Characteristics of NDM, ADM, NDPM

Characteristics	ADM	NDPM	NDM
Rank	No	No	Yes
Order	No	Yes	Yes
Document score	Yes	No	No
Considers irrelevant document	Yes	No	Yes
Equal Importance for first and last documents	Yes	Yes	No

#### 4. Experimental results

Phase 1 experiments show the importance of NDM metric. Effectiveness of an Information Retrieval System (IRS) depends on relevance and retrieval. [2] States that precision and recall are highly sensitive to the thresholds chosen.

Table 3: Document scores in six MLIR systems

	<i>D1</i>	<i>D2</i>	<i>D3</i>	<i>D4</i>	<i>D5</i>
<i>USER</i>	0.9	0.8	0.7	0.6	0.5
<i>MLIR1</i>	<b>0.8</b>	<b>0.7</b>	<b>0.6</b>	<b>0.5</b>	0.9
<i>MLIR2</i>	0.9	<b>0.7</b>	<b>0.6</b>	<b>0.8</b>	0.5
<i>MLIR3</i>	0.9	<b>0.6</b>	<b>0.8</b>	<b>0.7</b>	0.5
<i>MLIR4</i>	0.9	0.8	0.7	<b>0.5</b>	<b>0.6</b>
<i>MLIR5</i>	<b>0.8</b>	<b>0.9</b>	0.7	0.6	0.5
<i>MLIR6</i>	0.9	<b>0.7</b>	<b>0.8</b>	<b>0.5</b>	<b>0.6</b>

Precision and recall are not continuous therefore precision and recall are not sensitive to important changes to MLIR systems like giving importance to top relevant documents. ADM and NDPM metrics are continuous metrics. Thus we are comparing the NDM metric with ADM and NDPM.

Table 4: Compare NDM with ADM and NDPM

	<i>ADM</i>	<i>NDPM</i>	<i>NDM</i>
<i>MLIR1</i>	0.84	0.60	0.647
<i>MLIR2</i>	0.92	0.80	0.863
<i>MLIR3</i>	0.92	0.80	0.885
<i>MLIR4</i>	0.96	0.90	0.9507
<i>MLIR5</i>	0.96	0.90	0.9554
<i>MLIR6</i>	0.92	0.80	0.987

Table 3 represents the six MLIR system's score list. The scores of the document are converted into rankings to obtain NDM and NDPM metrics. The drawbacks of the ADM are stated in [3]. The drawbacks of ADM are corrected in NDM. [3] states the importance of ranking in performance measurement. Table 4 compares NDM metric with ADM and NDPM.

We ordered 6 MLIR systems in Table 3 in such a way that the bottom MLIR system performance is better than the top MLIR systems. In Table 4 the ADM and NDPM values of the 6<sup>th</sup> MLIR system is low even though its performance is better than 4<sup>th</sup> and 5<sup>th</sup> MLIR system. Distribution of relevant documents is slightly different in MLIR3 and MLIR4, so NDM values are slightly different but ADM and NDPM shows no difference in performance. In MLIR2 and MLIR3 2<sup>nd</sup>, 3<sup>rd</sup>, 4<sup>th</sup> documents are interchanged among themselves. MLIR1 gives bad performance because the 1<sup>st</sup> top document is placed at last position. Figure 1 represents the table 4.

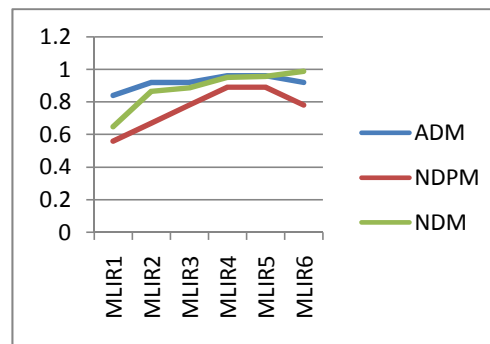


Fig. 1 The performance of NDM is compared with the ADM and NDPM

In the second phase of our experiments, we have measured the NDM values for four merging technique of a MLIR system. ADM value for the above MLIR system is 0.68 which remains constant for all 4 merging techniques. To obtain the performance of merging mechanisms of an MLIR we use NDM metric as follows. We took 9 documents from 3 languages and assigned document scores for 9 documents as shown in Table 5.

Table 5: Scores of 9 documents in three languages

Language 1	Language 2	Language 3
1.9	0.4	1.2
1.62	0.2	0.9
1.4		0.6
0.8		

We performed merging techniques for the above MLIR and the documents order is shown in the table 6. The ADM and NDM values for four merging mechanisms are shown in the Table 7.

Table 6: Rank lists of merging techniques

rank	Round robin	Raw score	Normalize with max(RSV)	Normalize with max(RSV) and min(RSV)
1	1.9	1.9	1	2
2	0.4	1.62	1	2
3	1.2	1.4	1	1.72
4	1.62	1.2	0.8	1.5
5	0.2	0.9	0.75	1.4
6	0.9	0.8	0.73	1.2
7	1.4	0.6	0.5	1
8	0.6	0.4	0.5	1
9	0.8	0.2	0.421	1.72

Table 7: NDM measure for 9 documents in three languages.

	ADM	NDM
Round Robin Merging	0.68	0.88
Raw Score Merging	0.68	0.84
Normalized score merging with max (RSV)	0.68	0.95
Normalized score merging with max (RSV) and min (RSV)	0.68	0.85

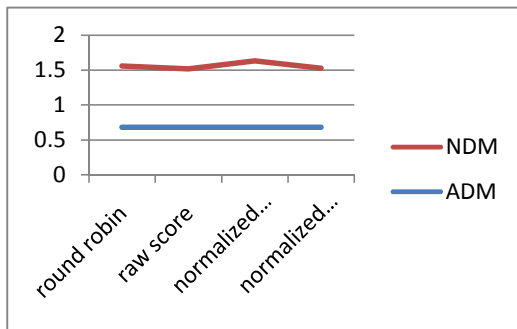


Fig 2: Graphical representation of table 7.

Fig 2 shows the variation of NDM metric for merging techniques, where ADM shows no difference. Characteristics of the NDM, ADM, NDPM shows that NDM considered many features.

## 5. Conclusions

This paper shows two phased experiment where first phase proposes a new metric for MLIR based on rank schema. It is shown that the new metric is better than old metrics like ADM and NDPM metrics. Characteristics that differentiate three metrics ADM, NDPM and NDM are tabularized. In the first phase we stated the benefits of NDM over ADM and NDPM in form of characteristics

and experiments. In the second phase NDM metric evaluates the performance of MLIR system when four different types of merging techniques are used.

## References

- [1] Bing Zhou, Yiyu Yao, "Evaluating Information Retrieval System Performance Based on User Preference", Journal of intelligent information systems, Springerlink, vol. 34, issue 3, pp. 227-248, June. 2010.
- [2] Kazuaki Kishida, "Technical issues of cross-language information retrieval: a review", Information Processing and Management international journal, science direct, vol. 41, issue 3, pp. 433-455, may. 2005.
- [3] Stefano Mizzaro, S, "A New Measure of Retrieval Effectiveness (Or: What's Wrong with Precision and Recalls)," In: International Workshop on Information Retrieval, pp. 43-52 .2001
- [4] Järvelin, K., Kekäläinen, J.: Cumulated Gain-based Evaluation of IR Techniques. ACM Transactions on Information Systems, vol. 20, Issue 4, 422-446 October (2002)
- [5] Yao, Y. Y. (1995). Measuring retrieval effectiveness based on user preference of documents. Journal of the American Society for Information Science, Volume 46 Issue 2, 133-145, March 1995.
- [6] W. C. LIN and H. H. CHEN, "Merging results by using predicted retrieval effectiveness," Lecture notes in computer science, pages 202-209, 2004.
- [7] Savoy, "Combining multiple strategies for effective monolingual and cross-lingual retrieval," IR Journal, 7(1-2):121-148, 2004.
- [8] Lin, W.C. & Chen, H.H. (2002b). Merging Mechanisms in Multilingual Information Retrieval. In Peters, C. (Ed.), Working Notes for the CLEF 2002 Workshop, (pp. 97-102).
- [9] Rita M. Aceves-Pérez, Manuel Montes-y-Gómez, Luis Villaseñor-Pineda, Alfonso Ureña-López. Two Approaches for Multilingual Question Answering: Merging Passages vs. Merging Answers International Journal of Computational Linguistics and Chinese Language Processing. Vol. 13, No. 1, pp 27-40, March 2008.
- [10] F. Martínez-Santiago, M. Martín, and L.A. Ureña. SINAI at CLEF 2002: Experiments with merging strategies. In Carol Peters, editor, Proceedings of the CLEF 2002 Cross-Language Text Retrieval System Evaluation Campaign. Lecture Notes in Computer Science, pages 103-110, 2002.
- [11] E. Airio, H. Keskustalo, T. Hedlund and A. Pirkola, Multilingual Experiments of UTA at CLEF2003 - the Impact of Different Merging Strategies and Word Normalizing Tools. CLEF 2003, Trondheim, Norway, 21-22 August 2003.
- [12] Wen-Cheng Lin and Hsin-Hsi Chen (2003). Merging Mechanisms in Multilingual Information Retrieval. In Advances in Cross-Language Information Retrieval: Third Workshop of the Cross-Language Evaluation Forum, CLEF 2002, Lecture Notes in Computer Science, LNCS 2785, September 19-20, 2002, Rome, Italy, pp. 175-186.
- [13] Anne Le Calvé, Jacques Savoy, "Database merging strategy based on logistic regression," Information Processing and Management: an International Journal, vol.36, p.341-359, May. 2000.



Chetana Sidige is presently pursuing M.Tech (Final year) in Computer Science of Engineering at Pondicherry University. She did her B.Tech in Computer Science and Information Technology from G. Pulla Reddy Engineering College, Sri Krishnadevaraya University. Currently the author is working on Multilingual Information retrieval evaluation.



Mukesh Kumar received his Bachelor of Technology degree in Computer Science and Engineering from Uttar Pradesh Technical University Lucknow, India in 2009. He is currently pursuing his master's degree in Network and Internet Engineering in the School of Engineering and Technology, Department of Computer Science, Pondicherry University, India. His research interests include Denial-of Service resilient protocol design, Cloud Computing and Peer to Peer Networks.



Pothula Sujatha is currently working as Assistant Professor and pursuing her PhD in Department of Computer science from Pondicherry University, India. She completed her Master of Technology in Computer Science and Engineering from Pondicherry University and completed her Bachelor of Technology in Computer Science and Engineering from

Pondicherry Engineering College, Pondicherry. Her research interest includes Modern Operating Systems, Multimedia Databases, Software Metrics and Information Retrieval. Her PhD research is on performance Evaluation of MLIR systems.



Raju Korra is presently pursuing Master of Technology in Computer Science and Engineering from Pondicherry University, India. He has completed his Bachelor of Technology in Computer Science and Engineering from Kakatiya University, Warangal. His research interest includes Genetic Algorithms, Software metrics, Data Mining,

Information Retrieval and MLIR. Currently he is working on metrics for evaluating MLIR systems.



Madarapu Naresh Kumar is presently pursuing Master of Technology in Computer Science with specialization in Network and Internet Engineering from Pondicherry University, India. He has completed his Bachelor of Technology in Computer Science and Engineering from JNTU Hyderabad. His research interest includes Cloud Computing, Web Services, Software Metrics, SOA and Information

Retrieval. Currently he is working on security issues in Cloud Computing.



# Brain Extraction and Fuzzy Tissue Segmentation in Cerebral 2D T1-Weighted Magnetic Resonance Images

Bouchaib CHERRADI<sup>1</sup>, Omar BOUATTANE<sup>2</sup>, Mohamed YOUSSEFI<sup>3</sup> and Abdelhadi RAIHANI<sup>4</sup>

<sup>1</sup> Faculty of Science and Technology, UFR: MCM&SCP, Hassan II University, Mohammedia, Morocco,

<sup>2</sup> E.N.S.E.T Institute, Department of Informatics, Hassan II University, Mohammedia, Morocco,

<sup>3</sup> Faculty of Science, Department of Information Processing, Mohammed V University, Rabat, Morocco.

<sup>4</sup> Faculty of Science, Department of Information Processing, Hassan II University, Mohammedia, Morocco.

## Abstract

In medical imaging, accurate segmentation of brain MR images is of interest for many brain manipulations. In this paper, we present a method for brain Extraction and tissues classification. An application of this method to the segmentation of simulated MRI cerebral images in three clusters will be made. The studied method is composed with different stages, first Brain Extraction from T1-weighted 2D MRI slices (TMBE) is performed as pre-processing procedure, then Histogram based centroids initialization is done, and finally the fuzzy c-means clustering algorithm is applied on the results to segment the image in three clusters. The introduction of this pre-processing procedure has been made in the goal to have a targeted segmentation method. The convergence speed for tissues classification has been considerably improved by avoiding a random initialization of the cluster centres and reduction of the volume of data processing.

**Keywords:** Clustering, Fuzzy c-means, histogram analysis, Brain Extraction, Image segmentation.

## 1. Introduction

Image segmentation is a key step toward image analysis and serves in the variety of applications including pattern recognition, object detection, and medical imaging [1], which is also regarded as one of the central challenges in image processing and computer vision. The task of image segmentation can be stated as the partition of an image into different meaningful regions with homogeneous characteristics using discontinuities or similarities of the image such as intensity, color, tone or texture, and so on [2]. Numerous techniques have been developed for image segmentation

and a tremendous amount of thorough research has been reported in the literatures [3–5]. According to these references, the image segmentation approaches can be divided into four categories: thresholding, clustering, edge detection and region extraction. In this paper, a clustering based method for image segmentation will be considered. Many clustering strategies have been used, such as the crisp clustering scheme and the fuzzy clustering scheme, each of which has its own special characteristics [6]. The conventional crisp clustering method restricts each point of the data set to exclusively just one cluster. However, in many real situations, for images, issues such as limited spatial resolution, poor contrast, overlapping intensities, noise and intensity inhomogeneities variation make this hard (crisp) segmentation a difficult task. Thanks to the fuzzy set theory [7], which involves the idea of partial membership described by a membership function, fuzzy clustering as a soft segmentation method has been widely studied and successfully applied to image segmentation [9, 10]. Among the fuzzy clustering methods, fuzzy c-means (FCM) algorithm [8] is the most popular method used in image segmentation because it has robust characteristics for ambiguity and can retain much more information than hard segmentation methods. Although the conventional FCM algorithm works well on most noise-free images, it has a serious limitation: it does not incorporate any information about spatial context, which cause it to be sensitive to noise and imaging artefacts. To compensate for this drawback of FCM, we have proposed in [11] the introduction of spatial information as decision by focusing on the neighbourhood (DFN) for the pixels not having a strong degree of membership after the fuzzy partition.

Intracranial segmentation commonly referred to as brain extraction or skull stripping, aims to segment the brain tissue (cortex and cerebellum) from the skull and non-brain intracranial tissues in magnetic resonance (MR) images of the brain. Brain extraction is an important pre-processing step in neuroimaging analyses because brain images must typically be skull stripped before other processing algorithms such as registration, or tissue classification can be applied. In practice, brain extraction is widely used in neuroimaging analyses such as multi-modality image fusion and inter-subject image comparisons [12]; examination of the progression of brain disorders such as Alzheimer's Disease, multiple sclerosis and schizophrenia, monitoring the development or aging of the brain; and creating probabilistic atlases from large groups of subjects. Numerous automated skull-stripping methods have been proposed [13-18]. The rest of this paper is organised as follows: in the next section we describe our proposed method for Brain Extraction from 2D MRI slices as pre-processing procedure; in section 3 the standard clustering fuzzy c-means algorithm is sketched. Histogram based centroids initialization is presented in section 4. The global proposed method of segmentation is presented in section 5. In section 6 we present different results obtained with this method. Final conclusions and future works are discussed in section 7.

## 2. Pre-processing.

### 2.1. Filtering.

This pre-processing stage performs a non linear mapping of the grey level dynamics for the image. This transform consists in the application of a 3x3 median filter. The use of median filtering derives from the nature of the noise distribution in the MR images. The main source of noise in this kind of images is due to small density variations inside a single tissue which tend to locally modify the RF emission of the atomic nuclei during the imaging process.

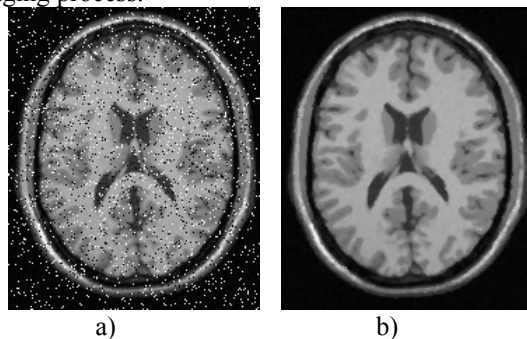


Figure 1: a) T1 MRI image with salt and paper noise, b) Median filtered image.

### 2.2. Brain Extraction: Threshold Morphologic Brain Extraction (TMBE).

The goal of this phase is to extract the brain from the acquired image: this will allow us to simplify the segmentation of the brain tissues. Our easy and effective method can be divided in five steps:

#### 2.2.1 Thresholding.

This step is based on global binary image thresholding using Otsu's method [19]. Figure 2-b shows a result of this operation.

#### 2.2.2 Greatest Connected Component Extraction.

A survey based on a statistical analysis of the existing connected components on the dilated image, permits to extract the region whose area is the biggest. Figure 2-c shows a result of this operation.

#### 2.2.3 Filling the holes.

The remaining holes in the binary image obtained in step 2, containing the greatest connected component, are filled using morphologic operation consisting of filling holes in the binary image. A hole is a set of background pixels within connected component. The result of this operation is shown in figure 2-d.

#### 2.2.4 Dilatation.

This morphologic operation consists of eliminating all remaining black spots on the white surface of the image. These spots are covered by the dilatation of the white parts. This carried out by moving a square structuring element of size (SxS) on binary image and applying logical OR operator on each of the (S<sup>2</sup>-1) neighbouring pixels (figure 2-e). In this paper we consider S=3.

#### 2.2.5 ROI Extracting.

The region of interest is the brain tissues. To extract this region we use the AND operator between the original filtered image and the binary mask obtained in last step. The non-brain region is obtained by applying AND operator between the image in figure 2-a and the logical complement of mask image in figure 2-e.

The figure 2-f shows the region of interest corresponding to the effective brain tissues in original MRI. The figure 2-g presents the non brain region.

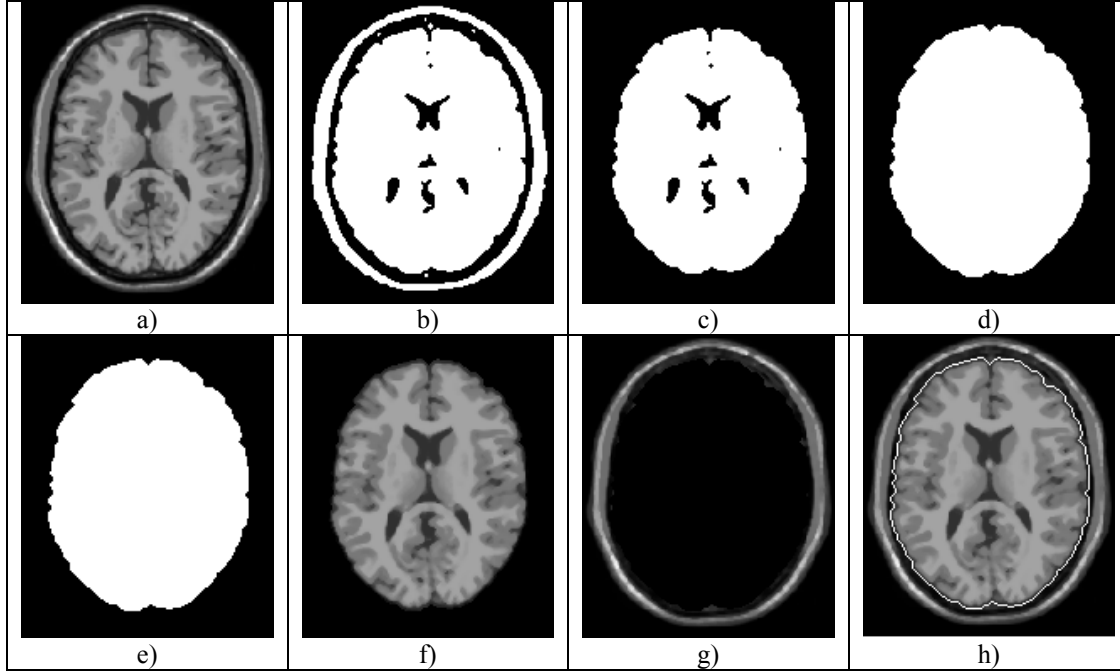


Figure 2: Brain Extraction steps on axial slice of number 84/181 in simulated data volume [21] with 5% uniform noise.

### 3. Standard FCM algorithm.

The fuzzy  $c$ -means (FCM) clustering algorithm was first introduced by DUNN [20] and later was extended by BEZDEK [8]. Fuzzy C-means (FCM) is a clustering technique that employs fuzzy partitioning such that a data point can belong to all classes with different membership grades between 0 and 1.

The aim of FCM is to find  $C$  cluster centers (centroids) in the data set  $X = \{x_1, x_2, \dots, x_N\} \subseteq R^p$  that minimize the following dissimilarity function:

$$J_{FCM} = \sum_{i=1}^c J_i = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m d^2(v_i, x_j) \quad (1)$$

$u_{ij}$ : Membership of data  $x_j$  in the cluster  $V_i$ ;

$V_i$ : Centroid of cluster  $i$ ;

$d_{(V_i, x_j)}$ : Euclidian distance between  $i^{\text{th}}$  centroid ( $V_i$ ) and  $j^{\text{th}}$  data point  $x_j$ ;

$m \in [1, \infty]$ : Fuzzy weighting exponent (generally equals 2).

$N$ : Number of data.

$C$ : Number of clusters  $2 \leq C < N$ .

$p$ : Number of features in each data.

With the constraints:

$$u_{ij} \in [0, 1], \forall i, j \quad (2a)$$

$$\sum_{i=1}^c u_{ij} = 1, \forall j = 1, \dots, N \quad (2b)$$

$$0 < \sum_{j=1}^N u_{ij} < N, \forall i = 1, \dots, C \quad (2c)$$

To reach a minimum of dissimilarity function there are two conditions.

$$V_i = \frac{\sum_{j=1}^N u_{ij}^m x_j}{\sum_{j=1}^N u_{ij}^m} \quad (3)$$

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left( \frac{d_{ij}}{d_{kj}} \right)^{2/(m-1)}} \quad (4)$$

This iterative algorithm is in the following steps.

**Step 0.** Randomly initialize the membership matrix ( $U$ ) according to the constraints of Equations 2a, 2b and 2c, Choose fuzzification parameter  $m$   $1 < m < \infty$ , Choose the number of clusters  $C$ , Choose the initial values of cluster centers  $V^{(0)}$  and *threshold*  $\epsilon > 0$ .

**At iteration  $N_i$**

{

**Step 1.** Calculate centroids vector ( $V_{N_i}$ ) using Equation (3).

**Step 2.** Compute dissimilarity function  $J_{N_i}$  using equation (1). If its improvement over previous iteration is below a *threshold*  $\epsilon$ , Go to Step 4.

**Step 3.** Compute a new membership matrix ( $U_{N_i}$ ) using Equation (4). Go to Step 1.

**Step 4. Stop.**  
 }

#### 4. Centroids initialization.

Clustering algorithms requires an initialisation of the clusters centres. Usually, this is randomly made. However, an adequate selection permits to improve the accuracy and reduces considerably the number of required iterations to the convergence of these algorithms.

The choice of the class number and initial correspondent centroids can be supervised or unsupervised. The supervised method consists in imposing a number and initial value of clusters according to the quantity of information that we want to extract from the image. The unsupervised method is based on the estimation of the number of clusters and initial cluster value in the image. Among the methods used in this domain we consider the histogram information analysis. This strategy consists in 4 stages:

Stage1: Histogram definition.

For image size  $S \times T$ , at point  $(s, t)$ ,  $f(s, t)$  is the gray value with  $0 \leq s \leq (S-1)$ ,  $0 \leq t \leq (T-1)$ . Let  $H(g)$  denote the number of pixels having gray level  $g$ . Therefore, the histogram function can be written as:

$$H(g) = \sum_{s=0}^{S-1} \sum_{t=0}^{T-1} \delta(f(s, t) - g) \quad (5)$$

where  $g \in G$ ,  $\delta(g=0) = 1$  and  $\delta(g \neq 0) = 0$ .

Stage 2: Histogram smoothing must be done to eliminate the parasitic peaks.

Stage 3: Detecting all local peaks.

Local peak at position  $g$  satisfy the condition  $H(g-1) < H(g)$  and  $H(g) > H(g+1)$ .

Stage 4: Eliminating weak peaks.

Among the detected peaks, there are some ones with weak height, they represent small non significant regions, and to eliminate these peaks we introduce adapted minimal amplitude  $Am$ .

The number of remaining peaks is the initial number of clusters  $C$  and correspondent's gray levels  $g_i$  are the initial centroids  $V^{(0)}$  for the clustering algorithm.

#### 5. Proposed method.

The proposed segmentation method is summarized as follows:

**Inputs** : MRI Gray level image  $I$  (size  $S \times T=N$ ), minimal amplitude  $Am$ , fuzzification parameter  $m$  ( $1 < m < \infty$ ) and Threshold  $\varepsilon > 0$ .

**Outputs:** Number of clusters  $C$ , Centroids of clusters vector  $V$ , correspondent fuzzy partition matrix  $U$  and segmented image  $Iseg$ .

**Pre-processing:**

*Step 1.* Noise removing: Median filter.

*Step 2.* Brain Extraction procedure (TMBE) (See section 2.2)

*Step 3.* Histogram computing for brain tissues using (5).

*Step 4.* Histogram Smoothing with appropriate 1D Gaussian filter.

*Step 5.* Detect all local peaks of the histogram.

*Step 6.* Eliminate weak peaks. The peaks whose the amplitude is  $< Am$  are eliminated.

The number of remaining peaks is  $C$  and correspondent gray levels are the initial centroids vector  $V^{(0)}$ .

**Fuzzy Clustering:**

At iteration  $N_i$  do

{

*Step 7.* Compute the membership function ( $U_{N_i}$ ) using (4).

*Step 8.* Compute the cluster centroids vector  $V^{(N_i)}$  using (3).

*Step 9.* Compute objective function  $J_{(N_i)}$  using (1).

*Step 10.* If  $abs(J_{(N_i)} - J_{(N_{i-1})}) < \varepsilon$ , go to step 11, otherwise, go to step 7.

}

**Region Labelling:**

*Step 11:* Defuzzification: Convert the final membership matrix  $U$  to crisp one using maximum procedure.

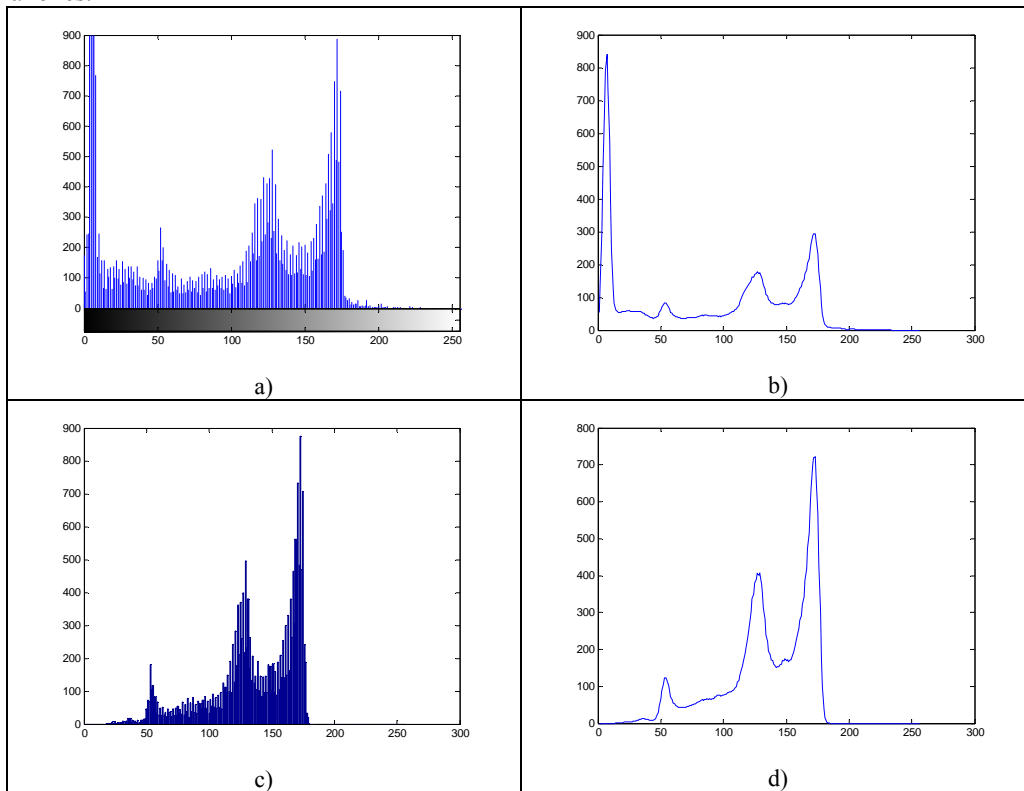
*Step 12:* Region labelling procedure to obtain  $Iseg$ .

#### 6. Results and discussion.

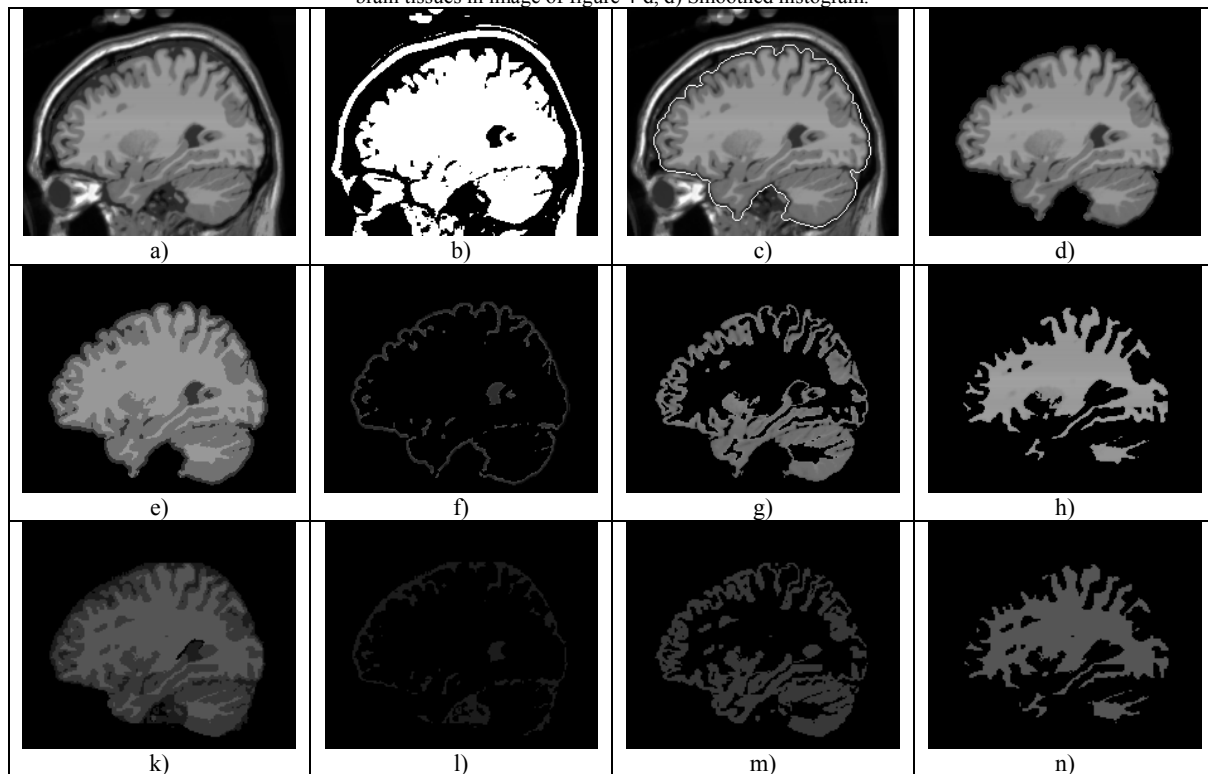
The method was implemented in MATLAB 7.8, validated on synthesized images, and then several simulated cerebral MRI images of different modalities (T1, T2 and PD) from the classical simulated brain database of McGill University [21] have been experimented.

The proposed method for brain extraction (TMBE) was tested separately on different magnetic resonance images

of different modalities of acquisition especially on healthy cerebral ones.



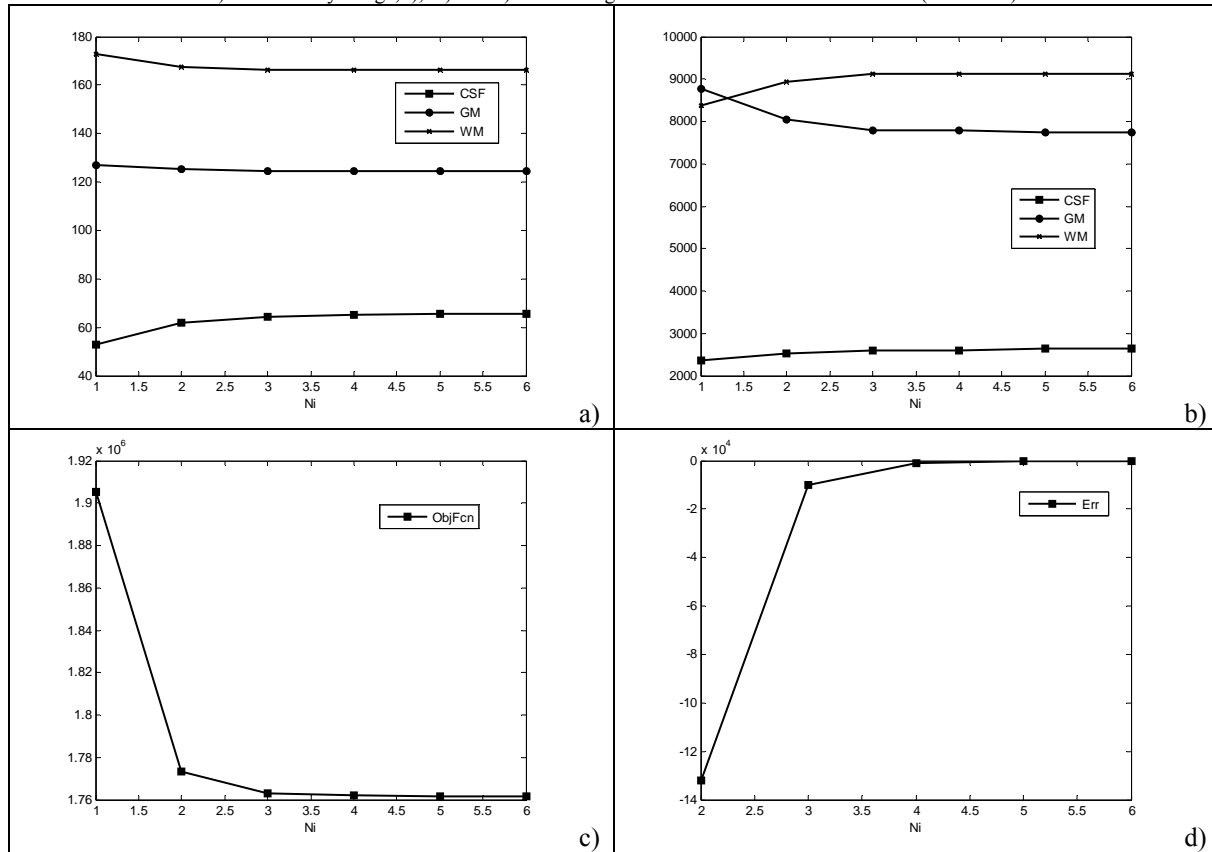
**Figure 3:** Histogram analysis for centroids initialization, a) Histogram of the image in figure 4-a, b) Smoothed histogram, c) Histogram of the extracted brain tissues in image of figure 4-d, d) Smoothed histogram.



**Figure 4:** Example of segmentation results. a)-d) Results of brain extraction proposed method.



e) Segmented image by the proposed method, f) Cerebrospinal fluid (CSF), g) Gray matter (GM) and h) White matter (WM).  
 k) Truth Verity image, l, m) and n) Manual segmentation of the same brain tissues (Brainweb).



**Figure 5:** Dynamic of different clustering parameters for image in figure 4-d.

a) Centroids starting from (C1: CSF, C2: GM, C3: WM) = (53, 127, 173) as results of histogram analysis, b) Cardinality of each tissue, c) Values of objective function  $J(N_i)$ , d) Values of  $Err$  ( $J(N_i) - J(N_i - 1)$ ).

The effectiveness of the method was tested on simulated MR images to extract the well known clusters (truth verity). Figure .3 shows the results of histogram analysis leading to a centroids initialisation of the extracted region of interest consisting of brain tissues that we want segment. It is about a sagittal T1-weighted slice number 120/181 in sagittal direction of TALAYRACH stereotaxic reference (volume of  $181 \times 217 \times 181$  voxels [21]).

Figure .4 shows an example of qualitative evaluation of our segmentation results with the provided manually segmentation results by the web site [21] corresponding to the same slice described above.

The segmentation aims to divide the image in three clusters: White matter (WM), gray matter (GM), and cerebrospinal fluid (CSF). The background pixels are removed from the image by thresholding (binarisation) before the clustering starts.

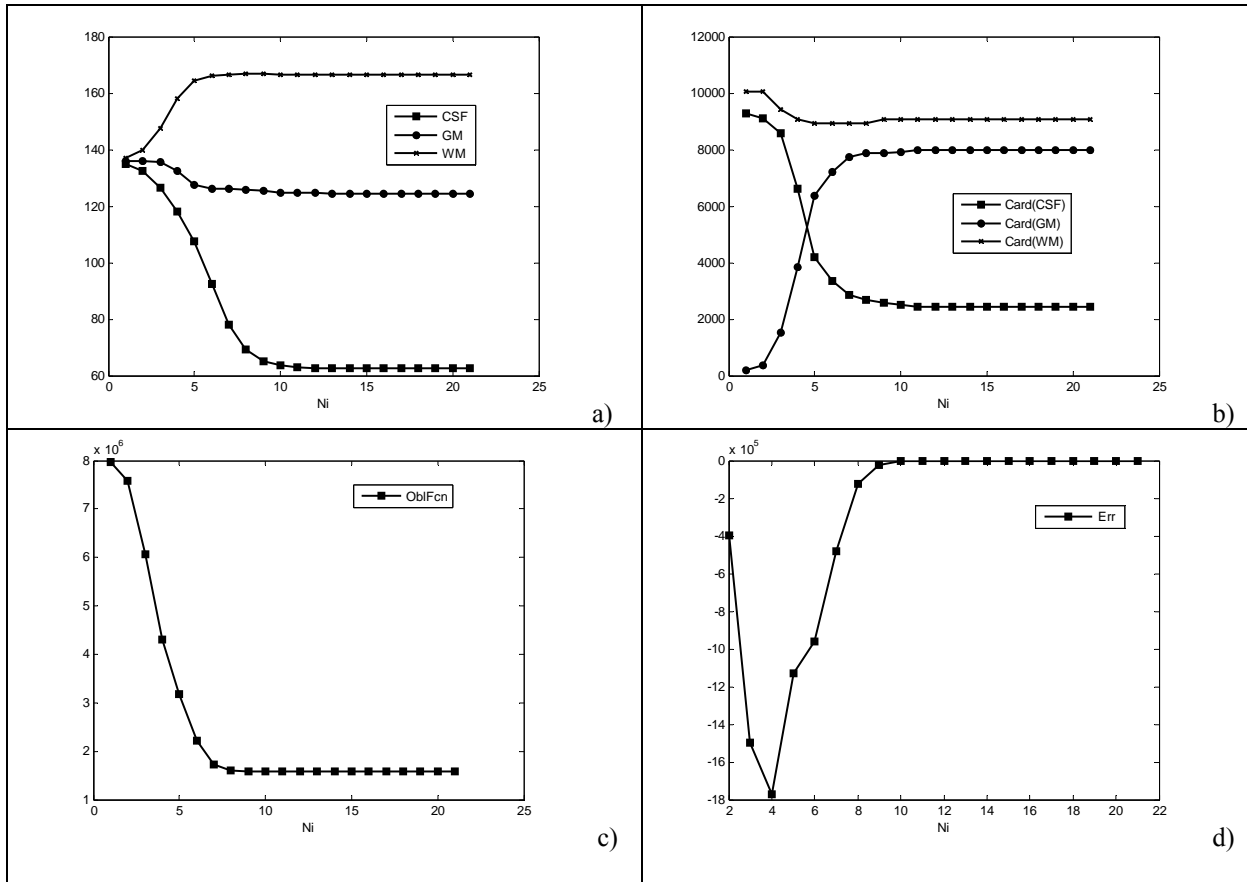
T1-weighted modality, that belong to the fastest MRI modalities available, are often preferred, since they offer

a good contrast between gray (GM) and white cerebral matter (WM) as well as between GM and cerebrospinal fluid (CSF). The advantages of using digital simulated images rather than real image data for validating segmentation methods is that it include prior knowledge of the true tissues types.

Comparison between figure .5 and figure .6 shows the effectiveness of the proposed method. Indeed, in figure. 5 we show that when we give adequate initially centroids, the iteratively clustering algorithm converge rapidly toward the effective clusters in the image (65, 124, and 166) in approximately about 6 iterations and with simple pace of the curves, but when the initialisation is made so far from the adequate values of the desired clusters, the convergence of the clustering algorithm is very slow, as is shown in figure 6 (approximately about 21 iterations), what gives a gain of approximately **70%** in time processing. In addition, the

pace of the curves present some anomalies that we have explain.

The accuracy of the proposed method for brain extraction is demonstrated with several MRI images in different modalities of acquisition, but its robustness in images of T1 modality is very remarkable, the figure 4 (a-d) shows an example.



**Figure 6:** Dynamic of different clustering parameters for image in figure 4-d.

a) Centroids starting from (C1:CSF, C2:GM, C3:WM) = (135.5, 136, 136.5) as manual initialization, b) Cardinality of each tissue, c) Values of objective function  $J(Ni)$ , d) Values of  $Err$  ( $J(Ni)-J(Ni-1)$ ).

## 7. Conclusion and perspectives.

In this paper, we have presented a complete image classification method. This method was applied to the segmentation of the MRI images. The use of the histogram analysis instead of a random initialization leads to an important improvement in the choice of the centers of classes (70%).

Unlike other brain segmentation methods described in the literature, the one described in this dissertation is truly automatic because it does not require a user to determine image-specific parameters, thresholds, or regions of interest.

The automatic proposed method for extracting the brain from the T1-weighted MRI head scans is based on a hybrid processing techniques including the adaptive thresholding and morphology mathematical operators.

Qualitative evaluation of the obtained results for the proposed brain extraction method show that the proposed method achieves important performance with synthetic Brainweb data, however it will be experimented with real database and quantitatively

evaluated and compared with the well known brain extraction techniques in the literature.

In perspective we will also study and characterise the comportment of centroids dynamic, it will follows a mathematical function. In addition we will explain the comportment of the *Err* curve that is observed in many essays when the clusters initialization is not adequately made (figure 6-d).

The robustness of the method up on the different artefacts usually present in magnetic resonance images such as noise and intensity inhomogeneity will be evaluated in future work. In other hand we are extending this method for 3D brain MRI and comparing it with some well known similar ones trough performance measure.

## References

- [1] J. Kim, J.W. Fisher, A. Yezzi, M.Cetin, A.S. Willsky, "A nonparametric statistical method for image segmentation using information theory and curve evolution", IEEE Transactions on Image Processing **14**(10), 2005, pp. 1486–1502.

- [2] G. Dong, M. Xie, "Color clustering and learning for image segmentation based on neural networks", *IEEE Transactions on Neural Networks* **16**(4), 2005, pp. 925–936.
- [3] R.M. Haralick, L.G. Shapiro, "Image segmentation techniques", *Computer Vision, Graphics and Image Processing* **29**(1), 1985, pp. 100–132.
- [4] N.R. Pal, S.K. Pal, "A review on image segmentation techniques", *Pattern Recognition* **26**(9), 1993, pp. 1277–1294.
- [5] D.L. Pham, C. Xu, J.L. PRINCE, "Current methods in medical image segmentation", *Annual Review of Biomedical Engineering* **2**(1), 2000, pp. 315–338.
- [6] WEINA WANG, YUNJIE ZHANG, YI LI, XIAONA ZHANG, "The global fuzzy c-means clustering algorithm", In *Proceedings of the World Congress on Intelligent Control and Automation*, Vol. 1, 2006, pp. 3604–3607.
- [7] L.A. Zadeh, "Fuzzy sets" *Information and Control*, Vol. 8, 1965, pp. 338–353.
- [8] J.C. Bezdek, "Pattern Recognition with Fuzzy Objective Function Algorithms", Plenum Press, New York 1981.
- [9] J.C. Bezdek, L.O. Hall, L.P. Clarke, "Review of MR image segmentation techniques using pattern recognition", *Medical Physics* **20**(4), 1993, pp. 1033–1048.
- [10] N. Ferahta, A. Moussaoui, K. Benmahammed, V.Chen, "New fuzzy clustering algorithm applied to RMN image segmentation", *International Journal of Soft Computing* **1**(2), 2006, pp. 137–142.
- [11] B.Cherradi and O.Bouattane. "Fast fuzzy segmentation method of noisy MRI images including special information". In the proceeding of *ICTIS'07 IEEE Morocco section*, ISBN 9954-8577-0-2, Fez, 3-5 Mars 2007, p 461-464.
- [12] R. P. Woods, M. Dapretto, N. L. Sicotte, A. W. Toga, and J. C. Mazziotta, "Creation and use of a Talairach-Compatible atlas for accurate, automated, nonlinear intersubject registration, and analysis of functional imaging data", *Human Brain Mapping*, vol. 8, pp: 73-79, 1999.
- [13] AM. Dale, B. Fischl, and MI. Sereno, "Cortical surface-based analysis. Segmentation and surface reconstruction", *NeuroImage*, vol. 9, pp: 179-194, 1999.
- [14] H. Hahn, and H-O. Peitgen, "The skull stripping problem in MRI solved by a single 3D watershed transform", *MICCAI*, vol. 1935, pp: 134-143, 2000.
- [15] S. Sandor, and R. Leahy, "Surface-based labeling of cortical anatomy using a deformable database", *IEEE Transactions on Medical Imaging*, vol. 16, pp: 41-54, 1997.
- [16] F. Segonne, A. M. Dale, E. Busa, M. Glessner, D. Salat, H. K. Hahn, and B. Fischl, "A hybrid approach to the skull stripping problem in MRI", *NeuroImage*, vol. 22, pp :1060-75, 2004.
- [17] S. M. Smith, "Fast robust automated brain extraction", *Human Brain Mapping*, vol. 17, pp: 143-55, 2002.
- [18] D.W. Shattuck, S.R. Sandor-Leahy, K.A. Shaper, D.A. Rottenberg, R.M. Leahy, "Magnetic resonance image tissue classification using a partial volume model". *NeuroImage*. 13 (5), 856–876. 2001.
- [19] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms". *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 9, No. 1, 1979, pp. 62-66.
- [20] J.C. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters", *Journal of Cybernetics* **3**(3), 1973, pp. 32–57.
- [21] <http://www.bic.mni.mcgill.ca/brainweb>

**Bouchaib CHERRADI** has born in 1970 at El JADIDA, Morocco. Received the B.S. degree in Electronics in 1990 and the M.S. degree in Applied Electronics in 1994 from the ENSET Institute, Mohammedia, Morocco. He received the DESA diploma in Instrumentation of Measure and Control from the University of EL JADIDA in 2004. He is now a Ph.D. student in MCM&SCP laboratory, Faculty of Science and Technology, Mohammedia. His research interests include Massively Parallel Architectures, Cluster Analysis, Pattern Recognition, Image Processing and Fuzzy Logic.

**Omar BOUATTANE** has born in 1962 at FIGUIG, south of Morocco. He has his Ph.D. degree in 2001 in Parallel Image Processing on Reconfigurable Computing Mesh from the Faculty of Science Ain Chock, CASABLANCA. He has published more than 30 research publications in various National, International conference proceedings and Journals. His research interests include Massively Parallel Architectures, cluster analysis, pattern recognition, image processing and fuzzy logic.

**Mohamed YOUSSEFI** has born in 1970 at OUARZAZATE, Morocco. He is now a teacher of computer science and researcher at the University Hassan II Mohammedia, ENSET Institute. His research is focused on parallel and distributed computing technologies, Grid Computing and Middleware's. Received the B.S. degree in Mechanics in 1989 and the M.S. degree in Applied Mechanics in 1993 from the ENSET Institute, Mohammedia, Morocco. He received the DEA diploma in Numeric Analysis from the University of RABAT in 1994. He received the Doctorate diploma in Computing and Numeric Analysis from the University MOHAMMED V of RABAT in 1996.

**Abdelhadi RAIHANI** has born in 1968 at El Jadida, Morocco. He is now a teacher of Electronics and researcher at ENSET Institute. His research is focused on parallel architectures and associated treatments. Recently, he worked on Wind Energy. Received the B.S. degree in Electronics in 1987 and the M.S. degree in Applied Electronics in 1991 from the ENSET Institute, Mohammedia, Morocco. He received the DEA diploma in information processing from the Ben M'sik University of Casablanca in 1994. He received the Doctorate diploma in Application of Parallel Architectures in

image processing from the Ain Chok University of Casablanca in 1998.



# A New Round Robin Based Scheduling Algorithm for Operating Systems: Dynamic Quantum Using the Mean Average

Abbas Noon<sup>1</sup>, Ali Kalakech<sup>2</sup>, Seifedine Kadry<sup>1</sup>

<sup>1</sup> Faculty of Computer Science, Arts Sciences and Technology University  
Lebanon

<sup>2</sup> Faculty of Business, Lebanese University  
Lebanon

## Abstract

Round Robin, considered as the most widely adopted CPU scheduling algorithm, undergoes severe problems directly related to quantum size. If time quantum chosen is too large, the response time of the processes is considered too high. On the other hand, if this quantum is too small, it increases the overhead of the CPU.

In this paper, we propose a new algorithm, called AN, based on a new approach called dynamic-time-quantum; the idea of this approach is to make the operating systems adjust the time quantum according to the burst time of the set of waiting processes in the ready queue.

Based on the simulations and experiments, we show that the new proposed algorithm solves the fixed time quantum problem and increases the performance of Round Robin.

**Keywords:** Operating Systems, Multi Tasking, Scheduling Algorithm, Time Quantum, Round Robin.

## 1. Introduction

Modern Operating Systems are moving towards multitasking environments which mainly depends on the CPU scheduling algorithm since the CPU is the most effective or essential part of the computer. Round Robin is considered the most widely used scheduling algorithm in CPU scheduling [8, 9], also used for flow passing scheduling through a network device [1].

CPU Scheduling is an essential operating system task, which is the process of allocating the CPU to a specific process for a time slice. Scheduling requires careful attention to ensure fairness and avoid process starvation in the CPU. This allocation is carried out by software known as scheduler and dispatcher [8, 9].

Operating systems may feature up to 3 distinct types of a long-term scheduler (also known as an admission scheduler or high-level scheduler), a mid-term or medium-term scheduler and a short-term scheduler (fig1).

The dispatcher is the module that gives control of the CPU to the process selected by the short-term scheduler [8].

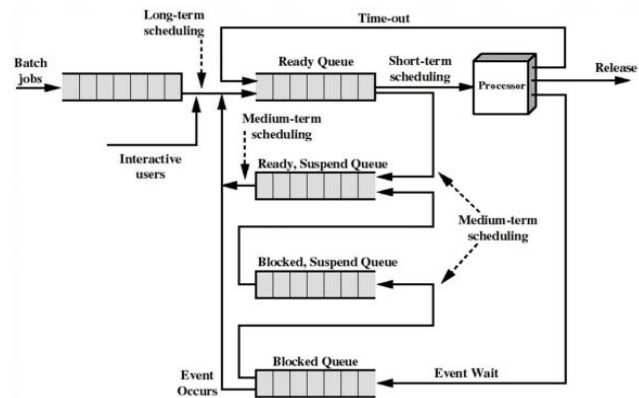


Figure 1: Queuing diagram for scheduling

There are many different scheduling algorithms which varies in efficiency according to the holding environments, which means what we consider a good scheduling algorithm in some cases which is not so in others, and vice versa. The Criteria for a good scheduling algorithm depends, among others, on the following measures [8]:

- Fairness: all processes get fair share of the CPU,
- Efficiency: keep CPU busy 100% of time,
- Response time: minimize response time,
- Turnaround: minimize the time batch users must wait for output,
- Throughput: maximize number of jobs per hour.

Moreover, we should distinguish between the two schemes of scheduling: preemptive and non preemptive algorithms. Preemptive algorithms are those where the burst time of a process being in execution is preempted when a higher

priority process arrives. Non preemptive algorithms are used where the process runs to complete its burst time even a higher priority process arrives during its execution time.

First-Come-First-Served (FCFS)[8, 9] is the simplest scheduling algorithm, it simply queues processes in the order that they arrive in the ready queue. Processes are dispatched according to their arrival time on the ready queue. Being a non preemptive discipline, once a process has a CPU, it runs to completion. The FCFS scheduling is fair in the formal sense or human sense of fairness but it is unfair in the sense that long jobs make short jobs wait and unimportant jobs make important jobs wait [8, 9].

Shortest Job First (SJF) [8, 9] is the strategy of arranging processes with the least estimated processing time remaining to be next in the queue. It works under the two schemes (preemptive and non-preemptive). It's provably optimal since it minimizes the average turnaround time and the average waiting time. The main problem with this discipline is the necessity of the previous knowledge about the time required for a process to complete. Also, it undergoes a starvation issue especially in a busy system with many small processes being run [8, 9].

Round Robin (RR) [8, 9] which is the main concern of this research is one of the oldest, simplest and fairest and most widely used scheduling algorithms, designed especially for time-sharing systems. It's designed to give a better responsive but the worst turnaround and waiting time due to the fixed time quantum concept. The scheduler assigns a fixed time unit (quantum) per process usually 10-100 milliseconds, and cycles through them. RR is similar to FCFS except that preemption is added to switch between processes [2, 3, and 8].

In this paper, we propose a new algorithm to solve the constant time quantum problem. The algorithm is based on dynamic time quantum approach where the system adjusts the time quantum according to the burst time of processes founded in the ready queue. The second section states some of previous works done in this field. Section III describes the proposed method in details. Section IV discusses the simulation done in this method, before concluding this paper in the last section.

## 2. Previous works

Round Robin becomes one of the most widely used scheduling disciplines despite of its severe problem which rose due to the concept of a fixed pre-determined time quantum [2, 3, 4, 5, 6, and 7]. Since RR is used in almost every operating system (windows, BSD, UNIX and Unix-

based etc...), many researchers have tried to fill this gap, but still much less than needs.

Matarneh [2] founded that an optimal time quantum could be calculated by the median of burst times for the set of processes in ready queue, unless if this median is less than 25ms. In such case, the quantum value must be modified to 25ms to avoid the overhead of context switch time [2]. Other works [7], have also used the median approach, and have obtained good results.

Helmy et al. [3] propose a new weighting technique for Round-Robin CPU scheduling algorithm, as an attempt to combine the low scheduling overhead of round robin algorithms and favor short jobs. Higher process weights means relatively higher time quantum; shorter jobs will be given more time, so that they will be removed earlier from the ready queue [3]. Other works have used mathematical approaches, giving new procedures using mathematical theorems [4].

Mohanty and others also developed other algorithms in order to improve the scheduling algorithms performance [5], [6] and [7]. One of them is constructed as a combination of priority algorithm and RR [5] while the other algorithm is much similar to a combination between SJF and RR [6].

## 3. AN Algorithm

In this paper, we present a solution to the time quantum problem by making the operating system adjusts the time quantum according to the burst time of the existed set of processes in the ready queue.

### 3.1 Methodology

When operating system is installed for the first time, it begins with time quantum equals to the burst time of first dispatched process, which is subject to change after the end of the first time quantum. So, we assume that the system will immediately take advantage of this method.

The determined time quantum represents real and optimal value because it based on real burst time unlike the other methods, which depend on fixed time quantum value. Repeatedly, when a new process is loaded into the ready queue in order to be executed, the operating system calculates the average of sum of the burst times of processes found in the ready queue including the new arrival process.

This method needs two registers to be identified:

- SR: Register to store the sum of the remaining burst times in the ready queue.

- AR: Register to store the average of the burst times by dividing the value found in the SR by the count of processes found in the ready queue.

When a process in execution finishes its time slice or its burst time, the ready queue and the registers will be updated to store the new data values.

- If this process finishes its burst time, then it will be removed from the ready queue. Otherwise, it will move to the end of the ready queue.
- SR will be updated by subtracting the time consumed by this process.
- AR will be updated according to the new data.

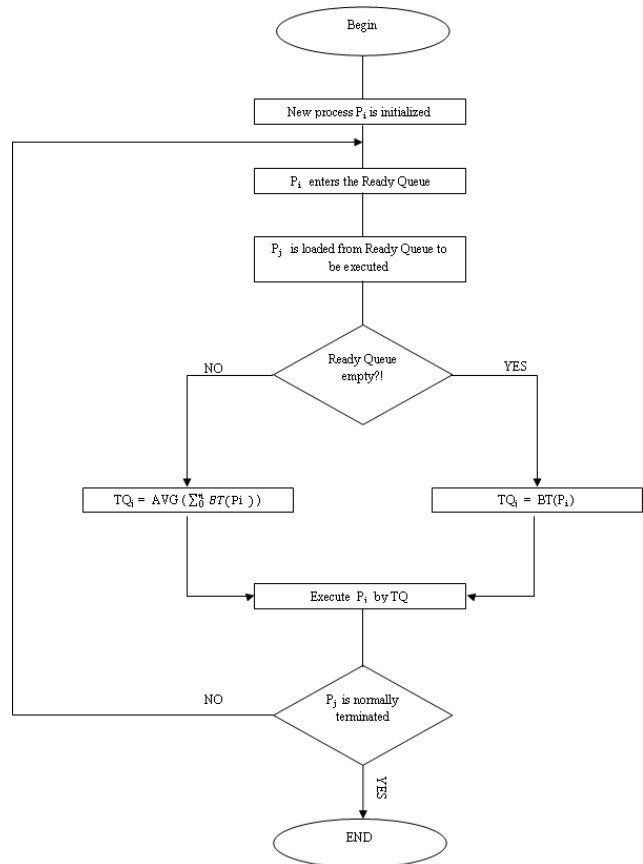
When a new process arrives to the ready queue, it will be treated according to the rules above in addition to updating the ready queue and the registers.

### 3.2 Pseudo Code and Flow Chart

The algorithm described in the previous section can be formally described by pseudo code and flow chart like follows:

```

New process P arrives
P Enters ready queue
Update SR and AR
Process p is loaded from ready queue
into the CPU to be executed
IF (Ready Queue is Empty)
    TQ ← BT (p)
    Update SR and AR
End if
IF (Ready Queue is not empty)
    TQ ← AVG (Sum BT of processes in
    ready queue)
    Update SR and AR
End if
CPU executes P by TQ time
IF (P is terminated)
    Update SR and AR
End if
IF (P is not terminated)
    Return p to the ready queue with
    its updated burst time
    Update SR and AR
End if
    
```



## 4. Simulations

In order to validate our algorithm (AN) over the existing Round Robin, we have built our simulator using MATLAB, since it presents the user data and solutions after fetching in a graphical representation which is not found in most other languages.

Using MATLAB 2010a, we built a simulator for AN algorithm that acquires a triplet (N, AT, BT) where:

- N: the number of processes
- AT: an array of arrival times of all processes
- BT: an array of burst times of all processes

The simulator calculates the average waiting time and the average turnaround time of the whole system consisting of N processes according to the AN algorithm.

We have also built a simulator for Round Robin algorithm that acquires a quadrant (Q, N, AT, BT) where:

- Q: The time quantum (assigned by the user)
- N: the number of processes
- AT: an array of arrival times of all processes
- BT: n array of burst times of all processes

Then the simulator calculates the average waiting time and the average turnaround time of the whole system consisting of N processes according to the Round Robin algorithm.

Finally, we have developed a simple function to compare among the two algorithms presenting graphical result, showing the efficiency of our algorithm over Round Robin. The function loads data from a text file consisting of 50 samples. Each sample is a 4 processes system (N=4). Arrival times and burst times were randomly chosen varying from 10 To 100 milliseconds. Note that we choose N = 4 since whatever N is, we will have the same result as will shown in the result below (figures 2 and 3).

We have chosen a fixed time quantum Q=10 ms in Round Robin it gives the results in fig2 and fig3. In these figures, the x-axis represents the different samples we have targeted, while the y-axis represents the TAT (average of turnaround times) in fig 2, and the WT (average of waiting times) in fig3. In the graphs below a higher vertex means a larger average turnaround time (fig2) and waiting time (fig3). As mentioned before a better algorithm is to minimize turnaround and waiting time, thus the better algorithm has the lowest vertex.

These figures clearly show that for all the tested cases, we obtain better results (lower TAT and WT) when using the AN algorithm.

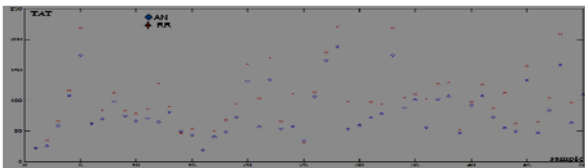


Figure 2: Average Turnaround time for time quantum = 10 ms

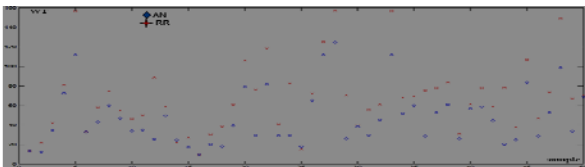


Figure 3: Average Waiting time for time quantum = 10 ms

The same process was done on TQ=15, 20, 25 and 30 ms to cover as much as possible fixed time quantum possibilities, and we always obtain the same results.

#### 4. Results and Observations

As a result of the simulation and hand solved examples we've reached to a conclusion that AN algorithm could improve the efficiency of Round Robin by changing the idea of fixed time quantum to dynamic calculated automatically without the interfere of user.

#### 4.1 Numerical Examples

To evaluate our proposed method and for simplicity seek we will take a group of four processes in four different cases with random burst, in fact the number of processes does not change the result because the algorithm works effectively even if it used with a very large number of processes. For each case, we will compare the result of our developed method with the traditional approach (fixed quantum = 20ms) and with the method proposed in [2]. We should mention here, the numerical values of the 4 different cases are taken from [2].

**Case 1:** Assume four processes arrived at time = 0, with burst time (P1 = 20, P2 = 40, P3 = 60, P4 = 80):

	Fixed Quantum=20ms	Dynamic method [2]	AN
Turn-around time	120	112.5	100
Waiting time	70	77.5	50
Context switch	9	6	5

**Case 2:** Assume four processes arrived at time = 0, with burst time (P1 = 10, P2 = 14, P3 = 70, P4 = 120):

	Fixed Quantum=20ms	Dynamic method [2]	AN
Turn-around time	100.5	96	85.5
Waiting time	47	42.5	32
Context switch	11	6	5

**Case 3:** Assume four processes arrived at different time, respectively 0, 4, 8, and 16, with burst time (P1 = 18, P2 = 70, P3 = 74, P4 = 80):

	Fixed Quantum=20ms	Dynamic method [2]	AN
Turn-around time	106	98.5	81
Waiting time	60	58.5	35
Context switch	10	4	5

**Case 4:** Assume four processes arrived at different time, respectively 0, 6, 13, and 21, with burst time (P1 = 10, P2 = 14, P3 = 70, P4 = 120):

	Fixed Quantum 20ms	Dynamic method [2]	AN
Turn-around time	90.5	46	75.5
Waiting time	37	30.5	22
Context switch	11	4	4

From the above comparisons, it is clear that the dynamic time quantum approach based on the average of processes bursts time is more effective than the fixed time quantum approach and the proposed method in [2] in round robin algorithm, where the dynamic time quantum significantly

reduces the context switch, turnaround time and the waiting time. In addition, the complexity calculation of the mean of the processes is very small.

#### 4.2 Improvements in waiting times and turnaround times

At the end of each run we calculated the percentage of improvement of AN algorithm over Round Robin by implementing a simple rule.

$I = (\text{Vertex [AN]} - \text{Vertex [RR]}) / \text{number of samples}$   
 We obtained the following results (table 1):

**Table 1: Improvement percentage of AN**

TQ	% I(wt[TQ])	% I(tat[TQ])
10 ms	20.1162	20.1162
15 ms	16.1163	16.1162
20 ms	13.8562	13.8562
25 ms	12.6113	12.6112
30 ms	10.4413	10.4412

#### 4.3 Success in Statistics

In addition to the improvement measure (%I), we added another measure of success over failure which is calculated by percentage of success samples over the failed ones. A succeed sample is sample where vertex of AN algorithm is less than vertex of RR.

$S = ((\text{number of succeed samples}) / (\text{total number of samples}))$  we obtained the following results (table 2).

**Table 2: Success over failure percentage of AN**

TQ	%S(tat[TQ])	%S(wt[TQ])
10 ms	96%	96%
15 ms	92%	90%
20 ms	90%	88%
25 ms	88%	88%
30 ms	86%	84%

#### 4.4 Improvement in Context Switches

As a result of our observations, 50% of the processes will be terminated through the first round and as time quantum is calculated repeatedly for each round, then 50% of the remaining processes will be terminated during the second round, with the same manner for the third round, fourth round etc...i.e., the maximum number of rounds will be less than or equal to 6 whatever the number of processes or their burst time (fig4). [2]

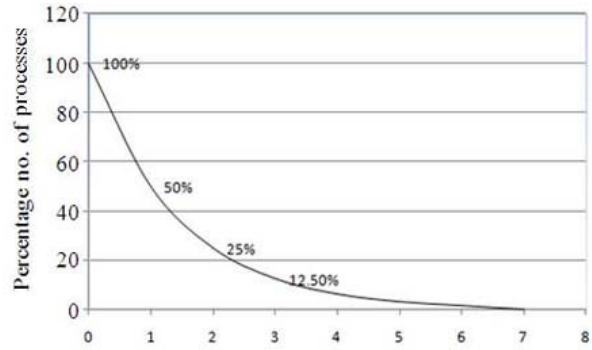


Figure 4: The rate of decrease in the number of processes in each round

The significant decrease of the number of processes will inevitably lead to significant reduction in the number of context switches, which may pose high overhead on the operating system in many cases. The number of context switches can be represented mathematically as follows:

$$Q_T = \left( \sum_{r=1}^r K_r \right) - 1$$

Where:

$Q_T$  = the total number of context switch

$r$  = the total number of rounds,  $r = 1, 2, \dots, 6$

$K_r$  = the total number of processes in each round

In other variants of round robin scheduling algorithm, the context switch occurs even if there is only a single process in the ready queue, where the operating system assigns to the process a specific time quantum  $Q$ [4]. When time quantum expires, the process is interrupted and again assigned the same time quantum  $Q$ , regardless whether the process is alone in the ready queue or not [2, 3], which means that there will be additional unnecessary context switches, while this problem does not occur at all in our new proposed algorithm; because in this case, the time quantum will equal to the remaining burst time of the process.

## 5. Conclusion

Time quantum is the bottleneck facing round robin algorithm and was more frequently asked question: What is the optimal time quantum to be used in round robin algorithm?

In light of the effectiveness and the efficiency of the RR algorithm, this paper provides an answer to this question by using dynamic time quantum instead of fixed time quantum, where the operating system itself finds the optimal time quantum without user intervention.

In this paper, we have discussed the AN algorithm that could be a simple step for a huge aim in obtaining an optimal scheduling algorithm. It will need much more efforts and researches to score a goal.



From the simulation study, we get an important conclusion; that the performance of AN algorithm is higher than that of RR in any system. The use of dynamic scheduling algorithm increased the performance and stability of the operating system and supports building of a self-adaptation operating system, which means that the system is who will adapt itself to the requirements of the user and not vice versa.

## References

- [1] Weiming Tong, Jing Zhao, "Quantum Varying Deficit Round Robin Scheduling Over Priority Queues", International Conference on Computational Intelligence and Security. pp. 252- 256, China, 2007.
- [2] Rami J. Matarneh, "Self-Adjustment Time Quantum in Round Robin Algorithm Depending on Burst Time of the Now Running Processes", American Journal of Applied Sciences, Vol 6, No. 10, 2009.
- [3] Tarek Helmy, Abdelkader Dekdouk, "Burst Round Robin as a Proportional-Share Scheduling Algorithm", In Proceedings of The fourth IEEE-GCC Conference on Towards Techno-Industrial Innovations, pp. 424-428, Bahrain, 2007.
- [4] Samih M. Mostafa, S. Z. Rida, Safwat H. Hamad, "Finding Time Quantum Of Round Robin Cpu Scheduling Algorithm In General Computing Systems Using Integer Programming", International Journal of Research and Reviews in Applied Sciences (IJRRAS), Vol 5, Issue 1, 2010.
- [5] Rakesh Mohanty, H. S. Beheram Khusbu Patwarim Monisha Dash, M. Lakshmi Prasanna , "Priority Based Dynamic Round Robin (PBDRR) Algorithm with Intelligent Time Slice for Soft Real Time Systems", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No.2, February 2011.
- [6] Rakesh Mohanty, H. S. Behera, Khusbu Patwari, Monisha Dash, "Design and Performance Evaluation of a New Proposed Shortest Remaining Burst Round Robin (SRBRR) Scheduling Algorithm", In Proceedings of International Symposium on Computer Engineering & Technology (ISCET), Vol 17, 2010.
- [7] Rakesh Mohanty, H. S. Behera, Debashree Nayak, "A New Proposed Dynamic Quantum with Re-Adjusted Round Robin Scheduling Algorithm and Its Performance Analysis", International Journal of Computer Applications (0975 – 8887), Volume 5– No.5, August 2010.
- [8] Silberschatz ,Galvin and Gagne, Operating systems concepts, 8th edition, Wiley, 2009.
- [9] Lingyun Yang, Jennifer M. Schopf and Ian Foster, "Conservative Scheduling: Using predictive variance to improve scheduling decisions in Dynamic Environments", SuperComputing 2003, November 15-21, Phoenix, AZ, USA.

# A Multi-Modal Recognition System Using Face and Speech

Samir Akrouf<sup>1</sup>, Belayadi Yahia<sup>2</sup>, Mostefai Messaoud<sup>2</sup> and Youssef chahir<sup>3</sup>

<sup>1</sup>Department of Computer Science, University of Bordj Bou Arréridj, Algeria  
El Anasser, 34030, BBA Algeria

<sup>2</sup>Department of Computer Science, University of Bordj Bou Arréridj, Algeria  
El Anasser, 34030, BBA Algeria

<sup>2</sup>Department of Computer Science, University of Bordj Bou Arréridj, Algeria  
El Anasser, 34030, BBA Algeria

<sup>3</sup>Department of Computer Science, University of Caen Lower Normandie, France  
Caen, State ZIP/Zone, France

## Abstract

Nowadays Person Recognition has got more and more interest especially for security reasons. The recognition performed by a biometric system using a single modality tends to be less performing due to sensor data, restricted degrees of freedom and unacceptable error rates. To alleviate some of these problems we use multimodal biometric systems which provide better recognition results. By combining different modalities, such as speech, face, fingerprint, etc., we increase the performance of recognition systems.

In this paper, we study the fusion of speech and face in a recognition system for taking a final decision (i.e., accept or reject identity claim). We evaluate the performance of each system differently then we fuse the results and compare the performances.

**Keywords:** Biometrics, data fusion, face recognition, automatic speaker recognition, data processing, decision fusion.

## 1. Introduction

Identity recognition is becoming more and more used in the last years. Demand is increasing for reliable automatic user identification systems in order to secure accesses to lots of services or buildings. Biometric Identification [1] is the area related to person recognition by means of physiological features (fingerprints, iris, voice, face, etc.).

A biometric person recognition system can be used for person identification or verification. For the verification, a user claims a certain identity ("I am X"). The system accepts or rejects this claim (deciding if really the user is who he claims to be). For identification, there is no identity claim. The system decides who the user is. In this paper we use two the biometrics which appears to be the most popular ones and are less restricting for person identification (voice and face). The major strength of

voice and face biometrics is their high acceptance by the society.

This multiple sensors capture different biometric traits. Such systems, known as multi-modal biometric systems [2], are more reliable due to the presence of multiple pieces of evidence. These systems are able to meet the stringent performance requirements imposed by various applications. Moreover, it will be extremely difficult for an intruder to violate the integrity of a system requiring multiple biometric traits.

In the literature we find that combining different biometric modalities enables to achieve better performances than techniques based on single modalities [3]–[10]. Combining different modalities allows to overcome problems due to single modalities. The *fusion* algorithm, which combines the different modalities, is a very critical part of the recognition system. So before the fusion one would ask what strategy do we have to adopt in order to make the final decision?

The sensed data (face and speech) are processed by different recognition systems: a face identification system and a speaker identification system. Each system, given the sensed data, will deliver a matching score in the range between zero (reject) and one (accept). The fusion module will combine the opinions of the different systems and give a binary decision: accept or reject the claim.

An identification scenario involving two modalities is shown in Fig. 1. The paper will address the issue of which binary classifier to use for the fusion of different expert "opinions."

The face recognition system will be presented in paragraph 2. The speaker recognition system based on text-dependent approach is discussed in paragraph 3.

The fusion [2]–[4] of different modalities is described in paragraph 5.

Finally we present the evaluation results and the main conclusions.

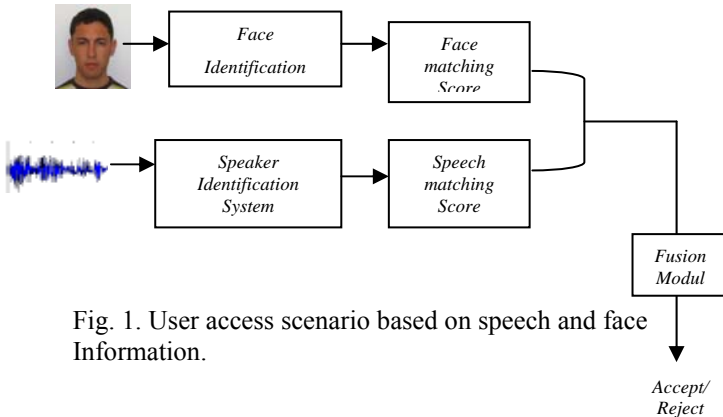


Fig. 1. User access scenario based on speech and face Information.

## 2. Face Recognition

This paper uses a hybrid method combining principal components analysis (PCA) [11] and the discrete cosine transform (DCT) [12] for face identification [13].

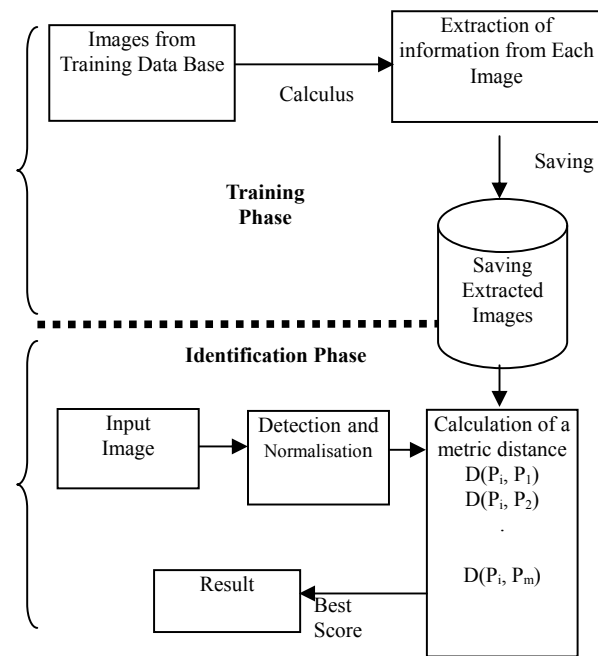


Fig. 2. Recognition Algorithm Stages.

### 2.1 Presentation of the Hybrid Method

PCA and DCT have certain mathematical similarities since that they both aim to reduce the dimensions of data. The use of a hybrid method combining these two techniques gave performances slightly higher than those obtained with only one method (experiments being made on three different image data bases). Its principle is very simple: each image is transformed into a coefficient vector (in the training and recognition phase). We first use the DCT method which produces a result used as entry for the PCA

method. We use PCA with coefficients vectors instead of pixels vectors. We notice that this technique requires more time than PCA (because of the calculation of the coefficients) in particular with data bases of average or reduced size but it should be noted that it requires less memory what makes its use advantageous with bases of significant size.

### 2.2 Experimental Results

The tests were performed by using the image data bases ORL, Yale Faces and BBAFaces. The latter was created at the University Center of Bordj Bou Arreridj in 2008. It is composed by 23 people with 12 images for each one of them (for the majority of the people, the images were taken during various sessions). The images reflect various facial expressions with different intensity variations and different light sources. To facilitate the tests, the faces were selected thereafter manually in order to get images of 124 X 92 pixels, we then convert them into gray levels and store them with JPG format. Fig. 3. represents a typical example of the data. It should be noted that certain categories of this data are not retained for the tests.

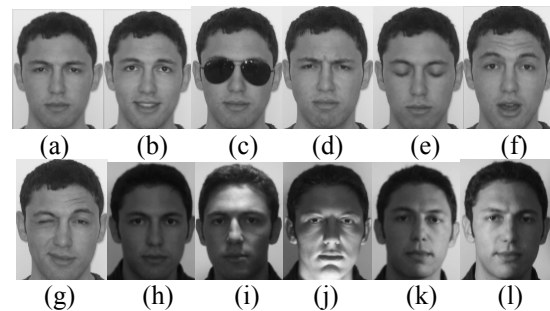


Fig. 3. Example from BBAFaces. (a): normal, (b): happy, (c): glasses, (d): sad, (e): sleepy, (f): surprised, (g): wink, (h): dark, (i): top light, (j): bottom light, (k): left light, (l): right light.

In the following we will expose the results obtained for the tests realized with Yale Faces and BBA Faces.

Table 1: Rates of Recognition

<i>Data Base</i>	<i>PCA</i>	<i>PCA + DCT</i>
BBA Faces	57.06 %	66.30 %
Yale Faces	62 %	72.77 %
ORL Base	71.38 %	72.77 %

Finally we conclude that the combination of PCA with DCT offers higher rates of recognition than those obtained with only one method which justifies our choice for the algorithm used in our system.

### 3. Speaker Recognition System

Nowadays The Automatic Treatment of speech is progressing, in particular in the fields of Automatic Speech Recognition "ASR" and Speech Synthesis.

The automatic speaker recognition is represented like a particular pattern recognition task. It associates the problems relating to the speaker identification or verification using information found in the acoustic signal: we have to recognize a person by using his voice. ASR is used in many fields, like domestic, military or jurisprudence applications.

In this work we use an automatic speaker recognition system presented an earlier paper [15]. We will use speaker recognition in text independent mode since we dispose of very few training data. We have to estimate with few data a robust speaker model to allow the recognition of the speaker.

#### 3.1 Basic System

A speaker recognition system comprises 4 principal elements:

1. An acquisition and parameterization module of the signal: to represent the message in an exploitable form by the system.
2. A training module: who is charged to create a vocal reference of the speaker starting from a sample of his voice «GMM Gaussian Mixture Models».
3. A resemblance calculus module: who calculates the resemblance between a sample signal and a given reference corresponding to a person.
4. A decision module: based on a strategy of decision.

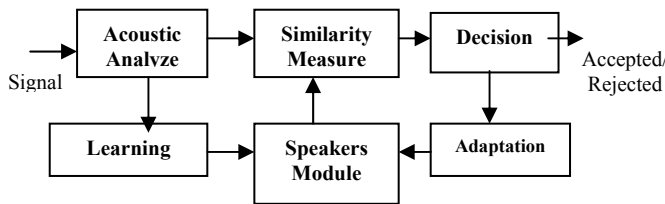


Fig. 4. Typical diagram of a checking speaker system

#### 3.2 Speaker Identification "SI"

The speaker identification consists in recognizing a person among many speakers by comparing his vocal expression with known references. From a diagrammatic point of view "see figure 4", a sequence of word is given in entry of the ASR system. For each known speaker, the sequence of word is compared with a characteristic reference of the speaker. The identity of the speaker whose reference is the nearest to the sequence of word will be the output datum of the system (ASR). Two modes of identification are possible: identification in a closed unit for which the speaker is identified among a known number of speakers

or identification in an open unit for which the speaker to be identified does not belong inevitably to this unit [16].

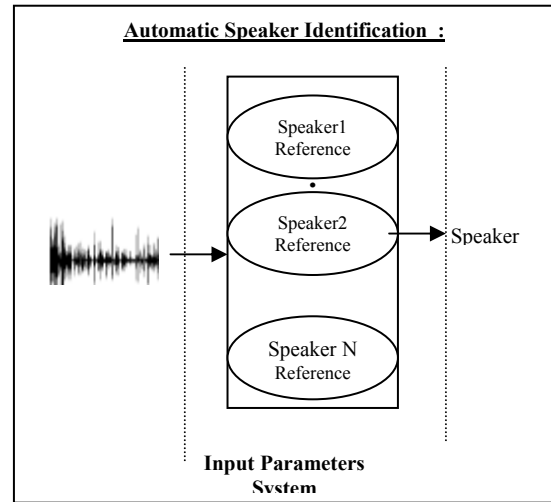


Fig.5. Automatic Speaker Identification

#### 3.4 Speaker Verification "SV"

The checking "or authentication" of the speaker consists in, after the speaker declines his identity, checking the adequacy of its vocal message with the acoustic reference of the speaker who it claims to be. A measurement of similarity is calculated between this reference and the vocal message then compared with a threshold. In the case the measurement of similarity is higher than the threshold, the speaker is accepted. Otherwise, the speaker is considered as an impostor and is rejected [16].

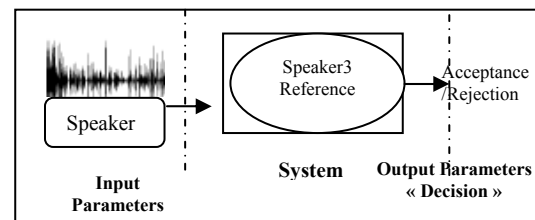


Fig. 6. Automatic Speaker Verification

#### 3.5 Text Dependent and independent mode

We distinguish between the speaker recognition independently of the contents of the sentence pronounced "text independent mode" and the speaker recognition with the pronunciation of a sentence containing a key word "text dependent mode". The levels of dependence to the text are classified according to the applications:

- Systems with free text "or *free-text*": the speaker is free to pronounce what he wants. In this mode, the sentences of training and test are different.
- Systems with suggested text "or *text-prompted*": a text, different on each session and for each person, is imposed to the speaker and is determined by the machine. The sentences of training and test can be different.
- Systems dependent on the vocabulary "or *vocabulary-dependent*": the speaker pronounces a sequence of words resulting from a limited vocabulary. In this mode, the training and the test are carried out on texts made up and starting from the same vocabulary.
- Personalized systems dependent on the text (or *to use-specific text dependent*): each speaker has his own password. In this mode, the training and the test are carried out on the same text.

The vocal message makes the task of ASR systems easier and the performances are better. The recognition in text mode independent requires more time than the text mode dependent [17].

### 3.6 Speaker Modeling

Here we briefly introduce the most usually used techniques in the speaker recognition. Here the problem (speaker recognition) can be formulated as a classification problem. Various approaches were developed; nevertheless we can classify them in four great families:

1. Vectorial approach: the speaker is represented by a set of parameter vectors in the acoustic space. The principal is he recognition containing "Dynamic Time Warping" DTW and by vectorial quantification.
2. Statistical approach: it consists in representing each speaker by a probabilistic density in the acoustic space parameters. It covers the techniques of modeling by the Markov hidden models, the Gaussian mixtures and statistical measurements of the second order.
3. The connexionnist approach: mainly consists in modeling the speakers by neuron networks.
4. Relative approach: here we model a speaker relatively with other reference speakers which models are well learned.

Finally we say that the automatic speaker recognition is probably the most ergonomic method to solve the access problems. However, the voice cannot be regarded as a biometric characteristic of a person taking into account intra-speaker variability. A speaker recognition system generally proceeds in three stages: acoustic analysis of the speech signal, speaker modeling and finally taking the decision. In acoustic analysis, the MFCC are the most used acoustic coefficients. As for the modeling, GMM

constitutes the state of the art in ASR. The decision of an automatic speaker recognition system is based on the two processes of speaker identification and/or checking whatever the application or the task is concerned with.

## 4. Performance of Biometric Systems

The most significant and decisive argument which makes the difference between a biometric system and another is its error rate, a system is considered ideal if its:

**False Rejection Rate= False Acceptance Rate= 0;**

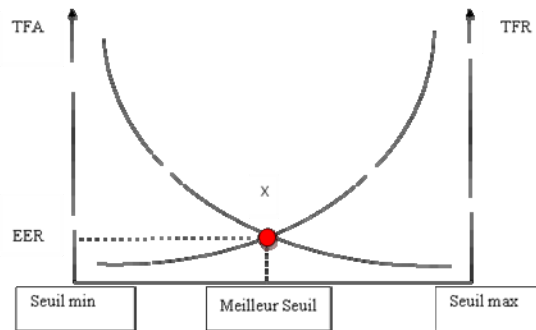


Fig. 7. Illustration of typical errors in a biometric system.

Consequently it is necessary to find a compromise between the two rates which are the junction of the curves (point X) where couple (TFR, TFA) is minimal.

## 5. Fusion by Decision Methods

Among the fusion of decision methods the most used one quotes:

### 5.1 Fusion by the AND operator:

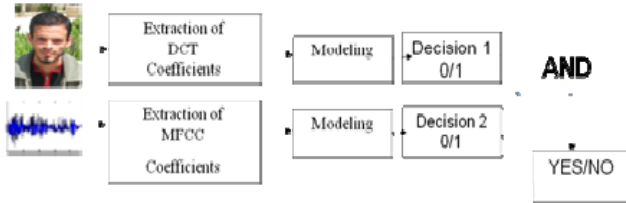
If all the systems decided 1 then the final decision is YES with the operator AND, a false acceptance occurs only if the result of each test is a false acceptance. The probability of false acceptance is thus the product of the probabilities obtained for each test.

$$P(FA) = P1(FA).P2(FA)$$

But in a symmetrical way, the probability of false rejections becomes:

$$P(FR) = P1(FR) + P2(FR) - P1(FR).P2(FR)$$





- ✓  $P(FA_1)=0.1.$
- ✓  $P(FR_1)=0.6.$
- ✓  $P(FA_2)=0.3.$
- ✓  $P(FR_2)=0.2.$

□ In the speaker recognition system we obtained:

When applying fusion operators AND and OR we obtain:

➤ **AND Operator:**

$$P(FR) = 0.12$$

$$P(FA) = 0.37$$

➤ **OR Operator :**

$$P(FA) = 0.03$$

$$P(FR) = 0.68$$

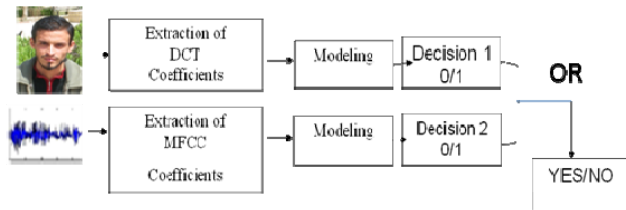
### 5.2 Fusion by OR operator

If one of the systems decided 1 then the final decision is **YES**. The user is accepted so at least one of the two tests is positive. In this configuration, a false rejection can exist only if the two tests produce a false rejection. The final probability of false rejection  $P(FR)$  is the product of the two probabilities of false rejection

$$P(FR) = P1(FR)*P2(FR)$$

The probability of false final acceptance is described by:

$$P(FA) = P1(FA) + P2(FA) - P1(FA)*P2(FA)$$



The tests carried out confirm not only the importance of biometric fusion but also the robustness and the effectiveness of the new system which makes its appearance much more through the real tests where the one modal systems had a fall of performances.

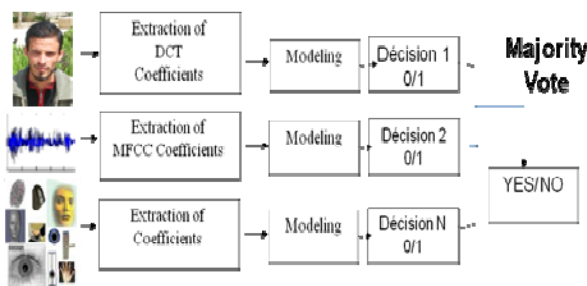
We noticed that Fusion give better results than those obtained by the first system.

We also noticed that the performances are closely related to the number of coefficients taken and the number of GMM. Finally we could say that the significant factor is the size of the base.

### 5.3 Fusion by the majority vote:

If the majority of the systems decided 1 then the final decision is **YES**.

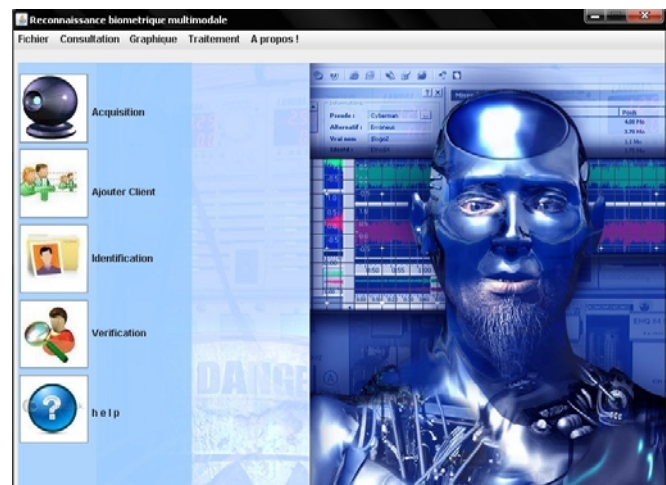
Majority Vote is a simple method to combine the exits of multiple sources and use a voting process. In this case, each source must provide a decision of its choice and the final decision is based on a majority rule.



## 6. Demonstration System

In the following we present some interfaces of our Multi-Modal Recognition system which was developed using a Pentium IV cadenced at 2 Ghz and using 1 Giga bytes of RAM. It was running under Windows XP professional edition and using Java 1.6 as programming language.

### 1. Main Interface

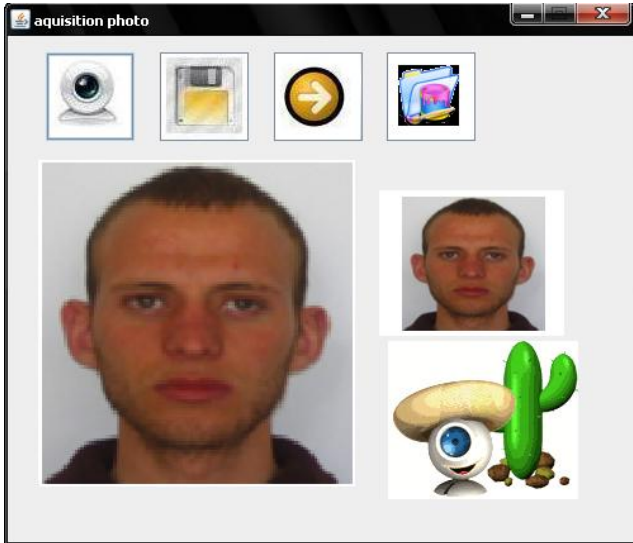


### 5.4 Experimental Results

In order to test our system we used ORL and TIMIT bases. We used 30 customers and 30 impostors with a base containing 100 elements. The face recognition system generated 13 false rejections and 6 false acceptances in an average time equal to 5.6 seconds whereas the speaker recognition system produced 7 false rejections and 12 false acceptances in an average time equal to 6.1 seconds.

□ In the face recognition system we obtained:

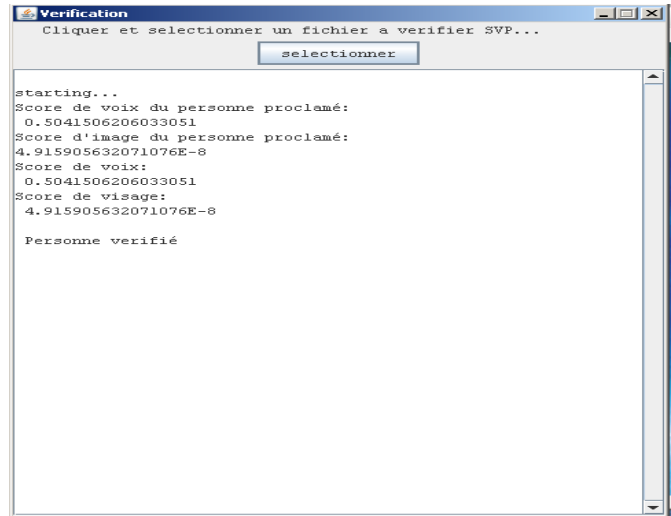
## 2. Acquisition Module for Face



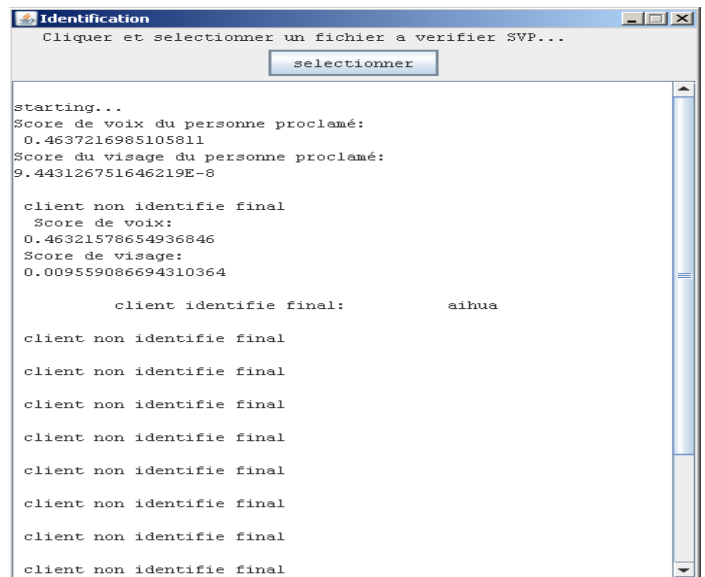
## 3. Acquisition Module for Speaker



## 4. Verification Process



## 5. Identification Process



## 7. Conclusions

This paper provides results obtained on a multi-modal biometric system that uses face and voice features for recognition purposes. We used fusion at the decision level with OR and AND operators. We showed that the resulting system (multi-modal) considered here provide better performance than the individual biometrics. For the near future we are collecting data corresponding to three

biometric indicators - fingerprint, face and voice in order to conceive a better multi-modal recognition system.

## Acknowledgments

Special thanks to Benterki Mebarka and Bechane Louiza for their contribution to this project.

Samir Akrouf thanks the Ministry of Higher Education for the financial support of this project (project code: B\*0330090009).

## References

- [1] A. K. Jain, R. Bolle, and S. Pankanti, *Biometrics: Personal Identification in Networked Society*. Boston, MA: Kluwer, 1998.
- [2] A. K. Jain, S. Prabhakar, and S. Chen, "Combining multiple matchers for a high security fingerprint verification system," *Pattern Recognition Letters*, vol. 20, pp. 1371-1379, 1999.
- [3] R. Brunelli and D. Falavigna, "Person identification using Multiple cues," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 17, pp. 955-966, Oct. 1995.
- [4] B. Duc, G. Maitre, S. Fischer, and J. Bigun, "Person Authentication by fusing face and speech information," in *1st Int. Conf. Audio- Video- Based Biometric Person Authentication AVBPA'97*, J. Bigun, G. Chollet, and G. Borgefors, Eds. Berlin, Germany: Springer-Verlag, Mar. 12-14, 1997, vol. 1206 of Lecture Notes in Computer Science, pp. 311-318.
- [5] E. Bigun, J. Bigun, B. Duc, and S. Fischer, "Expert conciliation for multi modal person authentication systems by Bayesian statistics," in *Proc. 1st Int. Conf. Audio-Video- Based Biometric Person Authentication AVBPA'97*. Berlin, Germany: Springer-Verlag, Lecture Notes in Computer Science, 1997, pp. 291-300.
- [6] L. Hong and A. K. Jain, "Integrating faces and fingerprint for Personal identification," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, 1997.
- [7] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On Combining classifiers," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, pp. 226-239, 1998.
- [8] A. K. Jain, L. Hong, and Y. Kulkarni, "A multimodal biometric system using fingerprints, face and speech," in *Proc. 2nd Int. Conf. Audio-Video Based Biometric Person Authentication*, Washington, D.C., Mar. 22-23, 1999, pp. 182-187.
- [9] T. Choudhury, B. Clarkson, T. Jebara, and A. Pentland, "Multimodal person recognition using unconstrained audio and video," in *Proc. 2nd Int. Conf. Audio-Video Based Person Authentication*, Washington, D.C., Mar. 22-23, 1999, pp. 176-180.
- [10] S. Ben-Yacoub, "Multimodal data fusion for person authentication using SVM," in *Proc. 2nd Int. Conf. Audio-Video Based Biometric Person Authentication*, Washington, D.C., Mar. 22-23, 1999, pp. 25-30.
- [11] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Science*, pages 71-86, 1991.
- [12] Ronny Tjahyadi, Wanquan Liu, Svetha Venkatesh.

- Application of the DCT Energy Histogram for Face Recognition. 2nd International Conference on Information Technology for Application (ICITA 2004) PP 305-310
- [13] Samir Akrouf, Sehili Med Amine, Chakhchoukh Abdesslam, Messaoud Mostefai and Youssef Chahir 2009 Fifth International Conference on Mems Nano and Smart Systems 28-30 December 2009 Dubai UAE.
  - [14] N Morizet, Thomas Ea, Florence Rossant, Frédéric Amiel Et Amara Amara, *Revue des algorithmes PCA, LDA et EBGM utilisés en reconnaissance 2D du visage pour la biométrie*, Tutoriel Reconnaissance d'images, MajecStic 2006 Institut Supérieur d'Electronique de Paris (ISEP).
  - [15] Akrouf Samir, Mehamel Abbas, Benhamouda Nacéra, Messaoud Mostefai An Automatic Speaker Recognition System, 2009 the 2<sup>nd</sup> International Conference on Advanced Computer Theory Engineering (ICACTE 2009) Cairo, Egypt September 25-27 2009
  - [16] Approche Statistique pour la Reconnaissance Automatique du Locuteur : Informations Dynamiques et Normalisation Bayésienne des Vraisemblances", October, 2000.
  - [17] Yacine Mami "Reconnaissance de locuteurs par localisation dans un espace de locuteurs de référence" Thèse de doctorat, soutenue le 21 octobre 2003.

**Samir Akrouf** was born in Bordj Bou Arréridj, Algeria in 1960. He received his Engineer degree from Constantine University, Algeria in 1984. He received his Master's degree from University of Minnesota, USA in 1988. Currently, he is an assistant professor at the Computer department of Bordj Bou Arréridj University, Algeria. He is an IACSIT member and is a member of LMSE laboratory (a research laboratory in Bordj Bou Arréridj University). He is also the director of Mathematics and Computer Science Institute of Bordj Bou Arréridj University. His main research interests are focused on Biometric Identification, Computer Vision and Computer Networks.

**Yahia Belayadi** was born in Bordj Bou Arréridj, Algeria in 1961. He received his Engineer degree from Setif University Algeria in 1987. He received his magister from Setif University Algeria in 1991. Currently, he is an assistant professor at the Computer department of Bordj Bou Arréridj University, Algeria. He also is the director of University Center of Continuous Education in Bordj Bou Arréridj.

**Messaoud Mostefai** was born in Bordj Bou Arréridj, Algeria in 1967. He received his Engineer degree from Algiers University, Algeria in 1990. He received a DEA degree en Automatique et Traitement Numérique du Signal (Reims - France) in 1992. He received his doctorate degree en Automatique et Traitement Numérique du Signal (Reims - France) in 1995. He got his HDR Habilitation Universitaire : Theme : « Adéquation Algorithmique /Architecture en traitement d'images » in (UFAS Algeria) in 2006. Currently, he is a professor at the Computer department of Bordj Bou Arréridj University, Algeria. He is a member of LMSE laboratory (a research laboratory in Bordj Bou Arréridj University). His main research interests are focused on classification and Biometric Identification, Computer Vision and Computer Networks.

**Youssef Chahir** is an Associate Professor (since '00) at [GREYC Laboratory CNRS UMR 6072, Department of Computer Science, University of Caen Lower-Normandy](#) France.

# A Temporal Neuro-Fuzzy Monitoring System to Manufacturing Systems

Rafik Mahdaoui<sup>1,2</sup>, Leila Hayet Mouss<sup>1</sup>, Mohamed Djamel Mouss<sup>1</sup>, Ouahiba Chouhal<sup>1,2</sup>  
1 Laboratoire d'Automatique et Productique (LAP) Université de Batna,  
Rue Chahid Boukhrouf 05000 Batna, Algérie  
1,2 Centre universitaire Khenchela Algérie,  
Route de Batna BP:1252, El Houria, 40004 Khenchela Algérie

## Abstract

Fault diagnosis and failure prognosis are essential techniques in improving the safety of many manufacturing systems. Therefore, on-line fault detection and isolation is one of the most important tasks in safety-critical and intelligent control systems.

Computational intelligence techniques are being investigated as extension of the traditional fault diagnosis methods. This paper discusses the Temporal Neuro-Fuzzy Systems (TNFS) fault diagnosis within an application study of a manufacturing system. The key issues of finding a suitable structure for detecting and isolating ten realistic actuator faults are described. Within this framework, data-processing interactive software of simulation baptized NEFDIAG (NEuro Fuzzy DIAGnosis) version 1.0 is developed.

This software devoted primarily to creation, training and test of a classification Neuro-Fuzzy system of industrial process failures. NEFDIAG can be represented like a special type of fuzzy perceptron, with three layers used to classify patterns and failures. The system selected is the workshop of SCIMAT clinker, cement factory in Algeria.

**Keywords:** *Diagnosis; artificial neuronal networks; fuzzy logic; Neuro-fuzzy systems; pattern recognition; FMEAC (Failure Mode, Effects and Criticality Analysis).*

## 1. Introduction

Several methods have been proposed in order to solve the fault detection and fault diagnosis problems. The most commonly employed solution approaches for fault diagnosis system include (a) model-based, (b) knowledge-based, and (c) pattern recognition-based approaches. Generally, analytical model-based methods can be designed in order to minimize the effect of unknown disturbance and perform the consistent sensitivity analysis.

Knowledge-based methods are used when there is a lot of experience but not enough details to develop accurate quantitative models. Pattern recognition methods are applicable to a wide variety of systems and exhibit real-time characteristics. [8]. Therefore the human expert in his mission of diagnosing the cause of a failure of a whole system, uses quantitative or qualitative information. On another side, in spite of the results largely surprising obtained by the ANN in monitoring and precisely in diagnosis they remain even enough far from equalizing the sensory capacities and of reasoning human being. Fuzzy logic makes another very effective axis in industrial diagnosis.

Also, can we replace the human expert for automating the task of diagnosis by using the Neuro-fuzzy approach? In addition, how did the human expert gather all relevant information and permit him to make their decision? Our objective consists of the following: making an association (adaptation) between the techniques of fuzzy logic and the temporal neural networks techniques (Neuro-fuzzy system), choosing the types of neural networks, determining the fuzzy rules, and finally determining the structure of the temporal Neuro-Fuzzy system to maximize the automation of the diagnosis task.

In order to achieve this goal we organize this article into three parts. The first part presents principal architectures of diagnosis and prognosis methods and principles for Temporal Neuro-Fuzzy systems operation and their applications (sections 2 and 3). The second part is dedicated to the workshop of clinker of cement factory (Section 4). Lastly, in the third part we propose a Neuro-Fuzzy system for system of production diagnosis. Machine Fault Prognosis

The literatures of prognosis are much smaller in comparison with those of fault diagnosis. The most obvious and normally used prognosis is to use the given current and past machine condition to predict how much time is left before a failure occurs. The time left before



observing a failure is usually called remaining useful life (RUL). In order to predict the RUL, data of the fault propagation process and/or the data of the failure mechanism must be available. The fault propagation process is usually tracked by a trending or forecasting model for certain condition variables. There are two ways in describing the failure mechanism. The first one assumes that failure only depends on the condition variables, which reflect the actual fault level, and the predetermined boundary (figure 1).

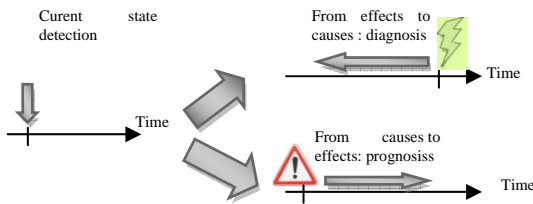


Figure 1. detection ,diagnosis and prognosis- the phenomenological aspect

The definition of failure is simply defined that the failure occurs when the fault reaches a predetermined level. The second one builds a model for the failure mechanism using available historical data. In this case, different definitions of failure can be defined as follows: (a) an event that the machine is operating at an unsatisfactory level; or (b) it can be a functional failure when the machine cannot perform its intended function at all; or (c) it can be just a breakdown when the machine stops operating, etc.

The approaches to prognosis fall into three main categories: statistical approaches, model-based approaches, and data-driven based approaches.

Data-driven techniques are also known as data mining techniques or machine learning techniques. They utilize and require large amount of historical failure data to build a prognostic model that learns the system behavior. Among these techniques, artificial intelligence was regularly used because of its flexibility in generating appropriate model.

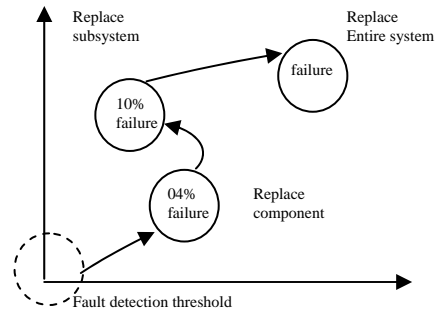


Figure 2. Prognosis technical approaches

Several of the existing approaches used ANNs to model the systems and estimate the RUL. Zhang and Ganesan [14] used self-organizing neural networks for multivariable trending of the fault development to estimate the residual life of bearing system. Wang and Vachtsevanos [13] proposed an architecture for prognosis applied to industrial chillers. Their prognostic model included dynamic wavelet neural networks, reinforcement learning, and genetic algorithms. This model was used to predict the failure growth of bearings based on the vibration signals. SOM and back propagation neural networks (BPNN) methods using vibration signals to predict the RUL of ball bearing were applied by Huang et al. in [12].

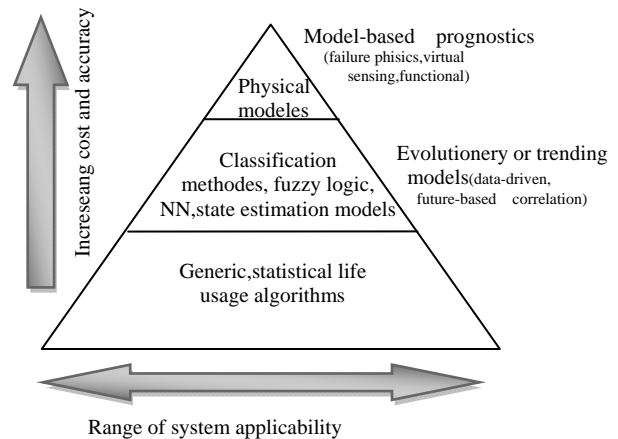


Figure 3. Prognosis technical approaches

Wang et al. [14] utilized and compared the results of two predictors, namely recurrent neural networks and ANFIS, to forecast the damage propagation trend of rotating machinery. In [15], Yam et al. applied a recurrent



neural network for predicting the machine condition trend. Dong et al. [16] employed a grey model and a BPNN to predict the machine condition. Altogether, the data-driven techniques are the promising and effective techniques for machine condition prognosis.

## 2. Temporal Neuro-Fuzzy Systems

Fuzzy neural network (FNN) approach has become a powerful tool for solving real-world problems in the area of forecasting, identification, control, image recognition and others that are associated with high level of uncertainty [2,7,10,11,14,23,24,23]

The Neuro-fuzzy model combines, in a single framework, both numerical and symbolic knowledge about the process. Automatic linguistic rule extraction is a useful aspect of NF especially when little or no prior knowledge about the process is available [3]. For example, a NF model of a non-linear dynamical system can be identified from the empirical data.

This model can give us some insight about the on linearity and dynamical properties of the system.

The most common NF systems are based on two types of fuzzy models TSK [5] [7] combined with NN learning algorithms. TSK models use local linear models in the consequents, which are easier to interpret and can be used for control and fault diagnosis [23]. Mamdani models use fuzzy sets as consequents and therefore give a more qualitative description. Many Neuro-fuzzy structures have been successfully applied to a wide range of applications from industrial processes to financial systems, because of the ease of rule base design, linguistic modeling, and application to complex and uncertain systems, inherent non-linear nature, learning abilities, parallel processing and fault-tolerance abilities. However, successful implementation depends heavily on prior knowledge of the system and the empirical data [25].

Neuro-fuzzy networks by intrinsic nature can handle limited number of inputs. When the system to be identified is complex and has large number of inputs, the fuzzy rule base becomes large.

NF models usually identified from empirical data are not very transparent. Transparency accounts a more meaningful description of the process i.e less rules with appropriate membership functions. In ANFIS [2] a fixed structure with grid partition is used. Antecedent and consequent parameters are identified by a combination of least squares estimate and gradient based method, called hybrid learning rule. This method is fast and easy to implement for low dimension input spaces. It is more prone to lose the transparency and the local model accuracy because of the use of error back propagation that is a global and not locally nonlinear optimization

procedure. One possible method to overcome this problem can be to find the antecedents & rules separately e.g. clustering and constrain the antecedents, and then apply optimization.

Hierarchical NF networks can be used to overcome the dimensionality problem by decomposing the system into a series of MISO and/or SISO systems called hierarchical systems [14]. The local rules use subsets of input spaces and are activated by higher level rules[12].

The criteria on which to build a NF model are based on the requirements for faults diagnosis and the system characteristics. The function of the NF model in the FDI scheme is also important i.e. Preprocessing data, Identification (Residual generation) or classification (Decision Making/Fault Isolation).

For example a NF model with high approximation capability and disturbance rejection is needed for identification so that the residuals are more accurate.

Whereas in the classification stage, a NF network with more transparency is required.

The following characteristics of NF models are important:

- Approximation/Generalisation capabilities
- transparency: Reasoning/use of prior knowledge /rules
- Training Speed/ Processing speed
- Complexity

Transformability: To be able to convert in other forms of NF models in order to provide different levels of transparency and approximation power.

Adaptive learning

Two most important characteristics are the generalising and reasoning capabilities. Depending on the application requirement, usually a compromise is made between the above two.

In order to implement this type of Neuro-Fuzzy Systems For Fault Diagnosis and Prognosis and exploited to diagnose of dedicated production system we have to propose data-processing software NEFDIAG (Neuro-Fuzzy Diagnosis).

The Takagi-Sugeno type fuzzy rules are discussed in detail in Subsection A. In Subsection B, the network structure of FENN is presented.

### 2.1 Temporal Fuzzy rules

Recently, more and more attention has paid to the Takagi-Sugeno type rules [9] in studies of fuzzy neural networks. This significant inference rule provides an analytic way of analyzing the stability of fuzzy control systems. If we combine the Takagi-Sugeno controllers together with the controlled system and use state-space equations to describe the whole system [10], we can get another type of rules to describe nonlinear systems as below:

Rule r: IF  $X_1$  is  $T_{x_1}^r$  AND ... AND  $X_n$  is  $T_{x_n}^r$  AND

$U_1$  is  $T_{u_1}^r$  AND ... AND  $U_M$  is  $T_{u_M}^r$

**THEN**  $X = A^r X + B^r U$

Where  $X = [x_1 \ x_2 \ \dots \ x_n]^T$  is the inner is the inner state vector of the nonlinear system,

$U = [u_1 \ u_2 \ \dots \ u_n]^T$  is the input vector to the system, and N, M are the dimensions;

$T_{x_i}^r, T_{u_i}^r$  are linguistic terms (fuzzy sets) defining the conditions for  $x_i$  and  $u_i$  respectively, according to Rule r;

$A^r = (a_{ij}^r)_{N \times N}$  is a matrix of  $N \times N$  and

$B^r = (b_{ij}^r)_{N \times M}$  Of  $N \times M$

When considered in discrete time, such as modeling using a digital computer, we often use the discrete state-space equations instead of the continuous version. Concretely, the fuzzy rules become:

Rule r:

IF  $X_1(t)$  is  $T_{x_1}^r$  AND ... AND  $X_n(t)$  is  $T_{x_n}^r$  AND

$U_1(t)$  is  $T_{u_1}^r$  AND ... AND  $U_M(t)$  is  $T_{u_M}^r$

**THEN**  $X(t+1) = A^r X(t) + B^r U(t)$

Where  $X = [x_1(t) \ x_2(t) \ \dots \ x_n(t)]^T$

is the discrete sample of state vector at discrete time t.

In following discussion we shall use the latter form of rules.

In both forms, the output of the system is always defined as:

$$Y = CX \text{ ( or } Y(t) = CX(t) \text{ )} \quad (1).$$

Where  $C = (c_{ij})_{P \times N}$  is a matrix of  $P \times N$ , and P is the dimension of output vector Y.

The fuzzy inference procedure is specified as below. First, we use multiplication as operation AND to get the firing strength of Rule r:

$$f_r = \prod_{i=1}^N \mu_{T_{x_i}^r} [x_i(t)] \cdot \prod_{i=1}^M \mu_{T_{u_i}^r} [u_i(t)] \quad (2)$$

Where  $\mu_{T_{x_i}^r}$  and  $\mu_{T_{u_i}^r}$  are the membership functions of  $T_{x_i}^r$  and  $T_{u_i}^r$  respectively? After normalization of the firing strengths, we get (assuming R is the total number of rules)

$$S = \sum_{r=1}^R f_r \quad , h_r = f_r / S \quad (3)$$

Where S is the summation of firing strengths of all the rules, and  $h_r$  is the normalized firing strength of Rule r. When the defuzzification is employed, we have

$$X^r(t+1) = A^r X(t) + B^r U(t),$$

$$X(t+1) = \sum_{r=1}^R h_r X^r(t+1) \quad (4)$$

$$= \sum_{r=1}^R h_r [A^r X(t) + B^r U(t)]$$

$$= (\sum_{r=1}^R h_r A^r) X(t) + (\sum_{r=1}^R h_r B^r) U(t)$$

$$= AX(t) + BU(t)$$

Where  $A = (\sum_{r=1}^R h_r A^r)$ ,  $B = (\sum_{r=1}^R h_r B^r)$

Using equation (4), the system state transient equation, we can calculate the next state of system by current state and input.

## 2.2 The structure of temporal Neuro-Fuzzy System

The main idea of this model is to combine simple feed forward fuzzy systems to arbitrary hierarchical models.

The structure of recurrent Neuro-fuzzy systems is presented in figure 3:

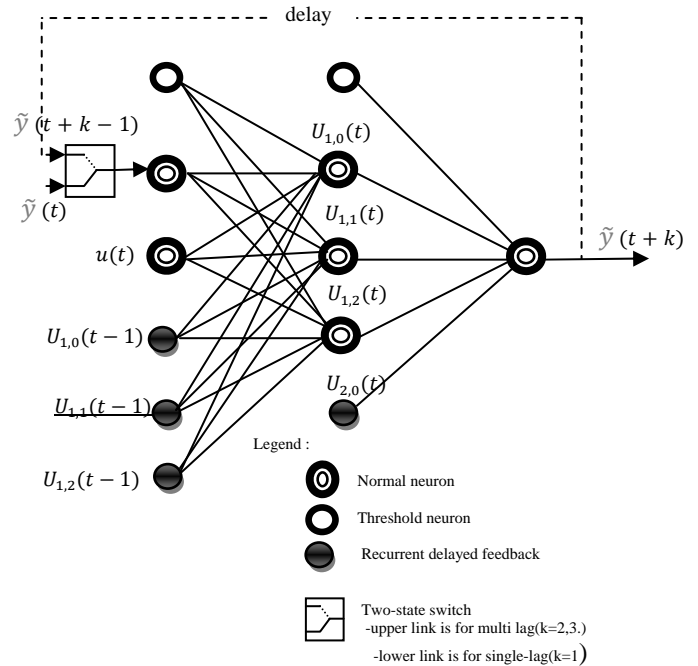


Fig 4. The structure of a simple TNFS

In this network, input nodes which accept the environment inputs and context nodes which copy the value of the state- space vector from layer 3 are all at layer 1 (the Input Layer). They represent the linguistic variables known as  $u_j$  and  $x_i$  in the fuzzy rules. Nodes at layer 2 act as the membership functions, translating the linguistic variables from layer 1 into their membership degrees. Since there may exist several terms for one linguistic variable, one node in layer 1 may have links to several nodes in layer 2, which is accordingly named as the term nodes. The number of nodes in the Rule Layer (layer 3) and the one of the fuzzy rules are the same - each node represents one fuzzy rule and calculates the firing strength of the rule using membership degrees from layer 2. The connections between layer 2 and layer 3 correspond with the antecedent of each fuzzy rule. Layer 4, as the Normalization Layer, simply does the normalization of the firing strengths. Then with the normalized firing strengths  $h_r$ , rules are combined at layer 5, the Parameter Layer, where A and B become available. In the Linear System Layer, the 6th layer, current state vector  $X(t)$  and input vector  $U(t)$  are used to get the next state  $X(t+1)$ , which is also fed back to the context nodes for fuzzy inference at time  $(t+1)$ . The last layer is the Output Layer, multiplying  $X(t+1)$  with C to get  $Y(t+1)$  and outputting it.

Next we shall describe the feed forward procedure of TNFS by giving the detailed node functions of each layer, taking one node per layer as example. We shall use notations like  $u_i^{[k]}$  to denote the  $i^{\text{th}}$  input to the node in layer k, and  $o^{[k]}$  the output of the node in layer k. Another issue to mention here is the initial values of the context nodes. Since TNFS is a recurrent network, the initial values are essential to the temporal output of the network. Usually they are preset to 0, as zero-state, but non-zero initial state is also needed for some particular case.

*Layer 1.* There is only one input to each node at layer 2. The Gaussian function is adopted here as the membership function:

$$o^{[1]} = e^{-\frac{(u^{[1]} - c^r)^2}{2(s^r)^2}} \quad (5)$$

where  $c^r$  and  $s^r$  give the center (mean) and width(variation) of the corresponding  $u^{[1]}$  linguistic term of input  $u^{[2]}$  in Rule r.

*Layer 2.* this layer has several nodes, one for figuring matrix A and the other for B. Though we can use many nodes to represent the components of A and B separately, it is more convenient to use matrices. So with a

little specialty, its weights of links from layer 4 are matrices  $A^r$  (to node for A) and  $B^r$  (to node for B). It is also fully connected with the previous layer. The functions of nodes for A and B are respectively.

$$for A \quad o^{[2]} = \sum_{r=1}^R u_r^{[2]} A^r, for B \quad o^{[2]} = \sum_{r=1}^R u_r^{[2]} B^r \quad (6)$$

*Layer 3.* the Linear System Layer has only one node, which has all the outputs of layer 1 and layer 2 connected to it as inputs. Using matrix form of inputs and output, we have [see (3)]

$$o^{[3]} = AX + BU = o_{for A}^{[2]} o_{context}^{[1]} + o_{for B}^{[2]} o_{context}^{[1]}$$

So the output of layer 3 is  $X(t+1)$  in (4).

This proposed network structure implements the dynamic system combined by our discrete fuzzy rules and the structure of recurrent networks. With preset human knowledge, the network can do some tasks well. But it will do much better after learning rules from teaching examples. In the next section, a learning algorithm will be put forth to adjust the variable parameters in FENN, such as  $c^r, s^r, A^r, B^r$ , and C.

### 3. Proposed Architecture for Fault diagnosis and Prognosis

Faults are usually the main cause of loss of productivity in the process industry. This section uses a straightforward architecture to detect, isolate and identify faults.

One of the most important types of systems present in the process industry is workshop of SCIMAT clinker . A fault in a workshop of SCIMAT clinker may lead to a halt in production for long periods of time. Apart from these economic considerations faults may also have security implications. A fault in an actuator may endanger human lives, as in the case of a fault in an elevator's emergency brakes or in the stems position control system of a nuclear power plant. The design and performance testing of fault diagnosis systems for industrial process often requires a simulation model since the actual system is not available to generate normal and faulty operational data needed for design and testing, due to the economic and security reasons that they would imply.

Figure 5 shows a view and the schematics of a typical industrial industrial process of manufacture of cement. This installation belongs to cement factory of Ain-Touta (SCIMAT) ALGERIA. This cement factory have a capacity of 2.500.000 t/an " Two furnaces " is made up of several units which determine the various phases of the manufacturing process of cement. The workshop of cooking gathers two furnaces whose flow clinker is of 1560 t/h. The cement crushing includes two crushers of 100t/h each one. Forwarding of cement is carried out starting from two stations, for the trucks and another for the coaches.

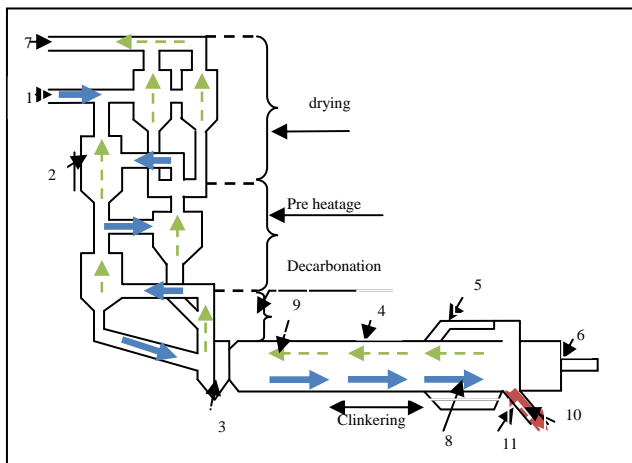


Fig 5. Workshop of SCIMAT clinker

### 3.1 Faults

The workshop of SCIMAT clinker may be affected by a number of faults. These faults are grouped into four major categories: heating tower faults, Kiln Cycling faults, cooler balloons faults and gas burner faults. Here only abrupt or incipient faults are considered.

This step has an objective of the identification of the dysfunctions which can influence the mission of the system. This analysis and recognition are largely facilitated using the structural and functional models of the installation. For the analysis of the dysfunctions we adopted the method for the analysis of the dysfunctions we adopted the method of Failure Modes and Effects Analysis and their Criticality (FMEAC).

While basing itself on the study carried out by [6], on the cooking workshop, we worked out an FMEAC by considering only the most critical modes of the failures (criticality >10), and for reasons of simplicity [46]. Therefore we have a Neuro-fuzzy system of 27 inputs and 11 outputs which were used to make a Prognosis of our system. The rules which are created with the system are knowledge a priori, a priori the base of rule. Each variable

having an initial partition will be modified with the length of the phase of training (a number of sets fuzzy for each variable). The reasoning for the diagnosis and prognosis is described in the form of fuzzy rules inside our Neuro-fuzzy system.

Table 4.1 faults description

Fault	Description	Inceptient/ Abrupt
F1	Chute de la jupe	I/A
F2	bouillage	I/A
F3	No break	I/A
F4	Transporteur à auget	I/A
F5	Presence anneaux	I
F6	Mauvaise homogénéisation	I/A
F7	Chute de croûtage	I/A
F8	Atteinte des briques réfractaires	I
F9	bouillage	I/A
F10	Moteur ventilateur tirage	I/A
F11	Courroies ventilateur tirage	I/A

Our TNFS must have a number of inputs equal to the number of variables sensor signals providing the ability to extend the timing window used for this problem have 27 inputs nodes comprised of 11 sensors signals at 4 successive time points at steps of 10 minutes, resulting in a temporal window of 40 minutes for each sensor .

The TNFS provides 14 outputs representing the 14 possible classes (faults): 11 process faults, 3 sensor faults and normal state.

### 3.2 Training TNFS

To train the TNFS ,we used scenario for each of the 11 possible faults. The process was simulated for 120 minutes, with the faults starting to appear after 40 minutes of normal operation. So, we had 9 different positions of the temporal window (0-40 mins,10-50 mins, etc..), providing 342 input/output vector pairs for training.

NEFDIAG(Neuro-Fuzzy Diagonosis) is a data processing program for interactive simulation. The NEFDIAG development was carried out within LAP (University of Batna), was primarily dedicated to the creation, the training, and the test of a Neuro-Fuzzy system for the classification of the breakdowns of a dedicated industrial process. NEFDIAG models a fuzzy classifier Fr with a whole of classes  $C = \{c1, c2, \dots, cm\}$ [45].

NEFDIAG makes its training by a set of forms and each form will be affected (classified) using one of the preset classes. Next NEFDIAG generates the fuzzy rules by: evaluating of the data, optimizing the rules via training and using the fuzzy subset parameters, and partitioned the data into forms «characteristic» and classified with parameters of the data. NEFDIAG can be used to classify a new observation. The system can be represented in the form of fuzzy rules

**If** symptom1(t) is  $A_1$       Symptom2(t-2) is  $A_2$   
       Symptom3(t) is  $A_3$       Symptom<sub>N</sub>(t-1) is  $A_n$   
**Then** the form  $(x_1, x_2, x_3, \dots, x_n)$  belongs to class «fault i».

For example  $A_1 A_2 A_3 A_n$  are linguistic terms represented by fuzzy sets. This characteristic will make it possible to complete the analyses on our data, and to use this knowledge to classify them. The training phase of the networks of artificial Neuro-Fuzzy systems makes it possible to determine or modify the parameters of the network in order to adopt a desired behavior. The stage of training is based on the decrease in the gradient of the average quadratic error made by network RNF[44].

The NEFDIAG system typically starts with a knowledge base comprised of a collection partial of the forms, and can refine it during the training. Alternatively NEFDIAG can start with an empty base of knowledge. The user must define the initial number of the functions of membership for partitioning the data input fields. And it is also necessary to specify the number K, which represents the maximum number of the neurons for the rules which will be created in the hidden layer. The principal steps of the training algorithm.

The data set used in this experiment contained 200 samples. Each data sample consisted of 27 features comprising the temperature and pressure measurements at various inlet and outlet points of the rotary kiln, as well as other important parameters as shown in Table 4.2. The heat transfer conditions were classified into two categories, i.e., the process of heat transfer was accomplished either efficiently or inefficiently.

From the database, there were 101 data samples (50.18%) that showed inefficient heat transfer condition, whereas 99 data samples (49.82%) showed efficient heat transfer condition in the rotary kiln. The data samples were equally divided into three subsets for training, prediction and test.

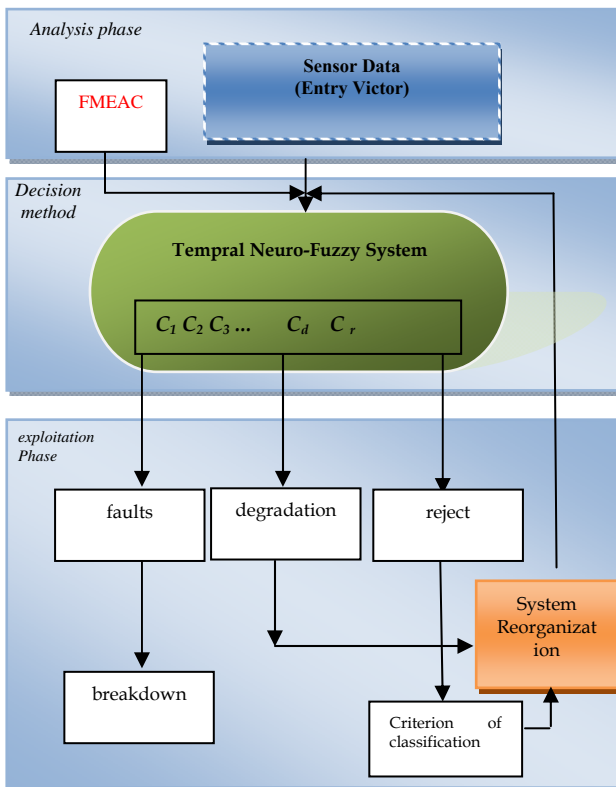


Fig. 6. The diagnosis by NEFDIAG.

Table 4.2 input and output variables for the rule compiling

Input var	Description
CO	CO in the first combustion chamber
Temp	Temp in the first combustion chamber
O2	O2 in the second combustion chamber
RPM	Rotary kiln rotating RPM
Press	Pressure in the first combustion chamber
output var	Description
$\Delta Burner$	Change in burner heating power. i.e. $burner(t)=burner(t-1)+\Delta Burner(t)$
$\Delta Air$	Change in input air quantity ; i.e. $Air(t)=air(t-1)+\Delta Air(t)$
$\Delta IDFan$	Change in induced fan inducing power i.e. $IDFan(t)=IDFan(t-1)+\Delta IDFan(t)$

Usually, the structure of TNFS is determined by trial-and-error in advance for the reason that it is difficult to consider the balance between the number of rules and desired performance [20]. In this study, to determine the structure of TNFS, first we convert numeric data into information granules by fuzzy clustering. The number



of clusters defines the number of fuzzy rules. By applying the fuzzy C-means clustering method [13,40] on the training data and checking the validity measure suggested in [13] it was identified that an adequate number of clusters is 4. Therefore 4 fuzzy rules were used for the basis for training and further refining. The clustering algorithm identified the following cluster centers for the presented data.

IF  $y(t-2)$  is A1 AND  $y(t-1)$  is B1 AND  $y(t)$  is C1 THEN  $y(t+1)$  is D1  
 IF  $y(t-2)$  is A2 AND  $y(t-1)$  is B2 AND  $y(t)$  is C2 THEN  $y(t+1)$  is D2  
 IF  $y(t-2)$  is A3 AND  $y(t-1)$  is B3 AND  $y(t)$  is C3 THEN  $y(t+1)$  is D3  
 IF  $y(t-2)$  is A4 AND  $y(t-1)$  is B4 AND  $y(t)$  is C4 THEN  $y(t+1)$  is D4

Initial fuzzy terms A1, A2, A3, A4 were created from the component  $y(t-2)$  of the cluster vectors 1, 2, 3, and 4, respectively. Similarly, terms B1, B2, B3, B4 – from  $y(t-1)$ , C1, C2, C3, C4 – from  $y(t)$ , and D1, D2, D3, D4 – from  $y(t+1)$ . The terms A1, A2, ..., B1, B2, ..., C1, C2, ..., D1, D2, ... are described linguistically.

Figure 7 and 8 show the response of the normal model output and the real output from five to fifteen minutes prediction horizon and figure 9 to 10 show the response of the fault model output and the real output from three to seven minutes prediction horizon for test data.

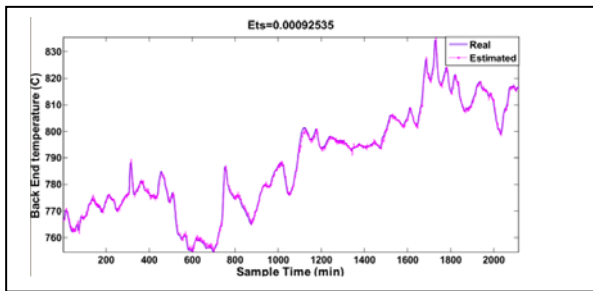


Fig. 7. Normal model with 5 min prediction horizon

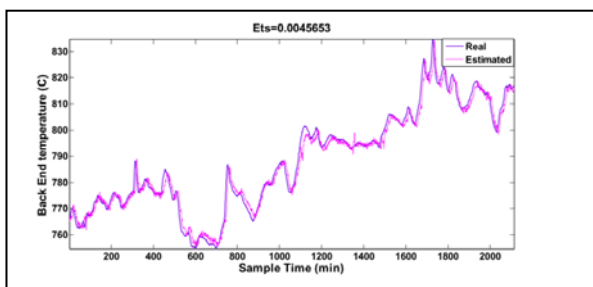


Fig. 8. Normal model with 10 min prediction horizon

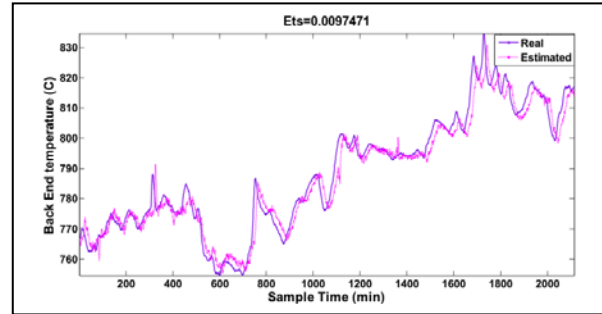


Fig. 9. Normal model with 15 min prediction horizon

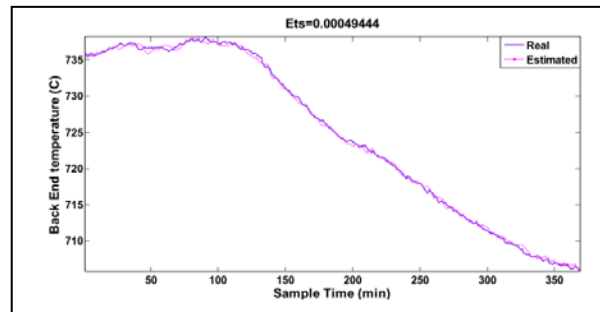


Fig.10 . Failure model with 10 or 15 min prediction

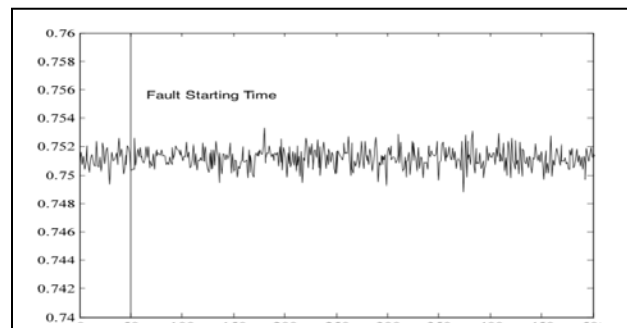


Fig. 11. Effect of incipient fault F10 on the Rotary kiln rotating RPM

## 4. Conclusion

The intelligent process operation aid system was developed to prevent the faults or errors in the process of manufacture of cement. This installation belongs to cement factory of Ain-Touta (SCIMAT) ALGERIA.

In order to do this work, the rule based and temporal Neuro-fuzzy system was implemented.

This TNFS was used for identification, prediction and detection of the fault process in the cement rotary kiln, back end temperature was used as the process monitor of the various conditions. The special character of this variable is that it can show the normal and abnormal conditions inside the kiln.

In spite of great importance of fuzzy neural networks for solving wide range of real-world problems, unfortunately, little progress has been made in their development.

We have discussed recurrent neural networks with fuzzy weights and biases as adjustable parameters and internal feedback loops, which allows capturing dynamic response of a system without using external feedback through delays. In this case all the nodes are able to process linguistic information.

As the main problem regarding fuzzy and recurrent fuzzy neural networks that limits their application range is the difficulty of proper adjustment of fuzzy weights and biases, we put an emphasize on the TNFS training algorithm.

## References

1. Ayoubi M (1994) Fault diagnosis with dynamic neural structure and application to a turbocharger. In: Proceedings of 1st IFAC Symposium SAFEPROCESS'94, Espoo, Finland, vol. 2, pp. 618-623
2. Martineau S, Gaura E, Burnham KJ and Haas OCL (2002) Neural network control approach for an industrial furnace. In: Proceedings of the 14th International Conference on Systems Science, Las Vegas, USA, pp. 227-233
3. Arton V and Palade V (2005) Human-like fault diagnosis using a neural network implementation of plausibility and relevance. *Neural Computing & Applications* 14(2):149-165
4. Babuska R (2002) Neuro-fuzzy methods for modeling and identification. In: Abraham A, Jain LC and Kacprzyk J (eds) *Recent Advances in Intelligent Paradigms and Applications*, pp. 161-186, Springer-Verlag, Heidelberg
5. Lopez-Toribio CJ, Patton RJ and Daley S (2000) Takagi-Sugeno Fault-Tolerant Control of an Induction Motor. *Neural Computation and Applications* 9: 19-28, Springer-Verlag, London, UK
6. Palade V, Patton RJ, Uppal FJ, Quevedo J and Daley S (2002) Fault Diagnosis of An Industrial Gas Turbine Using Neuro-Fuzzy Methods. In: Proceedings of the 15th IFAC World Congress, 21-26 July, Barcelona, pp. 2477-2482
7. Terstyánszky G and Kovács L (2002) Improving fault diagnosis using proximity and homogeneity measure. In: Preprints of the 15th IFAC World Congress, Barcelona, Spain
8. Himmelblau DM (1978) *Fault Detection and Diagnosis in Chemical and Petrochemical Processes*. Elsevier, Amsterdam.
9. Bocaniala CD, Sa da Costa J and Palade V (2005) Fuzzy-based refinement of the fault diagnosis task in industrial devices. *International Journal of Intelligent Manufacturing*, 16(6): 599-614
10. Maciej Grzenda · Andres Bustillo · Pawel Zawistowski (2010) A soft computing system using intelligent imputation strategies for roughness prediction in deep drilling *International Journal of Intelligent Manufacturing*, 22(2):120-131.
11. Koscielny JM and Syfert M (2003) Fuzzy logic applications to diagnostics of industrial processes. In: *SAFEPROCESS'2003*, Preprints of the 5th IFAC Symposium on fault detection, supervision and safety for technical processes, Washington, USA, pp. 771-776.
12. Xi F, Sun Q, Krishnappa G (2000) Bearing Diagnostics Based on Pattern Recognition of Statistical Parameters. *Journal of Vibration and Control* 6:375-392
13. Patton RJ, Frank PM and Clark RN (2000) *Issues of Fault Diagnosis for Dynamic Systems*. Springer, London
14. Chia-Feng Juang (2002) A TSK-Type Recurrent Fuzzy Network for Dynamic Systems Processing by Neural Network and Genetic Algorithms, *IEEE Transactions on Fuzzy Systems*, vol. 10, no. 2
15. Bocaniala CD, Sa da Costa J and Palade V (2004) A Novel Fuzzy Classification Solution for Fault Diagnosis. *International Journal of Fuzzy and Intelligent Systems* 15(3-4):195-206
16. Marinai L (2004) Gas path diagnostics and prognostics for aero-engines using fuzzy logic and time series analysis (PhD Thesis). School of Engineering, Cranfield University.
17. Bocaniala CD and Sa da Costa J (2004) Tuning the Parameters of a Fuzzy Classifier for Fault Diagnosis. Hill-Climbing vs. Genetic Algorithms. In: *Proceedings of the Sixth Portuguese Conference on Automatic Control (CONTROLO 2004)*, 7-9 June, Faro, Portugal, pp. 349-354.
18. Jing He (2006), neuro-fuzzy based fault diagnosis for nonlinear processes (PhD Thesis). the university of new brunswick.
19. kasuma bin ariffin (2007) on neuro-fuzzy applications for automatic control, supervision, and fault diagnosis for water treatment plant (PhD Thesis). Faculty of Electrical Engineering Universiti Teknologi Malaysia
20. jonas biteus (2005) distributed diagnosis and simulation based residual generators (PhD Thesis). Vehicular Systems Department of Electrical Engineering Linköping universitet, SE - 581 83 Linköping, Sweden
21. Faisal J Uppal & Ron J Patton (2002) , Fault Diagnosis of an Electro-pneumatic Valve Actuator Using Neural Networks With Fuzzy Capabilities
22. [David Henry](#) , [Xavier Olive](#) , [Eric Bornschlegel](#) (2011) A model-based solution for fault diagnosis of thruster faults: Application to the rendezvous phase of the Mars Sample Return mission, *European conference for aero-space sciences (EUCASS'2011)*, Russie, Fédération .
23. Lakhmi Jain, Xindong Wu (2006) , *Computational Intelligence in Fault Diagnosis* , Springer-Verlag London Limited.
24. george vachtsevanos, frank lewis, michael roemer, andrew hess, biquing wu (2006), *intelligent fault diagnosis and prognosis for engineering systems*. Published by John Wiley & Sons, Inc., Hoboken, New Jersey
25. Silvio Simani, Cesare Fantuzzi and Ron J. Patton (2002), model-based fault diagnosis in dynamic systems using identification techniques, Springer-Verlag Berlin
26. Christopher Edwards, Thomas Lombaerts, and Hafid Smaili (2010), *Fault Tolerant Flight Control: A Benchmark Challenge*, 2010 Springer-Verlag Berlin Heidelberg.
27. János Fodor and Janusz Kacprzyk, (2009), *Aspects of Soft Computing, Intelligent Robotics and Control*, Springer-Verlag Berlin Heidelberg.
28. Lixiang Shen , Francis E.H. Tay , Liangsheng Qu , Yudi Shen ,(2000), Fault diagnosis using Rough Sets Theory, *Computers in Industry*, elsevier.

29. Yunfei Zhou, Shuijin Li \*, Rencheng Jin,(2002), A new fuzzy neural network with fast learning algorithm and guaranteed stability for manufacturing process control, *Fuzzy Sets and Systems*, elsevier.
30. Krzysztof Patan ,( 2008) *Artificial Neural Networks for the Modelling and Fault Diagnosis of Technical Processes* , Springer-Verlag Berlin Heidelberg
31. Sio-Iong Ao, Burghard Rieger, Su-Shing Chen(2009), *Advances in Computational Algorithms and Data Analysis*, Springer Science+Business Media B.V.
32. Xinsheng Lou, Kenneth A. Loparo,(2004), Bearing fault diagnosis based on wavelet transform and fuzzy inference, *Mechanical Systems and Signal Processing* 18 1077–1095, elsevier
33. Venkat Venkatasubramanian, Raghunathan Rengaswamy, Surya N. Kavuri, Kewen Yin,(2003), A review of process fault detection and diagnosis Part III: Process history based methods, *Computers and Chemical Engineering* 27 327 / 346
34. Marcos E. Orchard and George J. Vachtsevanos (2007), A Particle Filtering Approach for On-Line Failure Prognosis in a Planetary Carrier Plate, *International Journal of Fuzzy Logic and Intelligent Systems* vol7 ,N 4 pp221-227.
35. nuran arzu yilmaz,(2003), a temporal neuro-fuzzy approach for time series analysis, (phd thesis). the department of computer engineering, the middle east technical university
36. michal knotek,(2006), fault diagnostics based on temporal analysis, (phd thesis). l'université joseph fourier
37. Marcin Witczak,(2008), *Modelling and Estimation Strategies for Fault Diagnosis of Non-Linear Systems From Analytical to Soft Computing Approaches*, Springer Berlin Heidelberg New York
38. M. Staroswiecki\*, G. Comtet-Varga,(2001), Analytical redundancy relations for fault detection and isolation in algebraic dynamic systems, *Automatica* 37 687-699, elsevier
39. Bin Zhang, Taimoor Khawaja, Romano Patrick, George Vachtsevanos, Marcos Orchard (2010), A novel blind deconvolution de-noising scheme in failure prognosis, *Transactions of the Institute of Measurement and Control* 32, 1 pp. 3–30
40. Nikola Kasabov, *Evolving Connectionist Systems: The Knowledge Engineering Approach*, Springer-Verlag London Limited 2007.
41. Nicolas PALLUAT,(2006) , *Méthodologie de surveillance dynamique à laide des réseaux neuro-flous temporels*, (phd thesis). université de franche comté.
42. Venkat Venkatasubramanian, Raghunathan Rengaswamy, Kewen Yin, Surya N. Kavuri(2003), A review of process fault detection and diagnosis Part I: Quantitative model-based methods, *Computers and Chemical Engineering* 27 293-311. elsevier
43. Rolf Isermann,(2006) *Fault-Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance*, Springer-Verlag Berlin Heidelberg.
44. Rafik Mahdaoui , leyla hayet mouss , med djamel mouss, chouhal ouahiba,(2011), *Temporal Neuro-Fuzzy Systems in Fault Diagnosis and Prognosis For Manufacturing Systems*, *International Review on Modelling and Simulations (I.R.E.M.O.S.)*, Vol. 04, n.1.
45. Rafik mahdaoui ,(2008) , le diagnostic industriel par approche Neuro-Floue, these de magister , université de batna.
46. Mahdaoui Rafik and Mouss Hayet Leyla and Chouhal Ouahiba and Houassi Hichem and kadri Ouahab ,(2009), *Industrial dynamics monitoring by Temporals Neuro-Fuzzy systems: Application to manufacturing system* ICGST International Journal on Automatic Control and System Engineering GVIP.
47. M. Aenis , E. Knopf, R. Nordmann,(2002), Active magnetic bearings for the identification and fault diagnosis in turbomachinery *Mechatronics* 12 1011–1021 , elsevier.



**First Authorname**

MAHDAOUI Rafik holds the engineer Diploma from the Computer Science department, Hadj LAKhdhar University of Batna, Algeria in 2001; he obtained the degree of Magister in 2008 at Batna University in Industrial engineering. Since 2009 he is an assistant Professor at Computer Science department of Khenchela university Algeria , where he teaches: Programming languages, graph theory, script languages and others matters. He supervises engineers and masters students on their final projects. he is a member of secure operating systems Group, at LAP Laboratory, Hadj LAKhdhar University of Batna, where he is preparing a PHD diploma. he's research interests are in Neuro-Fuzzy systems , Artificial intelligence, emergent technologies , prognosis and diagnosis , e-maintenance.

# An Efficient Stream Cipher Algorithm for Data Encryption

Majid Bakhtiari<sup>1</sup> Mohd Aizaini Maarof<sup>2</sup>

<sup>1</sup> Department of Computer Science & Information Systems, University Technology Malaysia,  
City Campus Jalan Semarak, 54100 Kuala Lumpur, Malaysia

<sup>2</sup> Department of Computer Science & Information Systems, University Technology Malaysia,  
Skudai Johor Bahru, 81310 Malaysia

## Abstract

Nowadays the data telecommunication security has been provided by most of well-known stream cipher algorithms which are already implemented in different secure protocols such as GSM, SSL, TLS, WEP, Bluetooth etc. These algorithms are A5/1, A5/2, E0 and RC4. On the other hand, these public algorithms already faced to serious security weakness such that they do not provide enough security of proportional plain data in front of cryptanalysis attacks. In this paper we proposed an efficient stream cipher algorithm which generates 23 random bits in each round of processing by parallel random number generator and 115 bits of Initial Vector. This algorithm can implement in high speed communication link more than 100Mb/s and it has passed all of standard cryptographic tests successfully, also it can resist in front of well-known attacks such as algebraic and correlation.

**Keywords:** *Stream Ciphers, GSM, SSL, WEP, A5/1, A5/2, E0, data telecommunication, cryptanalysis attacks.*

## 1. Introduction

Stream cipher algorithms are being used in a wide range of information processing applications. This kind of cryptography is symmetric encryption primitives which are widely applied for providing the confidentiality of different networks. Currently, the public communication security is supported by well-known secure protocols such as GSM, WEP, SSL, TLS, Bluetooth, etc. These protocols are supported by four stream cipher algorithms which are A5/x in GSM networks [1], E0 in Bluetooth standard [2] and RC4 in SSL, TLS and WEP (802.11 wireless LAN standard) [3]. On the other hand, there are many practical attacks discovered on all mentioned encryption algorithms [4-6].

Stream ciphers are always faster than block ciphers but due to the nature of random number generators which have been used in well-known stream ciphers, there are confronting with many threatening problems that permits unauthorized persons to easily access on public privacy. On the other hand, it is impossible to have infinite state random number generator to generate a truly random

sequence, since the finiteness forces the random sequence to be periodic. Therefore, the best that can do is using very long period sequences that called pseudo-random sequences.

With consider that all of linear random number generators are not enough strong in front of algebraic and correlation attacks, it is necessary to notice that the linear part of algorithm should isolate from the output part of algorithm which generates key stream. However, some encryption algorithms are not implemented that part of algorithm like as A5/x, E0, RC4. Currently, the mentioned algorithms have tried to solve the linearity weaknesses by applying nonlinear clocking to those cryptosystems. As a result those algorithms cannot resist in front of algebraic and correlation attacks. It should notice that A5/1, A5/2 and E0 do not completely protect the linear part of random generator which is threatening stream ciphers.

Basically, stream ciphers should have one part as internal state and some of update function to update internal state for each round of process. The internal state, mostly initialized by secret key and initial vector (IV) key, then generates long key-stream, that known as a pseudo-random sequence. The internal state must be located behind of output part of the algorithm that generates random sequence for plain data engagement. This consideration is an important subject that some of current stream ciphers do not follow. The non-compliance with this rule causes to generate serious security problems for those stream ciphers which do not follow up like as A5/x, E0 and RC4 faced to. However, the Boolean functions in the output part of random generators must have good non linearity properties in order to resist in front of many cryptanalysis attacks such as algebraic, correlation and known IV attacks [7].

Another problem that the most of stream ciphers faced to is that each of them generates just one random bit in each round of process as the output stream of cryptosystem. This feature increases the risk of algebraic and



correlation attacks against those cryptosystems. In this paper, an efficient stream cipher algorithm designed in such a way that can generate 115 random bits in one round of process. This feature increases the resistance in front of Berlekamp-Massey, algebraic and correlation attacks.

In this paper, three public stream ciphers algorithms (A5/1, A5/2 and E0) are explained briefly, then a new stream cipher algorithm has designed in two sections as Parallel Random Number Generator and Read Out Combiner Function. The designed algorithm can generate 23 random bits in a round of processing, also it can resist in front of algebraic and correlation attacks. The designed algorithm can easily implement by software and hardware. This algorithm has passed successfully all of important cryptographic tests that mentioned in National Institute of Standards and Technology (NIST).

## 2. Background

Some of the most popular stream cipher algorithms which now cover more than 80% of the world of telecommunication and cyber space are A5/1, A5/2, E0 and RC4. These algorithms are weak in front of different kinds of cryptanalysis attacks such as correlation and algebraic attacks and etc. In this section their structure of those algorithms which are working on the base of LFSRs briefly explained.

### 2.1 A5/1 Stream Cipher Algorithm

The A5/1 is one of the stream cipher algorithm that currently is using by the most countries around the world in order to ensure privacy of conversations on GSM mobile phones. The A5/1 consists of 3 shift registers named R1, R2 and R3 with method of majority clocking as shown in Figure 1. The initialization of registers will be done by 64-bit  $K_c$  and 22-bit frame number which these are first shifted into the left side of all 3 registers and XORed with the feedbacks. Then A5/1 is clocked by using the majority clocking for 100 cycles to initial mix the bits. Then, the next 114-bits of output from A5/1 is XORed with the plaintext to encrypt/decrypt.

There are several kinds of attacks are listed on A5/1 in section 1 in [8]. One of them is the method of Biryukov [9]. He found a known-key stream attack on A5/1 requiring about two second of the key stream and recovers  $K_c$  in a few minutes on a personal computer. The second one is the method of Barkan [8]. He proposed a cipher-text-only attack on A5/1 that can recover  $K_c$  by using only four frames of cipher text.

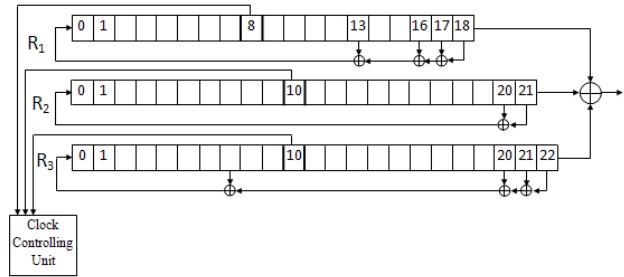


Figure 1: The A5/1 stream cipher algorithm.

Alex Biryukov, Adi Shamir and David Wagner presented that it is possible to find the A5/1 key in less than a second on a single PC, by analyzing the output of the A5/1 algorithm in the first two minutes of the conversation [9]. It is because the output key stream generated by just linear function named XOR.

In 2008, Timo Gendrusis and his team presented a guess-and-determine attack on the A5/1 stream cipher by running on the special-purpose hardware device named COPACOBANA [10]. It reveals the internal state of the cipher in less than 6 hours on average needing only 64 bits of known key stream [11].

### 2.2 A5/2 Stream Cipher Algorithm

The A5/2 is the 2nd stream cipher algorithm that currently support by GSM protocol in many countries. In 2006 Elad Barkan, Eli Biham and Nathan Keller demonstrated attacks against A5/1 and A5/2, that allow attackers to tap GSM mobile phone conversations and decrypt them either in real-time, or at any later time. The protocol weaknesses of GSM allow to recovery of the secret key. According to survey on the attacks against A5/2 stream cipher algorithm, it has been determined that exist linear relations among the output sequence bits and the vast majority of the unknown output bits can be reconstructed. Furthermore, some researcher have shown the time complexity of the attack is proportional to  $2^{17}$  [12]. While according on GSM declaration the complexity of A5/2 should be  $2^{64}$ .

In 2007 Ian Goldberg and David Wagner of the University of California at Berkeley published an analysis of the weaker A5/2 algorithm showing a work factor of  $2^{16}$ , or approximately 10 milliseconds. Elad Barkan, Eli Biham and Nathan Keller of Technion, the Israel Institute of Technology, have presented a cipher-text-only attack against A5/2 that requires only a few dozen milliseconds of encrypted off-the-air traffic. They also described new attacks against A5/1 and A5/3 [13].



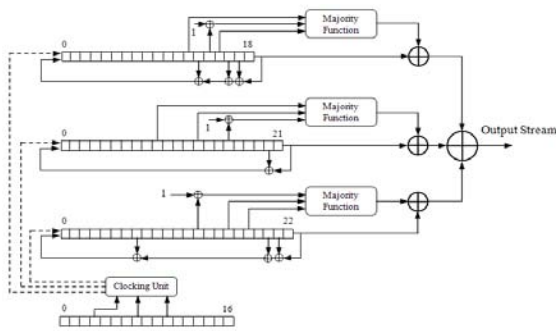


Figure 2: The A5/2 stream cipher algorithm.

With point of comparison to the previous algorithms, the description of KASUMI is public and based on the block cipher MISTY. However, the A5/3 algorithm is so far believed to be stronger than A5/1 and A5/2 but an attack successfully has done by Biham [8]. He presented that the key could be found faster than exhaustive key search [1].

### 2.3 E0 Stream Cipher Algorithm

Bluetooth protocol is an open standard for short-range digital radio. The goal of Bluetooth is to connect devices (PDAs, cell, phones, printers, faxes, etc.) together wirelessly in a small environment such as an office or home. The Bluetooth has three different encryption modes to support the confidentiality service as follows:

- Mode 1: No encryption is performed on any data.
- Mode 2: Broadcast traffic is not encrypted, but the individually addressed traffic is encrypted according to the individual link keys.
- Mode 3: All traffic is encrypted according to the master link key.

Bluetooth is working on the base of E0 algorithm. Until now, there are many known attacks on the encryption scheme E0 are available that can threaten the security of Bluetooth. The most well-known of them are algebraic attacks [14] and correlation attacks [15-16].

E0 generates a bit using four shift registers with differing lengths (25, 31, 33, 39 bits). The Figure 3 shows the involved algorithm use in the Bluetooth standard.

However, in E0 like A5/1 and A5/2, the last function that generates key stream is simple XOR. Due to the linear properties of XOR, the output key stream has linear relation with its inputs that it may threaten the whole of algorithm.

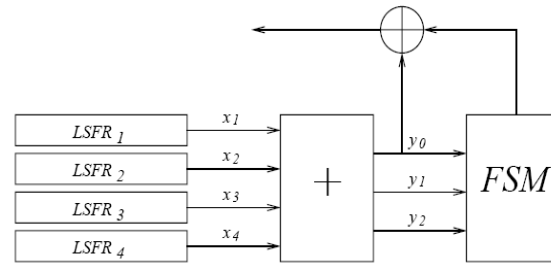


Figure 3: The encryption algorithm used in Bluetooth.

## 3. New Stream Cipher Algorithm

One of the important parameter that A5/1, A5/2 and E0 suffering from is using from XOR at the last section of algorithm to generate key stream. This method of random bit generation has caused that those algorithm faced to big security problems. In this paper, a new algorithm has designed in such a way that key stream can resist in front of correlation and algebraic attacks.

### 3.1 Parallel Random Number Generator

Linear feedback shift registers (LFSRs) are very applicable in parallel random number generators. Due to the simplicity of implementing the LFSRs in structure of hardware and software, LFSRs are use in many of random number generators. LFSRs can generate different sequences with good statistical properties and large length of period. With notice that, the equation of polynomial feedback plays very important role in LFSRs. If the feedback polynomial equation is primitive, it means that an LFSR with length  $n$  can generate maximal length of sequence equal to  $2^n - 1$ . Furthermore, due to feature of output linearity stream, the output sequences of LFSR are easily expectable and if the designers want to trap more than one stream as outputs of sequence, each bit is exactly equal to others bits with time delay (maximum delay is length of LFSR-bit or  $n$ ). This problem is threatening the system from different viewpoint especially from correlation attack. In this paper, one model of LFSR has designed in such a way that can solve this big problem; hence it is important in cryptography.

Designing a parallel random number generator (PRNG) by using one LFSR has the feature that one can construct a linear sequential system which is correctly initialized and for each clock cycle generates different consecutive stream of the sequences, while the normal LFSR would generate just one stream sequence. In fact, each bit-output of the finite state machine can be XORed together to form the key-stream output.

LFSRs are defined by characteristic polynomials which determine all properties of the sequences produced by an LFSR. Parallel Random Number Generators (PRNGs) are defined by very specific polynomials. The most properties of this kind of generators are that those have been used in practice and at large scale of encryption in symmetric cryptography. The estimation of the number of primitive polynomials in PRNGs related to LFSRs can be calculated from Eq. (1), where  $v$  is the number of sub-registers.

$$P(x) = \frac{2^{1+2\lfloor \frac{2n}{v} \rfloor} \cdot \varphi(2^n - 1)}{n \cdot 2^n} \quad (1)$$

This class of primitive polynomials is very strong properties on the parallel implementation of an LFSR. The basic idea of a parallel generator consists in generating the sub-sequences of a given sequence in parallel. However, this is a basic technique for taking advantage of parallel computer.

Let  $S = (S_0, S_1, S_2, \dots)$  be an unlimited binary sequence with period  $T$ , thus  $S_j \in \{0,1\}$  and  $S_{j+T} = S_j$  or all  $j \geq 0$ . For a given integer  $d$ , a  $v$ -decimation of  $S$  is the set of sub-sequences defined in Eq. (2).

$$S_v^i = (S_i, S_{i+v}, S_{i+2v}, \dots, S_{i+jv}, \dots) \quad (2)$$

where  $i \in \{0, d-1\}$  and  $j = 0,1,2, \dots$ . Consequently, the sequence  $S$  is completely described by the sub-sequences as follows:

$$\begin{aligned} S_v^0 &= (S_0, S_v, \dots) \\ S_v^1 &= (S_1, S_{v+1}, \dots) \\ S_v^2 &= (S_2, S_{v+2}, \dots) \\ &\vdots \\ S_v^{v-2} &= (S_{v-2}, S_{2v-2}, \dots) \\ S_v^{v-1} &= (S_{v-1}, S_{2v-1}, \dots) \end{aligned}$$

Usually the strong random generators are structured by combining of more than one LFSR which are working together with different methods to provide non-linearity in output stream. This paper do not follows the classical methods of generation such as  $n$ -sequences that explained by Colomb. Mostly, the classical methods for generating  $n$ -sequences is using a primitive polynomial to select those taps of an  $n$ -cell shift register which, if their contents are added modulo 2 and the summation used as input to the shift register, will result in a cycle length of  $2^n - 1$  steps. In this paper, the special combination of LFSR has designed in such a way that each output traps are not equal to others outputs. With consider that, this paper need 115 separated random sequences with maximum period of length which should be isolated from each others. This paper designed a LFSR in such a way that can provide 115 separated random stream sequences. On the other hand,

the speed of processing is important parameters that it is possible to implement designed PRNG by software and hardware to obtain suitable speed of processing. In this regard, the Eq. (3) initially designed as a primitive candidate polynomial equation that can be satisfied the form of Eq. (4).

$$\begin{aligned} P(x) = & x^{257} + x^{254} + x^{251} + x^{249} + x^{244} + x^{243} + x^{242} + x^{238} + \\ & x^{237} + x^{233} + x^{232} + x^{230} + x^{228} + x^{226} + x^{225} + x^{221} + x^{220} + \\ & x^{218} + x^{216} + x^{214} + x^{213} + x^{209} + x^{208} + x^{204} + x^{203} + x^{202} + \\ & x^{197} + x^{195} + x^{192} + x^{190} + x^{187} + x^{185} + x^{180} + x^{179} + x^{178} + \\ & x^{174} + x^{173} + x^{169} + x^{168} + x^{166} + x^{164} + x^{162} + x^{161} + x^{157} + \\ & x^{156} + x^{154} + x^{152} + x^{150} + x^{149} + x^{145} + x^{144} + x^{140} + x^{139} + \\ & x^{138} + x^{133} + x^{131} + x^{128} + x^{126} + x^{123} + x^{121} + x^{116} + x^{115} + \\ & x^{114} + x^{110} + x^{109} + x^{105} + x^{104} + x^{102} + x^{100} + x^{98} + x^{97} + \\ & x^{93} + x^{92} + x^{90} + x^{88} + x^{86} + x^{85} + x^{81} + x^{80} + x^{76} + x^{75} + x^{74} + \\ & x^{69} + x^{67} + x^{64} + x^{62} + x^{59} + x^{57} + x^{52} + x^{51} + x^{50} + x^{46} + x^{45} + \\ & x^{41} + x^{40} + x^{38} + x^{36} + x^{34} + x^{33} + x^{29} + x^{28} + x^{26} + x^{24} + x^{22} + \\ & x^{21} + x^{17} + x^{16} + x^{12} + x^{11} + x^{10} + x^5 + x^3 + 1 \end{aligned} \quad (3)$$

According to the output of parallel random number generator by LFSRs which should be on the base of Eq. (4), after analyzing the Eq. (3), we find that it is equal to Eq. (5). So, both of equation are equal to each other from period of length and sequence bit-randomness point of view.

$$\begin{aligned} P(x) = & x^n + (x^m + 1)^k * (x^j + 1)^l * (x^i + 1)^y \\ & ; ((m * k) + (j * l) + (i * y)) < n \end{aligned} \quad (4)$$

$$P(x) = x^{257} + (x^2 + 1)^{100} (x^3 + 1)^{13} (x^5 + 1)^3 \quad (5)$$

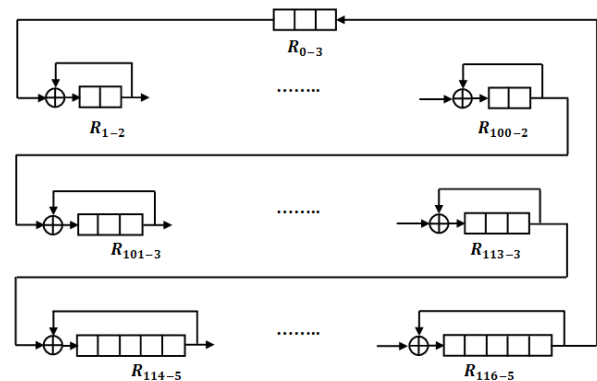


Figure 4: Diagram of Parallel Random Number Generator.

On the other hand, we need 23 random bit-streams that should be isolated from each other as the output of PRNG. The Eq. (5) has potential of parallel random bit-stream generation and it can generate 116 isolated bit-stream sequences. The period length of each sequence is equal to  $2^{257} - 1$ . It is because the characteristic polynomial equation of diagram that has shown in Figure 4 is primitive. In this regard, we have simulated Eq. (5) in

Figure 4. In fact, the diagram that shown in Figure 4 is equal to Eq. (5) or Eq. (3). Therefore, we can select just 115 traps as bit-stream output from Figure 4. However, it is advisable to implement a multiplexer for sequence selection to increase the nonlinearity of each stream. Finally, each sequence can be selected according to Eq. (6) as follows:

$$\begin{aligned} S_v^{Ri} &= (S_v^{Ri}, S_{v+1}^{Ri}, \dots) ; 1 \leq i \leq 31 ; 1 \geq v \geq 3 \\ S_w^{Rj} &= (S_w^{Rj}, S_{w+1}^{Rj}, \dots) ; 1 \leq j \leq 23 ; 1 \geq w \geq 5 \\ S_u^{Rk} &= (S_u^{Rk}, S_{u+1}^{Rk}, \dots) ; 1 \leq k \leq 5 ; 1 \geq u \geq 7 \end{aligned} \quad (6)$$

It is important to notice that if someone wants to implement one multiplexer to increase the degree of nonlinearity of stream sequence. It should apply on parameters of  $i, j, k$  in Eq. (6).

### 3.2 Read Out Combiner Function

The read out combiner as an important part of algorithm, plays very critical role in front of different kinds of attacks. This part is first part of algorithm that located in front of cryptanalyst. This part of algorithm designed in such a way that can resist in faced to strong methods of attacking such as correlation and algebraic attacks.

The first step is designing the truth table of read out combiner function. The truth table should be designed in such a way that can satisfy all of features related to cryptography point of view such balanced-ness, correlation immunity, algebraic degree and non-linearity. The truth table of this part of algorithm has shown in Appendix as Table 1 and the proportional Boolean function Eq. (7) is as follows:

$$\begin{aligned} f(x) = & \bar{A}\bar{B}\bar{C}\bar{D}\bar{E} + \bar{A}\bar{B}\bar{C}\bar{D}E + \bar{A}\bar{B}\bar{C}D\bar{E} + \bar{A}\bar{B}C\bar{D}\bar{E} + \\ & \bar{A}\bar{B}C\bar{D}E + \bar{A}\bar{B}CDE + \bar{A}BC\bar{D}\bar{E} + \bar{A}BCD\bar{E} + \\ & \bar{A}BCDE + A\bar{B}\bar{C}\bar{D}\bar{E} + A\bar{B}\bar{C}\bar{D}E + A\bar{B}\bar{C}D\bar{E} + A\bar{B}C\bar{D}\bar{E} + \\ & A\bar{B}C\bar{D}E + ABC\bar{D}\bar{E} + ABCD\bar{E} + ABED\bar{E} \end{aligned} \quad (7)$$

After designing the truth table as an output of Boolean function, the characteristic of designed table has simplified as an equation that has shown in Eq. (8).

$$\begin{aligned} f(x) = & \bar{A}\bar{B}\bar{C}\bar{D} + \bar{A}\bar{B}E(C \oplus D) + \bar{A}\bar{B}\bar{C}D + \\ & \bar{A}BC(\bar{D} \oplus \bar{E}) + \bar{A}\bar{B}\bar{C}\bar{E} + AC(D \oplus E) + \\ & ABCDE + ABC\bar{D}\bar{E} \end{aligned} \quad (8)$$

It should be notice that in this paper, the read out combiner consists of 23 functions that all of them are same together but all of their inputs are different from each others. Each function has five separate inputs from PRNG. Therefore, 115-bits stream from PNRG after XOR with Initial Vector (IV), feed to functions of read out combiner to generate 23

bits as output of algorithm. So, each bit of key stream output is simplifies as Eq. (8).

For convenient explanation of Eq. (8), the Figure 5 shows the Eq. (8) as function box with totally 115-bits. In fact, the functionality of Figure 5 is exactly Eq. (8). Therefore, the functionality of  $f(x)$  is a function with 5 input variables and 1-bit output that operates instead of Eq. (8). The important statistical cryptography tests have applied on  $f(x)$ .

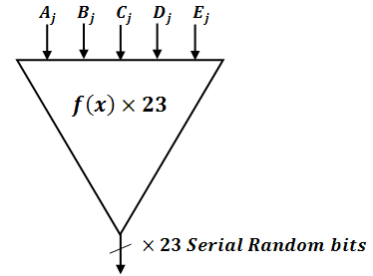


Figure 5: Read-Out Combiner Function.

#### 3.2.1 Balance Check

With consider that, a binary sequence is called balanced if its truth table has the same number of 1's and 0's. According to Table 1 in Appendix, the function of  $f(x)$  is balanced as shown in Eq. (8). The balanced-ness of a Boolean function is a significant cryptography property in the manner that the output of function should not leak any statistical information related to the crypto system.

#### 3.2.2 Nonlinearity Check

Both non-linear and linear functions are significance for block and stream ciphers. Non-linear functions are usually used to achieve confusion, while linear functions are employed to achieve diffusion. Non-linear functions are useful in protecting a cipher system from a differential cryptanalysis, and determining the key by solving equations and so on. The non-linearity is the number of bits which must change in the truth table of a Boolean function to reach the closest affine function.

There are different kinds of nonlinearity measurement methods available. In this paper, the non-linearity of Eq. (8) has calculated from affine function and Walsh spectrum. Therefore, the non-linearity of the Eq. (8) with 5-variable Boolean function  $N_f$  calculated as follows:

$$\begin{aligned} N_f &= 2^{n-1} - \frac{1}{2} \max |\omega_f(a)| , a \in \{0,1,2, \dots, 2^n-1\} \\ W_f &= 16,4,4,4, -4,4, -4,0,0,0,4,0,0,4,0,0, -4,4,4, \\ & 0,0, -4,4,4,0, -4,0,0,0,0, -4,0 \end{aligned}$$

$$N_f \leq 2^{n-1} - 2^{\frac{n}{2}-1} \Rightarrow N_f = 12 \quad (9)$$

The higher non-linearity in Boolean function means, that the designed function can protecting the cipher system in faced to some methods of attacks particularly algebraic attack. With consider that the non-linearity of designed function Eq. (8) that has shows as the read out combiner function in Figure 5, has calculated by Eq. (9). Also the maximum degree for five variables should be equal to 12, it is means that the designed function can protect the cipher system from two serious attacks such as correlation and algebraic. Currently well-known stream cipher algorithms are suffering from these kinds of attacks.

### 3.2.3 Correlation Immunity Check

Cryptographers care about correlation immunity because its absence in Boolean functions which are used in a cryptosystem can allow effective attacks on the system. According to the test result of designed function from probability point of view, the result for all of variables is equal to  $\frac{1}{2}$ . Therefore the designed function is correlation immune.

$$\begin{aligned} CI(A) &= \frac{1}{2}, & CI(B) &= \frac{1}{2}, & CI(C) &= \frac{1}{2}, \\ CI(D) &= \frac{1}{2}, & CI(E) &= \frac{1}{2} \end{aligned} \quad (10)$$

As it has shown in Eq. (10), which have derived from Table 1 (in Appendix), the result check of correlation immunity is excellent for designed function. On the other hand, from correlation calculations point of view, we have calculated all of possibility of designed function. All of results are equal to zero. It is because the correlation coefficients have boundaries of -1 and +1. A value of +1 indicates perfect positive linear relationship between two sequences, while -1 is a perfect negative linear relationship between them. A value of zero indicates no correlation between input variables or independent. In designed function, the correlations for five variables are excellent. Therefore highly non-linear balanced Boolean function with an excellent Correlation-Immunity is enough strong in faced to correlation attack.

### 3.2.4 Algebraic Degree Check

The algebraic degree is one of the nonlinearity measures of Boolean function. The Boolean functions with small algebraic degree are in general considered to be less suitable for cryptographic applications than those with higher degree. However there are large classes of cryptographically strong Boolean functions with small algebraic degree such as quadratic bent functions. It is

important that almost every balanced Boolean function has maximal or almost maximal algebraic degree.

The algebraic degree of Eq. (8) is equal to 4 which is the maximum level for 5 variables. Therefore, each output random bit of this function can successfully resist in faced to algebraic attack and Berlekamp-Massey attacks.

## 4. Practical Statistical Tests

According to National Institute of Standards and Technology (NIST), all important cryptography tests (Frequency test, Serial test, Run, Long Run test, Poker test, Auto-Correlation test, Maurer's Universal Test) have applied on designed stream cipher algorithm. All of tests passed successfully.

## 5. Conclusion

In this paper we introduce some weaknesses of well-known stream cipher algorithms in current industrial world which are threatening public interests in different cyber space networks. According to many sources and serious security weaknesses in well-known stream cipher algorithms which are already implemented in GSM, SSL, TLS, WEP, Bluetooth and so on, it is strongly advise not to rely on E0, A5/x and RC4 in field of data security communication.

Furthermore, an efficient designed stream cipher algorithm can be implemented in GSM, WEP, SSL, TLS and Bluetooth protocols. The new algorithm has designed base on parallel random number generator with the high speed of processing which can be implemented in high speed data/voice link of communication and it can resist in front of different kinds of attacks such as correlation and algebraic.

The designed algorithm has passed all of cryptographic tests in NIST standard successfully. The designed new algorithm can support the encryption/decryption with rate of 100 MB/s. The key variety of designed algorithm is equal to  $2^{257}$  and the length key of IV is equal to  $2^{115}$ . It can be implemented easily by hardware and software.

This paper designed a new stream cipher algorithm with key variety of  $2^{257}$  and 115-bit IV that is more secure than other public one from speed of processing and others viewpoint of security.

## Appendix

Table 1: Truth table of Nonlinear Function

<i>Input variety</i>	<i>Output</i>
00000	1
00001	1
00010	0
00011	1
00100	0
00101	1
00110	0
00111	0
01000	0
01001	0
01010	1
01011	1
01100	1
01101	0
01110	0
01111	1
10000	1
10001	0
10010	1
10011	0
10100	0
10101	1
10110	1
10111	0
11000	0
11001	0
11010	0
11011	1
11100	1
11101	1
11110	1
11111	0

## References

- [1] Briceno, M., I. Goldberg, and D. Wagner, *A pedagogical implementation of A5/1*. URL: <http://www.scard.org/gsm/a51.html>.
- [2] Bluetooth, S., *Specification of the Bluetooth system*. Core, version, 2005. 1: p. 2005-10.
- [3] Rivest, R., *The RC4 Encryption Algorithm*. RSA Data Security. Inc., March, 1992. 12.
- [4] Maximov, A., T. Johansson, and S. Babbage, *An Improved Correlation Attack on A5/1*, in *Selected Areas in Cryptography*, H. Handschuh and M. Hasan, Editors. 2005, Springer Berlin / Heidelberg. p. 1-18.
- [5] Stubblefield, A., J. Ioannidis, and A. Rubin, *A key recovery attack on the 802.11 b wired equivalent privacy protocol (WEP)*. ACM transactions on information and system security (TISSEC), 2004. 7(2): p. 319-332.
- [6] Golić, J., V. Bagini, and G. Morgari, *Linear Cryptanalysis of Bluetooth Stream Cipher*, in *Advances in Cryptology — EUROCRYPT 2002*, L. Knudsen, Editor. 2002, Springer Berlin / Heidelberg. p. 238-255.
- [7] Meier, W. and O. Staffelbach, *Nonlinearity Criteria for Cryptographic Functions*, in *Advances in Cryptology — EUROCRYPT '89*, J.-J. Quisquater and J. Vandewalle, Editors. 1990, Springer Berlin / Heidelberg. p. 549-562.
- [8] Barkan, E., E. Biham, and N. Keller, *Instant Ciphertext-Only Cryptanalysis of GSM Encrypted Communication*. Journal of Cryptology, 2008. 21(3): p. 392-429.
- [9] Biryukov, A., A. Shamir, and D. Wagner, *Real Time Cryptanalysis of A5/1 on a PC*, in *Fast Software Encryption*, G. Goos, et al., Editors. 2001, Springer Berlin / Heidelberg. p. 37-44.
- [10] Gendrullis, T., M. Novotný, and A. Rupp, *A Real-World Attack Breaking A5/1 within Hours*, in *Cryptographic Hardware and Embedded Systems – CHES 2008*, E. Oswald and P. Rohatgi, Editors. 2008, Springer Berlin / Heidelberg. p. 266-282.
- [11] Kumar, S., et al., *Breaking Ciphers with COPACOBANA –A Cost-Optimized Parallel Code Breaker*, in *Cryptographic Hardware and Embedded Systems - CHES 2006*, L. Goubin and M. Matsui, Editors. 2006, Springer Berlin / Heidelberg. p. 101-118.
- [12] Petrovic, S. and A. Fuster-Sabater. *An improved Cryptanalysis of the A5/2 Algorithm for Mobile Communications*. 2002.
- [13] Barkan, E., E. Biham, and N. Keller, *Instant Ciphertext-Only Cryptanalysis of GSM Encrypted Communication*, in *Advances in Cryptology - CRYPTO 2003*. 2003, Springer Berlin / Heidelberg. p. 600-616.
- [14] Armknecht, F. and M. Krause, *Algebraic Attacks on Combiners with Memory*, in *Advances in Cryptology - CRYPTO 2003*. 2003, Springer Berlin / Heidelberg. p. 162-175.
- [15] Hermelin, M. and K. Nyberg, *Correlation Properties of the Bluetooth Combiner*, in *Information Security and Cryptology - ICISC'99*, J. Song, Editor. 2000, Springer Berlin / Heidelberg. p. 17-29.
- [16] Lu, Y. and S. Vaudenay. *Faster correlation attack on Bluetooth keystream generator E0*. 2004: Springer.



# Rectangular Patch Antenna Performances Improvement Employing Slotted Rectangular shaped for WLAN Applications

Mouloud Challal<sup>1,2</sup>, Arab Azrar<sup>1</sup> and Mokrane Dehmas<sup>1</sup>

<sup>1</sup> Signals and Systems Laboratory, Institute of Electrical Engineering and Electronics, IGEE, University of Boumerdes Boumerdes, 35000, Algeria

<sup>2</sup> ICTEAM, Electrical Engineering, Université catholique de Louvain, Louvain-La-Neuve, 1348, Belgium

## Abstract

This paper describes the effect of inserting a rectangular shape defected ground structure (DGS) into the ground plane of the conventional rectangular microstrip patch antenna (CRMPA). The performances of the CRMPA are characterized by varying the dimensions of the rectangular slot (RS-DGS) and also by locating the RS-DGS at specific position. Simulation results have verified that the CRMPA including RS- DGS had improved the CRMPA without RS-DGS. The return loss (RL) enhances approximately of 100 %, and gain improvement of 0.8 dB.

*Keywords:* Conventional Rectangular Microstrip Patch Antenna (CRMPA), Rectangular Slot Defected Groud Structure (RS-DGS, Return Loss (RL), Gain, Radiation pattern.

## 1. Introduction

Recently, there has been a growing demand of microwave, and wireless communication systems in various applications resulting in an interest to improve antenna performances. Modern communication systems and instruments such as Wireless local area networks (WLAN), mobile handsets require lightweight, small size and low cost. The selection of microstrip antenna technology can fulfill these requirements [1]. WLAN in the 2.4 GHz band (2.4-2.483 GHz) has made rapid progress and several IEEE standards are available namely 802.11a, b, g and j [1]. Various design techniques using defected ground structure (DGS) in the patch antenna have been suggested in previous publications [2-4]. DGS is realized by etching a defect in the ground plane of planar circuits and antennas. This defect disturbs the shield current distribution in the ground plane and modifies a transmission line such as line capacitance and inductance characteristics [5]. Accordingly, a DGS is able to provide a wide band-stop characteristic in some frequency bands with a reduced number of unit cells. Due to their excellent pass and rejection frequency band characteristics [5], DGS

circuits are widely used in various active and passive microwave and millimeter-wave devices [6].

The purpose of this work is to enhance conventional rectangular microstrip patch antenna (CRMPA) performances operating at 2.4 GHz frequency band for WLAN applications using Rectangular Slot (RS) in the ground plane named RS-DGS. Configurations using RS-DGS located at different positions in the bottom of the substrate are considered and assessment of the new rectangular microstrip patch antennas performances achieved.

## 2. Antenna Design

A CRMPA is designed on a dielectric layer RO4003C substrate which has a relative permittivity and thickness of 1.524 mm. As shown in Figure 2.a, the patch antenna has a length (L) of 30 mm and a width (w) of 21 mm and its resonant frequency is 2.40 GHz. The resonant frequency, also called the center frequency, is selected as the one at which the return loss is minimum. An etched RS-DGS with different length values and a fixed width (3.5 mm) is then inserted into the ground plane of the original CRMPA shown in figure 1 (Ant.1) at different positions as shown in figure 2.a (Ant.2), figure 2.b (Ant.3) and figure 2.c (Ant.4).

In Figure 2, the RS-DGS is drawn with dash lines to indicate that it is located on the bottom of the substrate. Except the insertion of a rectangular shape slot to the ground plane, no other modification has been performed to the antenna patch and the feeding system.

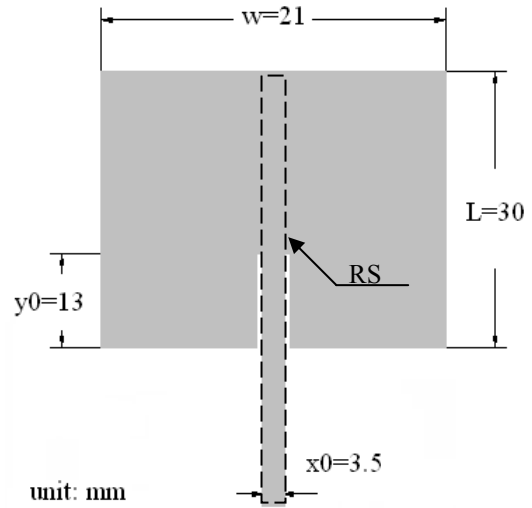
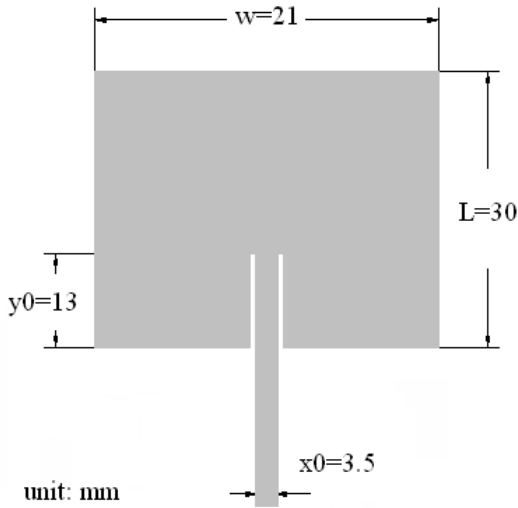


Fig. 1 Conventional Rectangular microstrip patch antenna (Ant. 1)

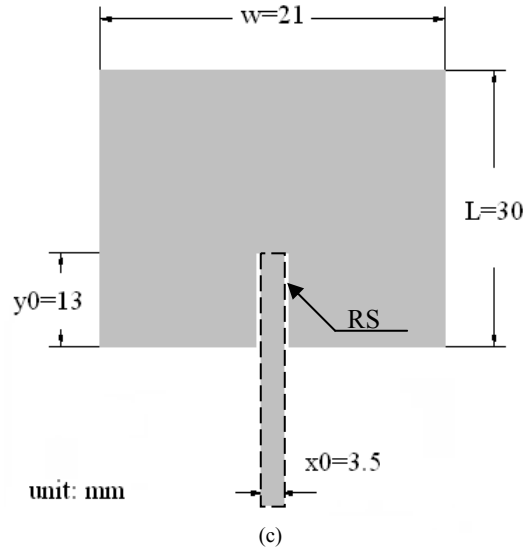
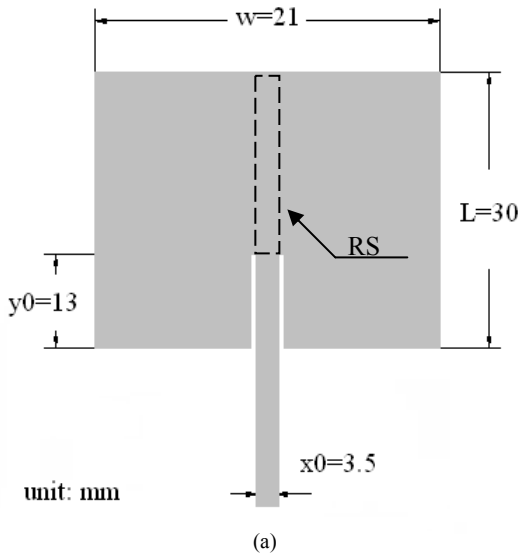


Figure 2. Rectangular microstrip patch antenna with rectangular slot (a) RS\_DGS\_1 (Ant. 2), (b) RS-DGS\_2 (Ant. 3) and (c) RS-DGS\_3 (Ant. 4)

The design and simulation are carried out over four RMPA types; CRMPA and the new modified model antenna by including RS-DGS located at different positions as shown in Figure 2a (Ant. 1), Figure 2b (Ant. 2), Figure 2c (Ant. 3) and, Figure 2d (Ant. 4). The simulations are carried out with IE3D from Zeland software which is based on the method of moments. The software is available in the microwave laboratory of UCL –Belgium.

With a specific resonant frequency ( $f_0$ ) and a characteristic impedance ( $Z_c$ ), the width ( $W$ ), length ( $L$ ) and the Feeding position of CRMSA are expressed as follows [7-8]:

$$W = \frac{c}{2f_0 \sqrt{\frac{\epsilon_r + 1}{2}}} \quad (1)$$

$$L = L_c - 2\Delta L \quad (2)$$

$$y_0 = \frac{L}{\pi} \times \sqrt{\arccos\left(\frac{R_{in}}{R_c}\right)} \quad (3)$$

where,

$$\Delta L = 0,412.h \cdot \frac{(\epsilon_e + 0.3)\left(\frac{W}{h} + 0.264\right)}{(\epsilon_e - 0.258)\left(\frac{W}{h} + 0.8\right)} \quad (4)$$

$$L_e = \frac{\lambda}{2} = \frac{\lambda_0}{2\sqrt{\epsilon_e}} = \frac{c}{2f_0\sqrt{\epsilon_e}} \quad (5)$$

$R_{in}$ ,  $L_e$  and  $\Delta L$  are, respectively, the input impedance, the effective and the extended lengths.

### 3. Results and discussion

Figure 3 shows the simulation result of the return loss (RL) of the CRMPA and the structures with inserted RS-DGS at different positions. This figure shows return losses of -15.72 dB, -14.99 dB, -26.92 dB and -31.87 dB at the resonant frequency of 2.4 GHz and respectively the CRMPA, Ant.2, Ant.3 and Ant. 4.

The simulation carried out with the structure with an RS-DGS implemented in the antenna (Ant.2) shows no significant difference as compared to the CRMPA except a slight shift up of the resonant frequency as illustrated in Figure 3. However, significant improvements are performed when the RS-DGS is implemented as shown in Figure 2.b (Ant. 3) and Figure 2.c (Ant. 4).

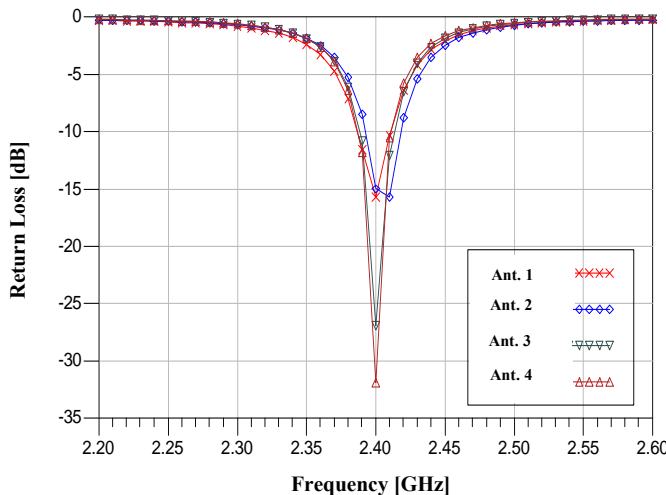


Fig. 3 Return Loss of the CRMPA and the antennas with RS-DGS

Another parameter, namely the gain, is also simulated and the results illustrated in Figure 4. This figure shows a gain

of 5.1 dB for the CRMPA and the insertion of RS-DGS's produces a gain of 5.9 dB for both Ant. 3 and Ant. 4 that is an improvement of 0.8 dB with respect to the antenna without RS-DGS. The gain enhancement justifies the impedance matching of the RL which makes in evidence an enhancement of the antenna efficiency.

Afterward, radiation patterns of the CRMPA in the E and H plane for both with and without RS-DGS are shown in Figure 5 and Figure 6 respectively. The CRMPA radiation patterns are simulated at a frequency of 2.4 GHz. It is observed from these figures that the antennas with RS-DGS have slightly higher lobe level due to the existence of the etched structure in the ground plane acting as a slot antenna resulting in a field distribution.

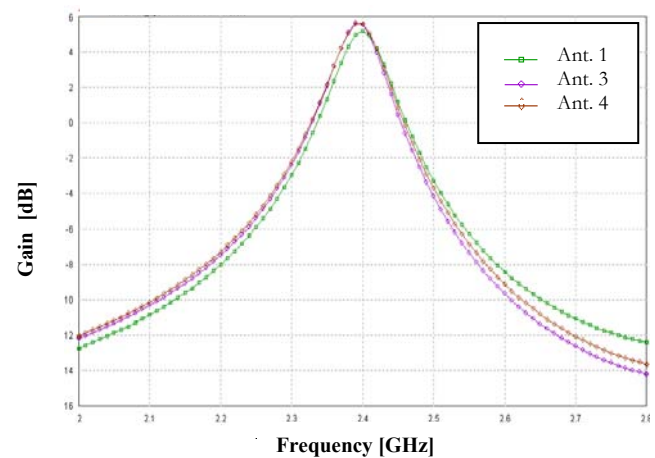


Fig.4 Gains of the CRMPA and the antennas with RS-DGS

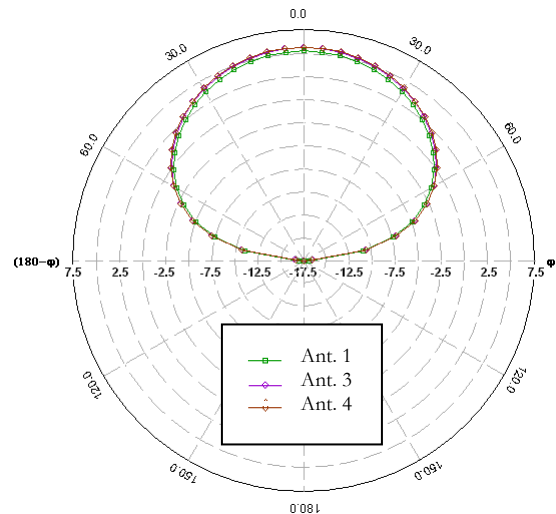


Fig. 5 E-plane radiation patterns of the CRMPA and the antennas with RS-DGS

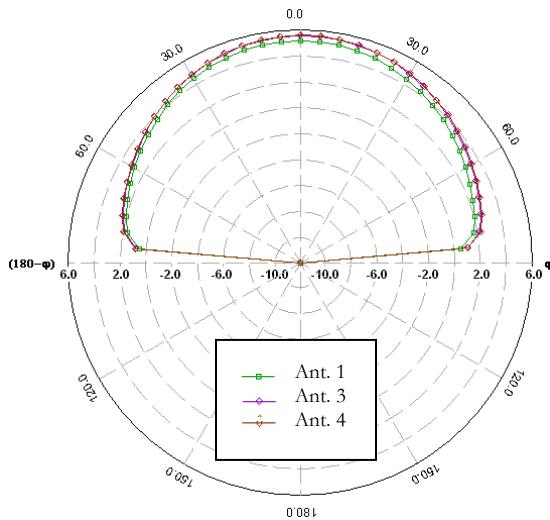


Fig. 6 H-plane radiation patterns of the CRMPA and the antennas with RS-DGS.

Table 1 summarizes the obtained simulation features of the designed antennas.

Table 1: The obtained simulations features

Antennas types	Resonance Freq. [GHz]	Material	RL [dB]	Gain [dB]
Ant. 1 : CRMPA	2.4	RO4003C $\epsilon_r = 3.4$ H=1.524 mm	-15.72	5.1
Ant. 3 : CRMPA with RS-DGS	2.4	RO4003C $\epsilon_r = 3.4$ H=1.524 mm	-26.92	5.9
Ant. 4 : CRMPA with RS-DGS	2.4	RO4003C $\epsilon_r = 3.4$ H=1.524 mm	-31.87	5.9

#### 4. Conclusions

A simple technique to improve conventional rectangular microstrip patch antenna (CRMPA) characteristics by adding an etched rectangular slot in the ground plane (RS-DGS) is presented in this paper. Simulation results have shown that inserting RS-DGS improves the antenna performances. For the considered CRMPA, the results show a 100 % enhancement of the return loss and a 0.8 dB improvement of the gain for the configurations named Ant. 3 and Ant. 4. A further work focusing on the effect of the RS-DGS position and parameters is essential to end up with an antenna configuration with optimal performances.

Moreover, an investigation of various shapes of DGSs is also planned.

#### References

- [1] Yen-Liang Kuo, Kin-Lu Wong, "Printed double-T monopole antenna for 2.4/5.2 GHz dual-band WLAN operations," IEEE Trans. Antennas Propagation, vol. 51, pp. 2187–2192, September, 2003
- [2] M. K. Mandal, P. Mondal, S. Sanyal, and A. Chakrabarty, "An Improved Design Of Harmonic Suppression For Microstrip Patch Antennas", Microwave and Optical Technology Letters, pp. 103-105 Vol. 49, No. 1, January 2007.
- [3] Haiwen Liu, Zhengfan Li, Xiaowei Sun, and Junfa, "Harmonic Suppression With Photonic Bandgap and Defected Ground Structure for a Microstrip Patch Antenna", IEEE Microwave and Wireless Components Letters, VOL. 15, NO. 2, Feb. 2005.
- [4] Y. J. Sung, M. Kim, and Y.-S. Kim, "Harmonics Reduction With Defected Ground Structure for a Microstrip Patch Antenna", IEEE Antennas and Wireless Propagation Letters, VOL. 2, 2003.
- [5] D. Ahn, J. S. Park, C. S. Kim, J. Kim, Y. Qian, and T. Itoh, "A design of the low-pass filter using the novel microstrip defected ground structure," IEEE Trans. Microwave Theory Tech., vol. 49, pp. 86–93, Jan. 2001.
- [6] C. S. Kim, J. S. Park, D. Ahn, and J. B. Lim, "A novel 1-D periodic defected ground structure for planar circuits," IEEE Microwave Guided Wave Lett., vol. 10, pp. 131–133, Apr. 2000.
- [7] C.A. Balanis, "Antenna theory: analysis and design," Third edition, John Wiley & sons Inc., 2005.
- [8] T. A. Milligan, "Modern antenna design," John Wiley & Sons, INC, 2005.



Mouloud Challal was born on March 06th, 1976, in Algiers, Algeria. He received the electronics and communication engineering degree from the Université des sciences et de la Technologie Houari Boumediene, Algiers, Algeria, in April 1999, and the M.Sc. degree in microwave and communication from the Ecole Nationale Polytechnique, Algiers, Algeria, in Dec. 2001. From 1998 to 1999, he acted as computer engineer in private company; in charge of maintenance, computer network installation (LAN), Algiers. From 1999 to 2002, he taught computer science in a public institute (Ex- ITEEM), Algiers. Since 2004, he is a lecturer and a researcher in the Institute of Electrical and Electronics Engineering, (IGEE, Ex. INELEC), University of Boumerdes (UMBB), Boumerdes, Algeria. Since 2007/2008 academic year, is registered as researcher/PhD student at both UMBB and Université catholique de Louvain (UCL), Louvain-a-Neuve, Belgium. His research interests include RF/Microwave circuits, design and analysis of microstrip filters, defected ground structures behaviors, wireless communication systems, microstrip antenna array analysis, synthesis and design. He is an IEEE, EuMA, IAENG and SDIWC Member.



**Arab AZRAR** was born in Takerboust, Bouira, Algeria, on August 2<sup>nd</sup>, 1971. He received the B.S. degree in Electrical and Electronic Engineering from National Institute of Electricity and Electronics of Boumerdes Algeria in 1995 and the MS and doctorate degrees from National Polytechnic school of El-Harrach, Algeria respectively in 1998 and 2004. Currently,

he is a lecturer in the institute of Electrical and Electronic Engineering of Boumerdes University and his fields of interest include Antennas, Propagation, and Microwaves.



**Mokrane DEHMAS** was born in April 1967 in Tizi-Ouzou, Algeria. He received the Engineer and Magister degrees in the National Institute of Electricity and Electronics (INELEC-Boumerdes, Algeria) respectively in 1991 and 1996. He is currently an associate professor in the Institute of Electrical and Electronic Engineering of the University of Boumerdes and a member of the research team in communication systems. His main

fields of interest are semiconductor devices modeling and microstrip radiating structures.



# Semantic annotation of requirements for automatic UML class diagram generation

Soumaya Amdouni<sup>1</sup>, Wahiba Ben Abdesslem Karaa<sup>2</sup> and Sondes Bouabid<sup>3</sup>

<sup>1</sup> University of tunis High Institute of Management  
Bouchoucha city, Bardo 2000, TUNISIA

<sup>2</sup> University of tunis High Institute of Management  
Bouchoucha city, Bardo 2000, TUNISIA

<sup>3</sup> University of tunis High Institute of Management  
Bouchoucha city, Bardo 2000, TUNISIA

## Abstract

The increasing complexity of software engineering requires effective methods and tools to support requirements analysts' activities. While much of a company's knowledge can be found in text repositories, current content management systems have limited capabilities for structuring and interpreting documents. In this context, we propose a tool for transforming text documents describing users' requirements to an UML model. The presented tool uses Natural Language Processing (NLP) and semantic rules to generate an UML class diagram. The main contribution of our tool is to provide assistance to designers facilitating the transition from a textual description of user requirements to their UML diagrams based on GATE (General Architecture of Text) by formulating necessary rules that generate new semantic annotations.

**Keywords:** *annotation, class diagram, GATE, requirements, semantic techniques, software engineering, UML model.*

## 1. Introduction

Increasing complexity of IS (information systems) and their quickly development prompted an increased interest in their study, in order to evaluate their performance in response to users' expectations. There is much and growing interest in software systems that can adapt to changes in their environment or their requirements in order to continue to fulfill their tasks. In fact, requirements specification is a fundamental activity in all process of software engineering. Many Researches [5] notice that many system failures can be attributed to a lack of clear and specific information requirements. Knowledge requirements are formally defined and transferred from some knowledge source to a computer program. It has been argued that requirements study and

knowledge acquisition are almost identical processes. Analysts can use several techniques necessary to extract relevant knowledge for software engineering. These knowledge define system expectations in terms of mission objectives environment, constraints, and measures of effectiveness and suitability. Thus, we need platforms and tools that enable the automation of activities involved in various life cycle phases of software engineering. These tools are very useful to extract functional and non-functional requirements from textual descriptions in order to develop graphic models of application screens, which will assist end-users to visualize how an application will look like after development. The aim of the work presented in this paper is to develop a tool that transforms a textual description to an UML class diagram. Our tool takes as input text data that represent textual user requirements descriptions. First, it identifies named entities (i.e., classes, properties and relationships between classes) and second it classifies them in a structured XML file.

The paper is organized into five sections. Section 2 reviews some related works. Section 3 gives an overview of GATE API. Section 4 discusses our system and the final section presents a conclusion.

## 2. Related works

In the last years several efforts have been devoted by researchers in the Requirements Engineering community to the development of methodologies for supporting designers during requirements elicitation, modeling, and analysis.

However, these methodologies often lack tool support to facilitate their application in practice and encourage companies to adopt them.

The present work is in the context of engineering models such as MDA (Model Driven Architecture) which is a process based on the transformation of models: model to model, code to model, model to code, etc. It presents an experience in the application of requirements specifications expressed in natural language into structured specifications.

The proposed application having an input text data that represent user requirements identifies named entities (entities, properties and relationships between entities ...) to classify them in a structured XML file. Several researchers have tried to automate the generation of an UML diagram from a natural language specification.

Kaiya et al. [8] proposed a requirements analysis method which based on domain ontologies. However, this work does not support natural language processing, it allows the detection of incompleteness and inconsistency in requirements specifications, measurement of the value of the document, and prediction of requirements changes.

In [2] Christiansen et al. developed a system to transform use case diagram to class diagram Definite Clause Grammars extended with Constraint Handling Rules. The grammar captures information about static world (classes and their relations) and subsequently the system generates the adequate class diagram. This work is very interesting but the problem that organization's requirements are not always modeled as use case diagram.

The work in [10] implemented a system named GeNLangUML (Generating Natural Language from UML) which generates English specifications from class diagrams. The authors translate UML version 1.5 class diagrams into natural language. This work was considered by most developers as an efficient solution for reducing the number of errors and verification and an early validation of the system but we need for all time to generate UML diagram from natural language. The system process is as follows:

- Grammatical labeling based on a dictionary wordnet to disambiguate the lexical structure of UML concepts.
- Sentences generation from the specification by checking attributes, operations and associations with reference to a grammar defining extraction rules.
- Checking if the generated sentences are semantically correct.
- Generating a structured document containing the natural specification of a natural class diagram.

Hermida et al. [7] proposed a method which adapts UML class diagrams to build domain ontologies. They describe the process and the functionalities of the tool that they have developed for supporting this process. The authors

have chosen a use case in the pharmacotherapeutic domain. The authors present a good approach however it is specific to a well defined area (pharmacotherapeutic).

In [6] authors proposed a tool NT2OD which derives an initial object diagram from textual use case descriptions using natural language processing (NLP) and ontology learning techniques. NT2OD consists of creating a parse tree for the sentence, identifying objects and relations and generating the object diagram.

In our work we propose a CASE tool (Computer-aided software engineering). We extract information from users' requirements to generate class diagram taking in account existing approaches. We propose a design tool which extracts UML concepts and generate UML class diagram according to different concepts (class, association, attribute). The idea is to use GATE API<sup>1</sup> and we extended it by new JAPE rules to extract semantic information from user requirements.

### 3. GATE overview

GATE "General Architecture for Text Engineering" is developed by the Natural Language Processing Research Group<sup>2</sup> at the University of Sheffield<sup>3</sup>. GATE is a framework and graphical development environment, which enables users to develop and deploy language engineering components and resources in a robust fashion [4]. GATE contains different modules to process text documents. GATE supports a variety of formats (doc, pdf, xml, html, rtf, email...) and multilingual data processing using Unicode as its default text encoding.

In the present work we use the information extraction tool **ANNIE plugin** (A Nearly-New **IE** system) (Fig. 1). It contains Tokeniser, Gazetteer (system of lexicons), Pos Tagger, Sentence Splitter, Named Entity Transducer, and OrthoMatcher.

- Tokeniser: this component identifies various symbols in text documents (punctuation, numbers, symbols and different types). It applies basic rules to input text to identify textual objects.
- Gazetteer: gazetteer component creates annotation to offer information about entities (persons, organizations...) using lookup lists.
- POS Tagger: this component produces a tag to each word or symbol.
- Sentence splitter: sentence splitter identifies and annotates the beginning and the end of each sentence.

---

<sup>1</sup> <http://gate.ac.uk/>

<sup>2</sup> <http://nlp.shef.ac.uk/>

<sup>3</sup> <http://www.shef.ac.uk/>

- Named Entity Transducer: the NE transducer applies JAPE rules to input text to generate new annotations [1].

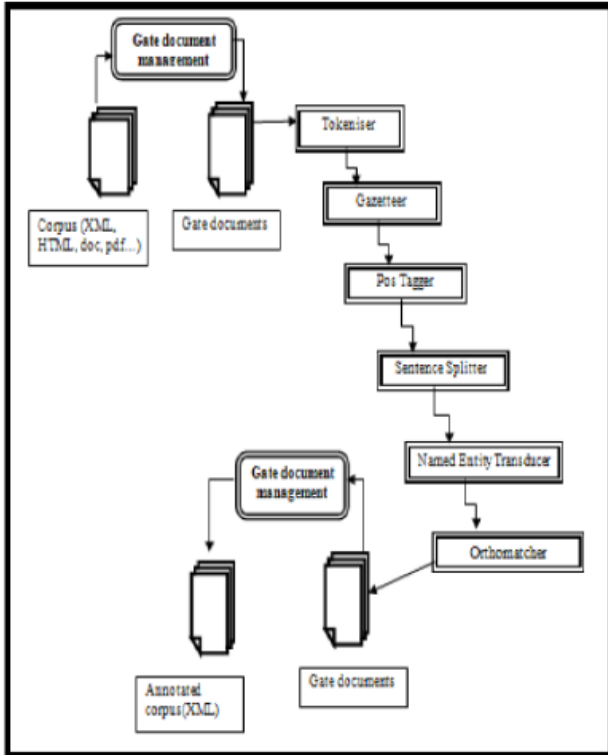


Fig. 1 ANNIE components in GATE.

#### 4. System description

Our system is a document analysis and annotation framework that uses efficient methods and tools adopted from markup domain. The approach discriminates between domains of the annotation process and hence allows an easy adaptation to different applications.

In fact, it uses GATE API and especially the following components: sentence splitter, pos tagger, gazetteer, named entity transducer. The entity recognition is the most interesting task for this reason we extended ANNIE tool with additional rules and additional lists to enhance entities' extraction. The following figure (fig. 2) describes the process we have proposed for the extraction UML concepts in order to generate an UML class diagram.

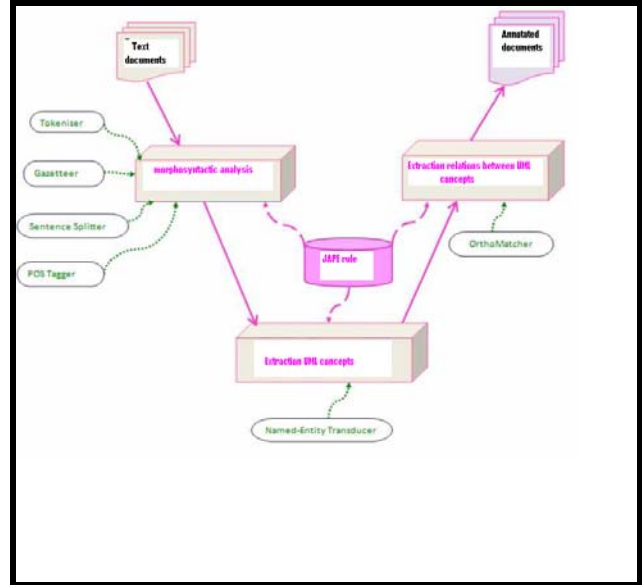


Fig. 2. System architecture.

#### 4.1 Morphological analysis

The first phase of our system consists in morphosyntactic analysis of users' requirements. The tool parses an input document according to a predefined grammar. The produced parse tree consists of structures such as paragraph, sentence, and token. In this step we use sentence splitter and Tokeniser component to extract sentences and basic linguistic entities. Then, we used Pos Tagger to associate with each word (token) grammatical category and to distinguish the morphology of various entities. For example below, the tagger identifies a verb (i.e., passe), two nouns (i.e., client, commande), an, and two prepositions (i.e., le, une ).

**Le (PRP) client (NN) passe (VB) une (PRP) commande (NN).**

#### 4.2 Semantic Extraction of UML concepts

The second phase is extraction of UML concepts. The system is based in the results generated by morphosyntactic analysis stage and uses the Named Entity Transducer component to perform the operation for extracting named entities (classes, attributes and associations) referring to new JAPE rules and Gazetteer lists.

JAPE rule (Java Annotation Patterns Engine) a variant adapted to the Java programming language consists in files containing a set of rules [3]. Gazetteer lists are lookup lists with one entry per line containing names of people large organizations, months of the year, days of the week, numbers, etc [10].

In class diagram usually have the following format:

**Noun+verb+Noun**

The example below demonstrates two classes (le client, une commande), and an association (passe).

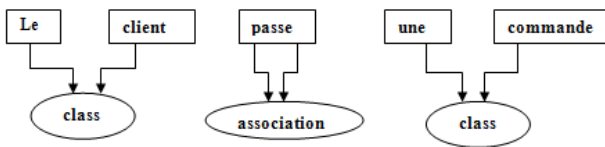


Figure 3 (fig. 3) describes a Jape rule called "Class" allowing the recognition of classes running a well-defined set of instructions. The actual treatment begins from line 7 of this figure by testing whether the word (token) under analysis belongs to a gazetteer list already defined. If this test is checked, the word in question will be annotated "Class" or there will be a passage to the following instructions from line 16.

```

1. Phase:IdentifClasse
2. Input: Lookup Token
3. Options: control =appelt
4. Rule: Classe
5. Priority: 20
6. (
7. {Lookup.majorType ==ClassMinuscule }
8. |
9. {Lookup.majorType==ClassJustPremierMaj}
10. |
11. {Lookup.majorType==ClassPremierMaj}
12. |
13. {Lookup.majorType==ClassMajuscule}
14. |
15. (
16. {Token.kind==word, Token.category==NNP}
17. {Token.kind==word, Token.category==V}
18. )
19. ):label
20. -->
21. :label.classe = {rule= Classe}
    
```

Fig. 3. JAPE rule for extracting UML class.

To extract association concept we use Jape rule illustrated in Figure 4 (fig. 4). If the token belong to gazetteer lists (lines 7, 9, 11, 13), it will be annotated as association otherwise the instructions from line 16 will be executed: if the token belongs to the class list, the second token is a "verb", and that the third word belongs to the list "Class", then the second word (token) will be annotated as an association.

```

1. Phase:IdentifAssociation
2. Input: Lookup Token
3. Options: control = appelt
4. Rule: Association
5. Priority: 20
6. (
7. {Lookup.majorType == AssociationMinuscule }
8. |
9. {Lookup.majorType== AssociationTTMajuscule}
10. |
11. {Lookup.majorType== AssociationMajJustD eb}
12. |
13. {Lookup.majorType== AssociationMajD eb }
14. |
15. (
16. {Lookup.majorType == class }
17. {Token.kind==word, Token.category==V}
18. {Lookup.majorType == class }
19. )
20. ):label
21. -->
22. :label.Association={rule=Association}
    
```

Fig. 4 JAPE rule for extracting association.

In addition, we execute instructions in figure 5 (fig. 5) to extract attribute. This rule is running as precedent ones (JAPE rule extracting class, and JAPE rule extracting association). If the token fits in attribute lists so it will have an attribute annotation. Else if the token is a name following by a verb and another name not belonging in the class list the latter is identified as an attribute.

```

1. Phase:IdentifAttribut
2. Input: Lookup Token
3. Options: control = appet

4. Rule: Attribut
5. Priority: 20
6. (
7. {Lookup.majorType ==attribut Mini}
8. |
9. {Lookup.majorType==attribut TT MAJ}
10.|
11.{Lookup.majorType==Attribut Maj Déb Chak Mot}
12.|
13.{Lookup.majorType==attribut justDébutMaj}
14.|
15.(
16. {Token.kind==word, Token.category==NNP}
17. {Token.kind==word, Token.category==V}
18. {Token.kind==word,
    Token.category==NNP},{Lookup.majorType != class }
19.)
20.): label
21.->
22.:label Attribut = {rule= Attribut}
    
```

Fig. 5 JAPE rule for extracting attribute.

In this step, we propose a graphical representation of JAPE rules set used by all modules of ANNIE components that we have integrated in our application (fig. 6).

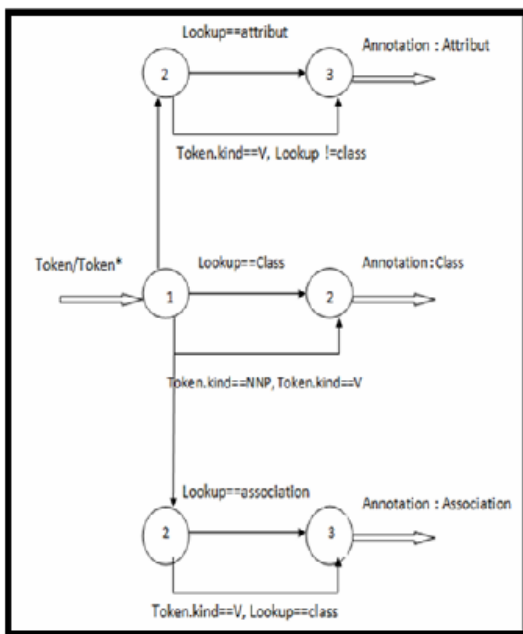


Fig. 6. General transducer

Figure 6 (fig. 6) illustrates a transducer describing extended JAPE grammar used in the context of our proposal: semantic extraction of UML concepts in order to create the corresponding UML class diagram.

### 4.3 Extraction relations between UML concepts

The third phase allows organizing relations between the entities (UML concepts) and gives not defined entities the corresponding annotation based on relations between the named entities that already exist. This phase presents a coreference resolution which is executed by Orthomatcher component. The tasks of recognizing relations are more challenging. The tool matches and annotates complex relations using annotations rules.

### 4.3 Test phase

In this phase, we have formed a corpus of users' requirements in different areas. Then, we have tested our system on this corpus. We applied GATE which generates an XML file containing all semantic tags. We clean the file by removing unnecessary tags like <sentence>, <token>... Figure 7 (fig. 7) shows an example of output GATE file. Our tool is robust and efficient and the error rate is very low, except that case studies are very complicated.

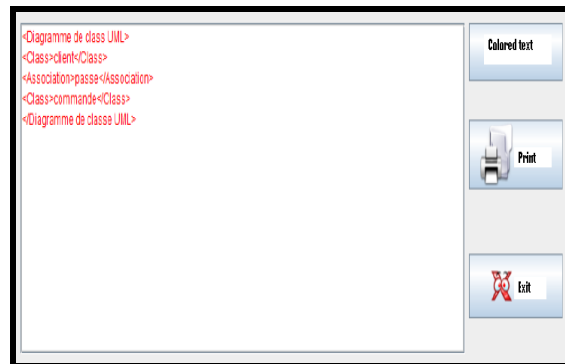


Fig. 7. GATE output file

## 5. Conclusion

Documents are central to Knowledge Management, but intelligent documents, created by semantic annotation, would bring the advantages of semantic search and interoperability. These benefits, however, come at the cost of increased authoring effort. Our system consists in semantic annotation of users' requirements based in GATE API. In fact, we have followed JAPE rules and Gazetteer lists elaboration to identify classes, associations and attributes in class diagrams. We assume that a chart generation of our XML file will be useful to ensure good



readability for the designer. This work is already underway.

## Acknowledgments

The authors would like to thank GATE users for their disponibilities and their helps.

## References

- [1] Alani, H., Kim, S., Millard, D.E., Weal, M.J., Hall, W., Lewis, P.H., and Shadbolt, N.R. (2003). Automatic Ontology-based Knowledge Extraction from Web Documents, *IEEE Intelligent Systems*, 18(1) (January-February 2003), pp 14-21.
- [2] Christiansen, H. and Have, C. T. (2007). From use cases to UML class diagrams using logic grammars and constraints. In *2007 International Conference on Recent Advances in Natural Language Processing*. 128-132.
- [3] Cunningham, H. and Maynard D. and Tablan V., *JAPE: a Java Annotation Patterns Engine* (Second Edition). Technical report CS--00--10, Univ. of Sheffield, Department of Computer Science, 2000.
- [4] Cunningham, Dr Hamish and Maynard, Dr Diana and Bontcheva, Dr Kalina and Tablan, Mr Valentin (2002). GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications. Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02), Philadelphia, US, 2002.
- [5] Ewusi-Mensah, K. "Critical Issues in Abandoned Information Systems Development Projects.," *Communication of the ACM* (40:9), 1997, pp. 74-80.
- [6] J Dreyer, S Müller, B Grusie, and A Zündorf (2010) NT2OD Online - Bringing Natural Text 2 Object Diagram to the web In: *ODiSE'10: Ontology-Driven Software Engineering Proceedings Reno/Tahoe, Nevada, USA: , oct (2010)*.
- [7] Hermida, J.M. and Romá-Ferri, M.T. and Montoyo, A. and Palomar, M. (2009) Reusing UML Class Models to Generate OWL Ontologies. A Use Case in the Pharmacotherapeutic Domain. In: *International Conference on Knowledge Engineering and Ontology Development (KEOD 2009)*, 6 Oct 2009 - 8 Oct 2009, Funchal, Portugal.
- [8] Kaiya, H. and M. Saeki: 2005, 'Ontology Based Requirements Analysis: Lightweight Semantic Processing Approach'. In: *Proc. of The 5th Int. Conf. on Quality Software*. pp. 223{230, IEEE Press.
- [9] D. Maynard, K. Bontcheva, and H. Cunningham. Automatic Language-Independent Induction of Gazetteer Lists. In *Proceedings of 4th Language Resources and Evaluation Conference (LREC'04)*, 2004.
- [10] F. Meziane, N. Athanasakis, S. Ananiadou (2008) Generating Natural Language Specifications from UML Class Diagrams. *Requir. Eng.* 13(1) 1-18.

in computer science in April 2010 from High Institute of Management. Her research interest includes natural language processing, semantic annotation, and web service.

**Wahiba Ben Abdesslem** is an assistant professor in the Department of Computer and Information Science at university of Tunis High Institute of Management. She received the Master Degree in 1992 from Paris III, New Sorbonne, France, and PhD, from Paris 7 Jussieu France in 1997. Her research interest includes Modelling Information System, Natural language processing, document annotation, information retrieval, text mining. She is a member of program committee of several International Conferences: ICCA'2010, ICCA'2011, RFIW 2011, and a member of the Editorial Board of the International Journal of Managing Information Technology (IJMIT).

**Sondes Bouabid** is a student in master in computer science at Paris dauphine. She obtained her degree in June 2010 from University of Tunis High Institute of Management. Her research interests are: textmining, and information system.

**Soumaya Amdouni** is a PhD candidate at the University of tunis High Institute of management. She received her master's degree

# Blind speech separation based on undecimated wavelet packet-perceptual filterbanks and independent component analysis

Ibrahim Missaoui<sup>1</sup>, Zied Lachiri<sup>1,2</sup>

<sup>1</sup> National School of Engineers of Tunis  
BP. 37 Le Belvédère, 1002 Tunis, Tunisia

<sup>2</sup> National Institute of Applied Science and Technology  
BP. 676 centre urbain cedex Tunis, Tunisia

## Abstract

In this paper, we address the problem of blind separation of speech mixtures. We propose a new blind speech separation system, which integrates a perceptual filterbank and independent component analysis (ICA) and using kurtosis criterion. The perceptual filterbank was designed by adjusting undecimated wavelet packet decomposition (UWPD) tree in order to accord to critical band characteristics of psycho-acoustic model. Our proposed technique consists on transforming the observations signals into an adequate representation using UWPD and Kurtosis maximization criterion in a new preprocessing step in order to increase the non-Gaussianity which is a pre-requirement for ICA.

Experiments were carried out with the instantaneous mixture of two speech sources using two sensors. The obtained results show that the proposed method gives a considerable improvement when compared with FastICA and other techniques.

**Keywords:** *Perceptual Filter-Bank, Undecimated Wavelet Packet Decomposition, Independent Component Analysis, Blind speech separation.*

## 1. Introduction

The blind source separation has become an interesting research topic in speech signal processing. It is a recent technique which provides one of the feasible solutions for recover the speech signals from their mixture signals without exploring any knowledge about the source signals and the mixing channel. This challenging research problem has been investigated by many researchers in the last decades, who have proposed many methods and it has been applied in various subjects including speech processing, image enhancement, and biomedical signal processing [1], [4].

Independent Component Analysis (ICA) is one of the popular BSS methods and often used inherently with them. It is a statistical and computational technique in which the goal is to find a linear projection of the data that the source

signals or components are statistically independent or as independent as possible [1].

There are many algorithms which have been developed, using ICA method, to address the problem of instantaneous blind separation [3] such as approaches based on the mutual information minimization [9], [28], maximization of non-Gaussianity [1], [12], [10] and maximization of likelihood [9], [20]. Among these approaches, SOBI algorithm [13] is the second order blind identification which consists to diagonalize a set of covariance matrix and Jade algorithm [14] based on higher order statistics and seek to achieve the separation of the signals by using a Jacobi technique in order to performed a joint diagonalization of the cumulant matrices.

Some researchers aim to improve the performance of BSS system by combining the ICA algorithm with other techniques. For example, the approach developed in [25] combines binary time-frequency masking technique inspired from computational auditory scene analysis system [2] with ICA algorithm. Others techniques decomposes the observed signals using for example subband decomposition [26] or discrete wavelet transform [11] and then apply the separation step in each sub band. In [27], [29], a preprocessing step is employed in wavelet domain but the separation is done in time domain. The idea behind employing wavelet transformation as a preprocessing step is to improve the non-Gaussianity distribution of independent components that is a pre-requirement for ICA and to increase their independency [27], [29], [24]. Inspired from this idea, we propose a new blind separation system, in the instantaneous mixture case, to extract the speech signals of two-speakers from two speech mixtures. The proposed technique uses a perceptual filterbank which is designed by adjusting undecimated wavelet packet decomposition (UWPD) tree, according to critical band characteristics of psycho-acoustic model [15], for the transformation of the two mixtures signals into

adequate representation to emphasize the non-Gaussian nature of mixture signals.

This paper is organized as follows. Section 2 introduces the blind speech separation problem and describes the FastICA algorithm. Section 3 presents the principle of the undecimated wavelet packet decomposition and perceptual filterbank. Then in section 4, the proposed method is described. Section 5 exposes the experimental results. Finally, Section 6 concludes and gives a perspective of our work.

## 2. Blind Speech Separation

### 2.1 Problem Statement

The objective of Blind Speech Separation is to extract the original speech signals from their observed mixtures without reference to any prior information on the sources signals or the observed mixtures. The latter contain a different combination of the source signals and can be mathematically described by:

$$X(t) = AS(t) \quad (1)$$

Where  $X(t)=[x_1(t)...x_n(t)]^T$  is a vector of mixture signals,  $S(t)=[s_1(t)...s_m(t)]^T$  is the unknown vector of sources signals and  $A$  is the unknown mixing matrix having dimension  $(m*n)$ .

Independent Component Analysis is a typical BSS method which tends to solve this problem. The purpose of the ICA is to find a separating matrix or an unmixing matrix  $W=A^{-1}$ , which is used to calculate the estimated signal  $S(t)$  of source signals as  $S(t)=WX(t)$ . To estimate  $W$ , we have to make some fundamental assumptions and impose certain restrictions [1]: The components  $s_i(t)$  of  $S(t)$  (i.e. the sources) are assumed to be statistically independent with non-gaussian distribution.

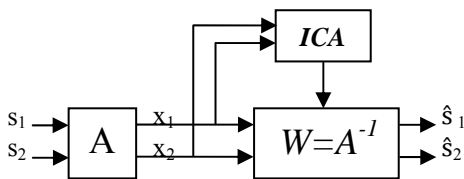


Fig. 1 Principle of ICA.

In other words, ICA can be defined as a method that researches a linear transformation, which maximizes the non-Gaussianity of the components of  $S(t)$ . To measure the non gaussianity, kurtosis or differential entropy called negentropy can be employed. FastICA algorithm [12], [1], [8] is one of the most popular algorithms performing

independent component analysis. The Principle of ICA can be depicted as in Figure 1.

### 2.2 FastICA Method

The FastICA algorithm (A Fast Fixed-Point algorithm of independent component analysis) is a technique proposed and developed by Aapo Hyvarinen and al [1], which is characterized by a high order convergence rate. In this approach, the separation task is based on a point iteration scheme in order to find the maximum of the non-Gaussianity of a projected component. The non-Gaussianity, which is the function of contrast of FastICA algorithm, can be measured with the differential entropy, as known as negentropy [12]. The latter is defined as the difference between the entropy of a Gaussian random vector  $y_{gauss}$  of same covariance matrix as  $y$  and the random vector  $y$ :

$$J(y) = H(y_{gauss}) - H(y) \quad (2)$$

Where  $H(y)$  is the differential entropy of  $y$  and it is computed as follows:

$$H(y) = -\int f(y) \log(f(y)) dy \quad (3)$$

The negentropy can be considered as the optimal measure of the non gaussianity. However, it is difficult to estimate the true negentropy. Thus, several approximations are used and developed such the one developed by Aapo Hyvarinen et al [1], [12]:

$$J(y) = \sum_{i=1}^p k_i (E[g_i(y)] - E[g_i(v)])^2 \quad (4)$$

Where  $k_i$ ,  $g_i$  and  $v$  are respectively positive constants, the non quadratic functions and Gaussian random variable. The separating matrix  $W$  is calculated using a fundamental fixed-point iteration which performed by using the following expression:

$$W_i(k) = E\{\hat{X}_i g(W_i^T \hat{X}_i)\} - E\{g(W_i^T \hat{X}_i)\} W_i \quad (5)$$

## 3. Undecimated Wavelet Packet-Perceptual Filterbank

### 3.1 Wavelet Transform

Wavelet Transform [5], [18], represents an alternative technique for the processing of non-stationary signals which provides a linear powerful representation of signals. The discrete wavelet transforms (DWT) is a multi-resolution representation of a signal which decomposes signals into basis functions. It is characterized by a higher time resolution for high frequency components and a higher frequency resolution for low frequency components. The DWT consists on filtering the input signal by two

filters H (a low-pass filter) and G (a high-pass filter), leading two sub-bands called respectively approximations and details, followed by a decimation factor of two. This filtering process is then iterated only for the approximation sub-band at each level of decomposition [6].

The wavelet packet decomposition (WPD), viewed as a generalization of the discrete wavelet transform (DWT), aims to have a more complete interpretation of the signal, in which the filtering process is applied to decompose on both approximations and details sub-bands and still decimates the filters outputs [7].

To provide a denser approximation and to preserve the translation invariance, the undecimated wavelet packet transform (UWPT) has been introduced and was invented several times with different names as algorithm à trous (algorithm with holes) [17], shift invariant DWT [22] and redundant wavelet transform [16]. The UWPT is computed in a similar manner as the wavelet packet transform except that the downsampling operation after each filtering step is suppressed.

### 3.2 Perceptual filterbank

In the proposed blind speech separation system, we use a perceptual filterbank which is designed using undecimated wavelet packet decomposition [15]. The decomposition tree consists on five levels full UWPD tree using Daubechies 4 (db4) of an 8 kHz speech signal. This decomposition tree structure is adjusted in order to accord to critical band characteristics. The result tree was called critical bands-undecimated wavelet package decomposition (CB-UWPD) tree. Indeed, the audible frequency range of human auditory is 20 to 20000 Hz which can be approximated with 25 barks. However, the sampling frequency chosen is 8 kHz leading to a bandwidth of 4 kHz. As shown in table 1, this bandwidth contains approximately 17 critical bands (barks). The tree structure of CB-UWPD obtained according to the results critical bandwidths (CBW) is depicted in fig 1.

The following equation for each node of the tree is given the corresponding to the critical bandwidths (CBW):

$$cbw(i, j) = 2^{-j}(F_s - 1) \quad (6)$$

Where  $i=(0,1,...,5)$  and  $j=(0.., 2^j-1)$  are respectively the number of levels and the position of the node and  $F_s$  is the sampling frequency.

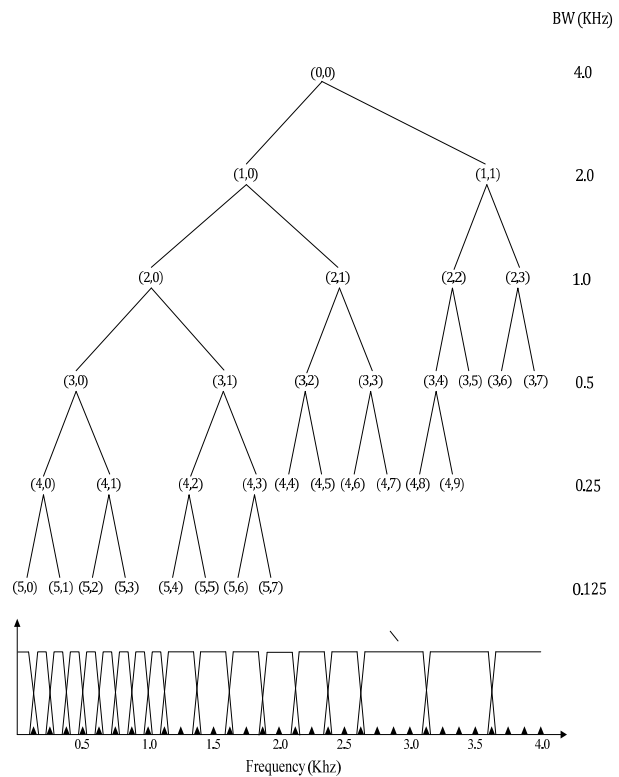


Fig. 1 The CB-UWPD tree and its corresponding frequency bandwidths(perceptual filterbank).

Table 1: Critical Band Characteristics

Critical bands (barks)	Center frequency (Hz)	Critical bandwidth (CBW) (Hz)
1	50	100
2	150	100
3	250	100
4	350	100
5	450	110
6	570	120
7	700	140
8	840	150
9	1000	160
10	1170	190
11	1370	210
12	1600	240
13	1850	280
14	2150	320
15	2500	380
16	2900	450
17	3400	550

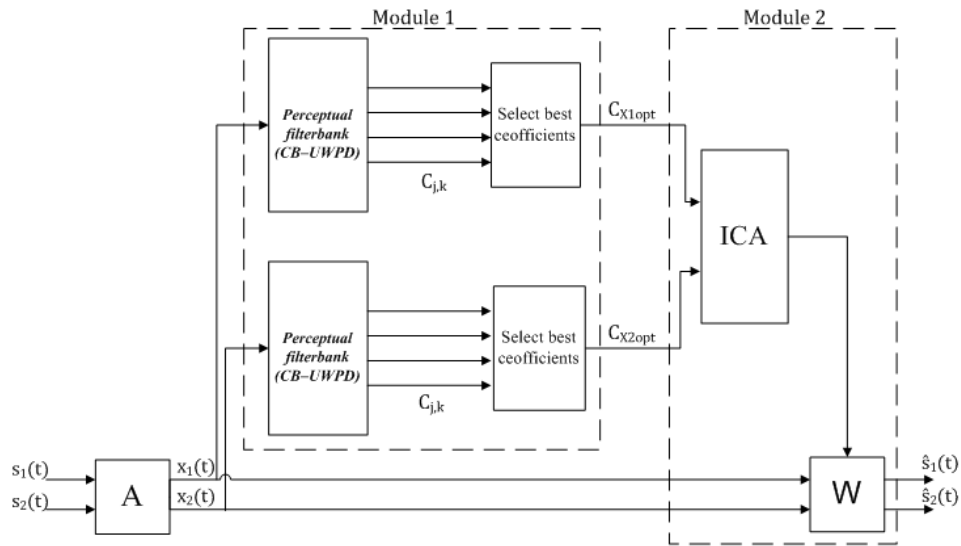


Fig. 2 The framework of proposed speech separation system

#### 4. The proposed Method

We suggest extracting the speech signals of two speakers from two speech mixtures. The proposed speech separation system, as depicted in Figure 1, contains two modules shown in dotted boxes. In the first module, the speech mixtures  $x_1(n)$  and  $x_2(n)$  are processed by a perceptual filterbank which designed by adjusting undecimated wavelet packet decomposition (UWPD) tree, according to critical bands of psycho-acoustic model of human auditory system. In order to increase the non Gaussianity that is a pre-requirement for ICA, we select the appropriate coefficients of the two mixtures which having the high non-Gaussian nature of distribution. The two result signals are then used as two new inputs of the second module. The latter performs the source separation using FastICA algorithm. The description of each module is given in the following sub-sections.

##### 4.1 Preprocessing Module

In this section, we explain the preprocessing module that decomposes the observed signals by perceptual filterbank. This filterbank is designed by adjusting undecimated wavelet packet decomposition tree to accord critical band characteristics of psycho-acoustic model [15]. Each result coefficients of the two mixtures  $x_1(n)$  and  $x_2(n)$  can be viewed as an appropriate signal. In order to increase the non Gaussianity of the signals that is a pre-requirement for ICA, we need to find the best coefficients of each mixture which have the highest non-Gaussian nature of distribution. Thus, the performance of source separation task will be improved. The selection of the best coefficients can be

performed using Shannon entropy criterion [27], [29]. In our case, we chose to use the kurtosis (forth order cumulant) as a criterion to select the best coefficients instead of Shannon entropy criterion. The procedure of the selection algorithm is give as follows:

- Step 1: Decompose the mixture signals into undecimated wavelet packet.
- Step 2: Calculate the kurtosis of each node  $C_{j,k}$  of UWPD tree.
- Step 3: Select the node which has the highest kurtosis.

The Kurtosis (forth order cumulant) for each node can be estimated by using the fourth moment. It is defined as follows:

$$kurt(y(i)) = E[y(i)^4] - 3(E[y(i)^2])^2 \quad (7)$$

Where  $y(i)$  is a vector of UWPD coefficients at each node. We assume that  $y(i)$  is zero-mean and have unit energy.

The forth order cumulant (Kurtosis) represents the classical measure of non gaussianity of signal [1]. Therefore seek to maximize the kurtosis correspond to find the representation of signal which own high non-Gaussian nature of distribution. Consequently, during the application of ICA that exploits the non-Gaussianity in separation task, we will have a significant gain.

##### 4.2 Separation Module

This module is the separation module. It can be devised into two steps. The first step consists of generating the unmixing matrix  $W$  using the FastICA algorithm. The



select UWPD coefficients of two mixtures signals  $x_1(n)$  and  $x_2(n)$  which obtained in preprocessing module are used as two inputs signals of FastICA algorithm. In the second step, the separated signals are obtained by taking into account the original mixtures signals.

## 5. Results and Evaluation

To evaluate the performance of the proposed blind speech separation Method, described in section 4. We use some sentences taken from TIMIT database, this database consists of speech signals of a total of 6300 sentences formed by 10 sentences spoken by each of 630 speakers from 8 major dialect regions of the United States [23]. We consider two speech mixtures composed of two speakers, so we mixes in instantaneous two speech signals, which are respectively pronounced by male and female speaker, two female speakers and two male speakers. The two speech mixtures are generating, artificially, using mixing matrix as:

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \quad (8)$$

The performance evaluation of our work includes different performance metrics such as the blind separation performance measures used in BSS EVAL [19], [30], including the signal to interference ratio SIR and the signal to distortion ratio SDR measures. The principle of these measures consists on decomposing the estimated signal  $s_i(n)$  into the following component sum:

$$s_i(n) = s_{target}(n) + s_{interf}(n) + s_{artefact}(n) \quad (9)$$

where  $s_{target}(n)$ ,  $e_{interf}(n)$  and  $e_{artefact}(n)$  are, respectively, an allowed deformation of the target source  $s_i(n)$  an allowed deformation of the sources which takes account of the interference of the unwanted sources and an artifact term which represents the artifacts produced by the separation algorithm. The two performance criteria SIR and SDR are computed using the last decomposition as following:

$$SIR = 20 \log \frac{\|s_{target}(n)\|^2}{\|s_{interf}(n)\|^2} \quad (10)$$

$$SDR = 20 \log \frac{\|s_{target}(n)\|^2}{\|s_{interf}(n)\|^2 + \|s_{artefact}(n)\|^2} \quad (11)$$

In addition, the recovered speech signals are evaluated with the segmental, overall signal to noise ratio (SNR) and the Perceptual Evaluation of Speech Quality(PESQ). The PESQ is defined in the ITU-T P.862 standard [21] and represents an objective method for evaluating the speech quality. The resulting of PESQ measurement is equivalent

to the subjective "Mean Opinion Score" (MOS) measured score.

In the previous experiments, we compare our system with FastICA algorithm [12] and two well-known algorithms Jade [14] and SOBI [13].

The experimental results are shown in three tables which reports the evaluation measures obtained for three example cases of mixture signal. Table 2 lists the separate performance measures including ratio SIR and SDR obtained after separation by Sobi, Jade, FastICA and the proposed method. We observed that the  $SIR \approx SDR$  and their values is better for the proposed method than that of FastICA, jade and SOBI in the majority of cases for the two signals. The SIR average where we have a mixture composed with two female speakers (or experiment 2) for exemple, is 14.06 for SOBI, 43.12 db for Jade, 39.80 for FastICA and 45.10 db for proposed method. The improvement in the SIR and SDR ratio average is particularly significant in the case of mixture observed formed by two male speaker signals. The improvement average in this case between the proposed method and FastICA is 15.45 db.

Table 3 and table 4 shows that the estimated signals obtained by using the proposed method is better than those obtained by FastICA and the two algorithms Jade and SOBI for the three experiments. We have obtained, for exemple, seg SNR egale to 33.90 db using proposed method and 29.14 db using FastICA.

In order to have a better idea about the quality of estimated signal obtained, PESQ has been used. It is regarded as one of the reliable methods of subjective test. It returns a score from 0.5 to 4.5. Table 5 illustrates the PESQ score obtained. We see that the proposed method is still more effective in terms of perceptual quality than FastICA, jade and SOBI.

Table 2: Comparison of SIR and SDR using SOBI, Jade, Fast-ICA and proposed Method (PM)

		<i>SOBI</i>	<i>Jade</i>	<i>FastICA</i>	<i>PM</i>
Experiment 1 Female (F)+Male (M)	SIR (F.speaker)	26.92	54.72	44.39	51.11
	SIR (M.speaker)	26.29	45.63	51.68	60.75
	SDR (F.speaker)	26.92	54.72	44.39	51.11
	SDR (M.speaker)	26.29	45.63	51.68	60.75
	Average	26.60	50.17	48.03	55.93
Experiment 2 Female (F)+ Female (F)	SIR (F.speaker 1)	14.39	41.37	44.57	51.62
	SIR (F.speaker 2)	13.74	44.87	35.04	38.59
	SDR (F.speaker 1)	14.39	41.37	44.57	51.62
	SDR (F.speaker 2)	13.74	44.87	35.04	38.59
	Average	14.06	43.12	39.80	45.10
Experiment 3 Male (M)+Male (M)	SIR (M.speaker 1)	18.46	65.02	46.20	72.22
	SIR (M.speaker 2)	19.57	37.32	48.37	53.25
	SDR (M.speaker 1)	18.46	65.02	46.20	72.22
	SDR (M.speaker 2)	19.57	37.32	48.37	53.25
	Average	19.01	51.17	47.28	62.73

Table 3: Comparison of segmental SNR using SOBI, Jade, FastICA and proposed Method (PM)

		<i>SOBI</i>	<i>Jade</i>	<i>FastICA</i>	<i>PM</i>
Experiment 1 Female (F)+Male (M)	Seg SNR (F.speaker)	22.58	33.56	30.79	32.79
	Seg SNR (M.speaker)	20.47	29.40	31.15	33.03
Experiment 2 Female (F)+ Female (F)	Seg SNR (F.speaker 1)	15.19	32.01	32.73	33.76
	Seg SNR (F.speaker 2)	12.27	32.12	28.37	30.07
Experiment 3 Male (M)+Male (M)	Seg SNR (F.speaker 1)	13.47	33.20	29.14	33.90
	Seg SNR (F.speaker 2)	20.89	30.88	33.56	34.10

Table 4: Comparison of overall SNR using SOBI, Jade, FastICA and proposed Method (PM)

		<i>SOBI</i>	<i>Jade</i>	<i>FastICA</i>	<i>PM</i>
Experiment 1 Female (F)+Male (M)	Overall SNR (F.speaker)	26.92	54.72	44.39	51.11
	Overall SNR (M.speaker)	26.29	45.63	51.68	60.75
Experiment 2 Female (F)+ Female (F)	Overall SNR (F.speaker 1)	14.37	41.37	44.57	51.62
	Overall SNR (F.speaker 2)	13.82	44.87	35.04	38.59
Experiment 3 Male (M)+Male (M)	Overall SNR (F.speaker 1)	18.47	37.32	46.20	72.22
	Overall SNR (F.speaker 2)	19.55	30.88	48.37	53.25

Table 5: Comparison of PESQ using SOBI, Jade, FastICA and proposed Method (PM)

		<i>SOBI</i>	<i>Jade</i>	<i>FastICA</i>	<i>PM</i>
Experiment 1 Female (F)+Male (M)	PESQ (F.speaker)	2.58	3.29	3.25	3.29
	PESQ (M.speaker)	3.45	4.14	4.27	4.38
Experiment 2 Female (F)+ Female (F)	PESQ (F.speaker 1)	1.53	4.20	4.27	4.42
	PESQ (F.speaker 2)	0.88	3.65	3.40	3.52
Experiment 3 Male (M)+Male (M)	PESQ (F.speaker 1)	1.53	2.24	2.06	2.24
	PESQ (F.speaker 2)	1.20	4.23	4.42	4.47

## 6. Conclusions

In this paper, we proposed a new blind speech separation system in the instantaneous case. This system consists on a combination of ICA algorithm with undecimated wavelet packet transform. The latter is used as a preprocessing module using a Kurtosis maximization criterion in order to increase the non-Gaussian nature of the signals. The results signals are then employed to perform a preliminary separation leading to the inverse matrix  $W$  used to separate the signals in the time domain. The experimental results show that the proposed approach yield to a better separation performance compared to FastICA and two well-known algorithms.

For future work, we aim to separate the convolutive mixtures with the proposed system.

## References

- [1] A. Hyvärinen, J. Karhunen, and E. Oja, Independent Component Analysis, New York: Wiley-Interscience, 2001.
- [2] L. Wang, and G. J. Brown, Computational Auditory Scene Analysis: Principles, Algorithms, and Applications, Hoboken NJ :Wiley/IEEE Press, 2006.
- [3] S. Haykin, Neural Networks and Learning Machines (third ed.), Prentice-Hall, 2008.
- [4] A. Cichocki, and S. Amari, Adaptive Blind Signal and Adaptive Blind Signal and Image Processing. New York: John Wiley and Sons, 2002.
- [5] C.S. Burrus, R.A. Gopinath, and H. Guo, Introduction to Wavelets and Wavelet Transform: A Primer, Prentice Hall, 1998.
- [6] S. Mallat, A Wavelet Tour of Signal Processing: The Sparse Way, 3rd ed, London: Academic Press, 2008.
- [7] R. R. Coifman, Y. Meyer, and M. V. Wickerhauser, "Wavelet analysis and signal processing, in Wavelets and their applications. Boston, MA: Jones and Bartlett, pp.153-178, 1992.
- [8] P. Comon, "Independent components analysis: A new concept?", Signal Processing, Vol. 36, No. 3, 1994, pp. 287-314.
- [9] A.J. Bell and T.J. Sejnowski, "An information maximization approach to blind separation and blind deconvolution", Neural Computation, Vol. 7, 1995, pp. 1004-1034.
- [10] F. S. Wang, H. W. Li, R. Li, "Novel Non Gaussianity Measure Based BSS Algorithm for Dependent Signals", Lecture Notes in Computer Science, Vol. 4505, 2007, pp. 837-844.
- [11] W. Xiao, H. Jingjing, J. Shijiu, X. Antao, and W. Weikui, "Blind separation of speech signals based on wavelet transform and independent component analysis", Transactions of Tianjin University, Vol. 16, No. 2, 2010, pp 123-128.
- [12] A. Hyvrinen, "Fast and robust fixed-point algorithms for independent component analysis". IEEE Transactions on Neural Networks, Vol. 10, No.3, 1999, pp. 626-634.
- [13] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, A blind source separation technique using second order statistics, IEEE Transactions on Signal Processing, Vol. 45, 1997, pp. 434-444.
- [14] J.F. Cardoso, "Higher-order contrasts for independent component analysis", Neural Computation, Vol. 11, 1999, pp.157-192.
- [15] H. Tasmaz, and E. Ercebebi, "Speech enhancement based on undecimated wavelet packet-perceptual filterbanks and MMSE-STSA estimation in various noise environments", Digital Signal process, Vol. 18, No. 5, 2008, pp. 797-812.
- [16] J. Fowler, "The redundant discrete wavelet transform and additive noise", IEEE Signal Processing Letters, Vol. 12, No. 9, 2005, pp. 629-632.
- [17] M. Shensa, "The discrete wavelet transform: Wedding the à trous and Mallat algorithms", IEEE Transactions on Signal Processing, Vol. 40, No. 10, 1992, pp. 2464-2482.
- [18] C. Gargour, M. abrea, , V. Ramachandran, and J.M. Lina, "A short introduction to wavelets and their applications", IEEE Circuits and Systems Magazine, Vol. 9, No. 2, 2009, pp. 57-58.
- [19] E. Vincent, R. Gribonval, and C. Fevotte, "Performance Measurement in Blind Audio Source Separation", IEEE Transactions on Audio, Speech, and Language Processing, Vol. 14, No. 4, 2006, pp 1462-1469.
- [20] J. T. Chien, B. C. Chen, "A New Independent Component Analysis for Speech Recognition and Separation", IEEE transactions on audio, speech and language processing, Vol. 14, No. 4, 2006, pp. 1245 - 1254.
- [21] ITU-T P.862, "Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs", International Telecommunication Union, Geneva, 2001.
- [22] A. T. Walden, and C. Contreras, "The phase-corrected undecimated discrete wavelet packet transform and its application to interpreting the timing of events", in Proceedings of the Royal Society of London, 1998, Vol. 454, pp. 2243-2266.
- [23] W. Fisher, G. Dodington, and K. Goudie-Marshall. "The TIMIT-DARPA speech recognition research database: Specification and status", In. DARPA Workshop on Speech Recognition, 1986.
- [24] K. Usman, H. Juzoji, I. Nakajima, and M.A. Sadiq, "A study of increasing the speed of the independent component analysis (ICA) using wavelet technique", in Proceedings of International Workshop on Enterprise Networking and Computing in Healthcare Industry (HEAL THCOM 2004), 2004, pp. 73-75.
- [25] M.S. Pedersen, D.L. Wang, J. Larsen, and U. Kjems "Overcomplete blind source separation by combining ICA and binary time-frequency masking", in Proceedings of IEEE Workshop on Machine Learning for Signal Processing, 2005, pp. 15-20.
- [26] T. Tanaka and A. Cichocki, "Subband decomposition independent component analysis and new performance criteria", in Proceedings of the IEEE Conference on Acoustics, Speech and Signal Processing, 2004, pp. 541-544.
- [27] R. Moussaoui, J. Rouat and R. Lefebvre, "Wavelet Based Independent Component Analysis for Multi-Channel Source Separation", in Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 2006, Vol. 5, pp. 645-648.

- [28] J. T. Chien, H. L. Hsieh, and S. Furui, "A new mutual information measure for independent component analysis", in Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 2008, 1817-1820.
- [29] M.R. Mirarab, M.A Sobhani, and AA Nasiri, "A New Wavelet Based Blind Audio Source Separation Using Kurtosis", in International Conference on Advanced Computer Theory and Engineering, 2010, Vol. 4, pp. 36-39.
- [30] C. Fevotte, R. Gribonval, and E. Vincent, BSS EVAL toolbox user guide, IRISA, Rennes, France, Technical Report 1706, 2005.

**Ibrahim Missaoui** was born in Tunisia. He received his M.S degree in automatic and Signal processing from the National School of engineering of Tunis (ENIT) in 2007. He started preparing his Ph.D degree in Electrical Engineering in 2008. Her current research area is the blind speech separation.

**Zied Lachiri** was born in Tunis, Tunisia. He received the M.S. degree in automatic and signal processing and the Phd. degree in electrical engineering from the National School of Engineer of Tunis (ENIT- Tunisia), in 1998 and 2002, respectively. In 2000, he joined the Applied Sciences and Technology National Institute (INSAT), as research Assistant and became Assistant Professor in 2002. He is currently an Associate Professor at the Department of Physic and instrumentation (INSAT) and member of Systems and Signal Processing Laboratory (LSTS-ENIT)). His research interests include signal processing, image processing and pattern recognition, applied in biomedical, multimedia, and man machine communication. He is Member of the EURASIP, European Association for Signal, Speech and Image Processing.

# A Neural Network Model for Construction Projects Site Overhead Cost Estimating in Egypt

Ismaail ElSawy<sup>1</sup>, Hossam Hosny<sup>2</sup> and Mohammed Abdel Razek<sup>3</sup>

<sup>1</sup> Civil Engineering Department, Thebes Higher Institute of Engineering  
Corniche Maadi, Cairo, Egypt

<sup>2</sup> Construction Engineering Department, Zagazig University  
Zagazig, Egypt

<sup>3</sup> Construction and Building Department, Arab Academy for Science, Technology and Maritime Transport  
Cairo, Egypt

## Abstract

Estimating of the overhead costs of building construction projects is an important task in the management of these projects. The quality of construction management depends heavily on their accurate cost estimation. Construction costs prediction is a very difficult and sophisticated task especially when using manual calculation methods. This paper uses Artificial Neural Network (ANN) approach to develop a parametric cost-estimating model for site overhead cost in Egypt. Fifty-two actual real-life cases of building projects constructed in Egypt during the seven year period 2002-2009 were used as training materials. The neural network architecture is presented for the estimation of the site overhead costs as a percentage from the total project price.

**Keywords:** Construction Projects, Project Site Overhead Cost, Egypt, Artificial Neural Network.

## 1. Introduction

Applications of ANN (Artificial Neural Network) in construction management in general go back to the early 1980's. These applications cover a very wide area of construction issues. Neural network models have been developed internationally to assist the managers or contractors in many crucial construction decisions. Some of these models were designed for cost estimation, decision making, predicting the percentage of mark up, predicting production rate ...etc.

The objective of this research is to develop a neural network (NN) model to assess the percentage of site overhead costs for building projects in Egypt. This can

assist the decision makers during the tender analysis process.

Cost Estimating is one of the most significant aspects for proper functioning of any construction company. It is the lifeblood of the firm and can be defined as the determination of quantity and the prediction or forecasting, within a defined scope, of the costs required to construct and equip a facility.

The significance of construction cost estimating is highlighted by the fact that each individual entity or party involved in the construction process have to make momentous financial contribution that largely affects the accuracy of a relevant estimate. The importance and influence of cost estimating is supported by scores of researches.

Carty (1995) and Winslow (1980), for example, have documented the importance of cost estimating, mentioning it as a key function for acquiring new contracts at right price and hence providing gateway for long survival in the business. According to Larry, D. (2002) cost estimating is of paramount importance to the success of a project [1].

Alcabas (1988), articulated that, estimating departments is responsible for the preparation of all estimates, estimating procedures, pricing information, check lists and applicable computerized programs. He also insists on the fact that accurate cost categorization, cost reporting, and profit calculation are the heart of the construction business. In order to achieve a financial engineered estimating methodology, it is imperative that different techniques should be evaluated [3].

Hegazy and Moselhi (1995), conducted several surveys studies in Canada and the United States to determine the elements of costs estimation. The survey was carried out with the participation of 78 Canadian and U.S.A building construction contractors in order to elicit current practices



with respect to the cost elements used to compile a bid proposal and to identify the types of methods used for estimating these elements. Their results indicated that direct cost and project overhead costs are estimated by contractors primarily in a detailed manner, which is contrary to the estimation of the general overhead costs and the markup [9].

Assaf, S. A. et al. (2001), investigated the overhead cost practices of construction companies in Saudi Arabia. They show how the unstable construction market makes it difficult for construction companies to decide on the optimum level of overhead costs that enables them to win and efficiently administer large projects [4].

Cost estimating models and techniques provides a well defined engineered calculation methods for the evaluation and assessment of all items of office overhead, project overhead, profit anticipation, total project cost estimation, and the assessment of overhead costs for construction projects that leads to competitive bidding in the construction industry [11].

This paper presents the steps followed to develop a proposed model for site overhead cost estimating. The necessary information and the required projects data were collected on two successive yet dependent stages:

- I. Comparison between the list of site overhead factors collected from previous studies and the applied Egyptian site overhead list of factors that is adapted by the first and second categories of construction firms in Egypt; and
- II. Collection of all required site overhead cost data for a sample of projects in Egypt to be used during the analysis phase and site overhead cost assessment model development.

## 2. Research Methodology

The findings from the survey conducted on all the previous researches served as key source in the identification of the main factors affecting site overhead costs for building construction projects. Based on an extensive review for the previous studies conducted in this area of work, the survey for such factors mainly include projects need for specialty contractors, percentage of sub-contracted works, consultancy and supervision, contract type, firm's need for work, type of owner/client, site preparation, projects scheduled time, need for special construction equipment, delay in projects duration, firms previous experience with projects type, legal environmental and public policies for the home country, projects cash-flow plan, project size, and projects location. Hence, the study shed a great deal of light on the area of site overhead costs for building construction projects in Egypt. Through seeking the experts opinions regarding the

development of a list for the main factors affecting the building projects overhead costs. They will be used during the development of the model. Such factors were mainly identified based on the expert's opinions from selected groups of prominent industrial professionals and qualified academicians from the most prominent universities in Egypt. The principal objective of this survey study was to reinforce the potential model, based on the expert's opinions from the aforementioned expert professionals [12].

Expert opinion included the reviews from nineteen prominent industrial professionals and sixteen qualified academicians from the American University in Cairo and the Arab Academy for Science and Technology and Maritime Transport. Reviews from experienced industrial professionals were essential for developing the overall model as these professionals are directly associated with the leading Egyptian building construction firms.

Each expert from both contractor and academic background were approached based on their personnel experiences. Half of the responses were obtained via personnel interviews and the other half were obtained through delivering the questionnaire and collecting back the same, E-mail or Fax.

As this phase of seeking expert's opinion consist of the walk-through observations of the selected specified industrial professionals and academicians connected to the construction industry. These reviews provided us with qualified remarks and suggestions, which will lead to making the necessary alterations on the list of the previously identified overhead cost factors to make it adaptable to the Egyptian building construction industry market. This is an essential step to have a more firm and yardstick final model for the assessment of overhead costs for building construction projects, in Egypt [12].

## 3. Data Collection

This phase is divided into two stages; first stage is to perform a comparison between the overhead cost factors from the comprehensive literature study and the Egyptian construction industry. Hence, the main factors affecting site overhead costs can be clearly identified. The second stage is to collect data for 50 projects from several construction companies that represent the first and the second categories of construction companies, in Egypt [12].

### 3.1 The questionnaire

In the first section of the data collection process, a questionnaire is prepared to investigate the main factors affecting site overhead cost for building construction projects in Egypt.

The questionnaire consisted of three sections, the first section contained nine yes or no questions to confirm or eliminate any of the constituent factors that have been collected previously from the literature study. The second section is where the experts illustrate the factors currently accounted for by construction firms in Egypt. The third section is where the experts are asked for their own opinions for the factors that are not accounted for and should be considered in order to stroll with the construction industry in Egypt. The characteristics of the participating experts, the contractors and the academicians are setting the basis for the findings of this study. The mentioned characteristics of contractors include their personnel professional experience and size of the firm they are associated with. The distinctiveness of academicians described includes their designation, area of specialization and essentially their experience.

Experts for this extensive research are very scrupulously identified to obtain comprehensive and precise results. The highly capable experts were selected among the practicing, experienced contractor's professionals in Egypt and the highly qualified academicians from the two renowned universities not only in Egypt but in the entire region [12].

### 3.1.1 Academicians

Academicians are the professionals, who have strong influence on national research and scientific work. As part of this thesis, expert appraisals from faculty members belonging to Construction Engineering and Management or Civil Engineering fields from two prestigious universities in Egypt. The Academicians engaged for this research are icons from academia. Their expertises are articulated by the fact that, seventy percent of the respondents are either Professor or Associate Professor in the two renowned universities. Along with the aforementioned colossal qualification levels, the traits of the participating academic professionals include their experience, classified based on the number of years in academia. Thirty one percent of the interviewed experts are dedicating their services to the academic discipline from more than 20 years. Another forty four percent of the academic experts have 10-20 years of practicing experience (twenty five percent have from 15-20 years and nineteen percent have from 10-15) and twenty five percent have less than 10 years of professional experience in academia (Fig. 1) [12].

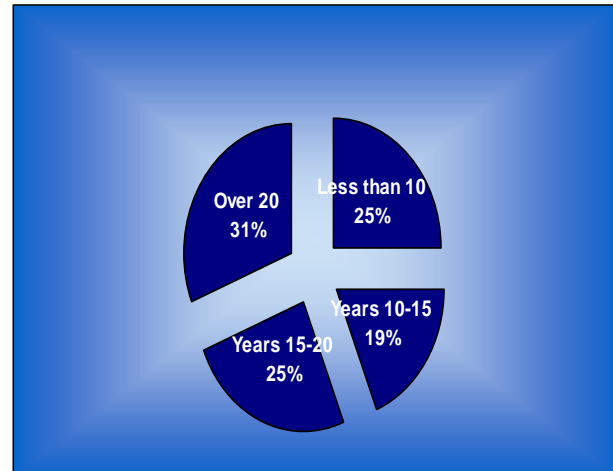


Fig. 1. Academicians Years of Experience. [12]

### 3.1.2 Contractors

The participating contractors (Cost Estimating Engineers) are highly experienced professionals from the construction industry. About fifty percent of the experts have more than 20 years of professional experience in the construction business. The remaining has experience less than 20 years. These vastly experienced industry professionals occupy senior and highly ranked administrative positions within their firms. Seventy percent of the experts are ranked as General Manager Engineers. The remaining thirty percent work as project cost estimation engineers. The participants work for successful construction firms belonging to the first and second categories. Twelve experts work for first category construction companies, five experts work for second category construction companies, and two experts work for a major construction consultancy firm all within Egypt, (Fig. 2).

The views of the contracting experts from firms of different grades were sought to get a more diversified & comprehensive review [12].

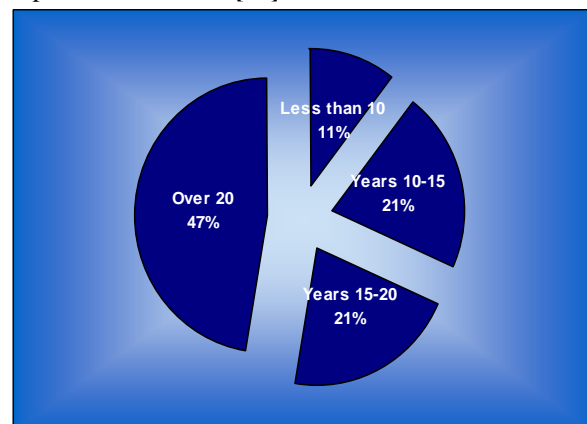


Fig. 2. Contractors Years of Experience. [12]

The analysis of the collected questionnaires illustrated that there is a difference between the factors that govern the assessment of building construction site overhead cost in Egypt and the international building construction industry trend. Many factors are not accounted for in Egypt due to its insignificance in the local market while it is a great contributor in both Europe and North/South America construction markets. Moreover, in Egypt there is a trend between contractors to combine two or more contributing items in one main factor. The academicians contravened with that behavior and characterized it to be an unprofessional attitude because it depends entirely on the person that is performing the task and his/her experience with the projects on hand (personalization). So after cross-matching and making the necessary alterations on the questionnaires collected from both the contractors and academicians in Egypt, a final list of factors were generated that represent both the parties and it can accurately represent the factors that contribute to building construction site overhead cost in the Egyptian construction market (Table 1) [12].

Table 1: Factors Contributing to Construction Site Overhead Cost Percentage in Egypt

	Factor
1	Construction Firm Category.
2	Project Size.
3	Project Duration.
4	Project Type.
5	Project Location.
6	Type-Nature of Client.
7	Type of Contract.
8	Contractor-Joint Venture.
9	Special Site Preparation Requirements.
10	Project need for Extra-man Power.

#### 4. Site Overhead Cost Data

A comparative analysis was performed between building construction site overhead cost and each constituent of site overhead regarding building construction projects, with the aid of (52) completed building construction projects. These projects were executed during the seven year period from 2002 to 2009. The comparison is made in terms of cost influence for each factor of projects site overhead on the percentage of projects site overhead cost in order to recognize and understand the governing relationship between each factor and the percentage of site overhead cost [12].

It must be illustrated that for all the collected projects the adapted construction technology was typical traditional reinforced concrete technology. This may be due to the participating experts opinion, because that technology represents over (95%) of the adopted building construction technology in Egypt. Contrarily, if any specific

construction technique is required for a certain project it must be accounted for by the construction firm cost estimating department in an exceptional manner [12].

#### 5. Comparative Analysis Results

The major and minor findings of the entire research were summarized in this part of the research. Based on the findings the current and further recommendations are developed as the base for further research in the very context of building construction projects overhead cost for the first and the second categories of construction companies, in Egypt [12].

The analysis illustrated many facts that needed to be clarified and understood about the percentage of site overhead costs for building construction projects in Egypt. These facts will be the structure (backbone) for the development of a model for the assessment of site overhead cost as a percentage from the total contract amount for building construction projects, in Egypt. This can be simply summarized in the following two facts: [12]

- A. Through the literature review and the expert's opinions potential factors that are found to influence the percentage of site overhead costs for building construction projects in Egypt, ten factors were identified.
- B. The analysis of the collected data gathered from fifty-two real life building construction projects from Egypt during the seven year period from 2002 to 2009, illustrated that project's duration, total contract value, projects type, special site preparation needs and projects location are identified as the top five factors that affect the percentage of site overhead costs for building construction projects in Egypt.

#### 6. Neural Network Model

The guidelines of N-Connection Professional Software version 2.0 (1997), users manual were used to obtain the best model. Moreover, for verifying this work the traditional trial and error process was performed to obtain the best model architecture [11].

The following sections present the steps performed to design the artificial neural network model, ANN-Model. Neural network models are generally developed through the following basic five steps [8]:

1. Define the problem, decide what information to use and what network will do;
2. Decide how to gather the information and represent it;
3. Define the network, select network inputs and specify the outputs;
4. Structure the network;
5. Train the network; and

6. Test the trained network. This involves presenting new inputs to the network and comparing the network's results with the real life results, (Fig. 3).

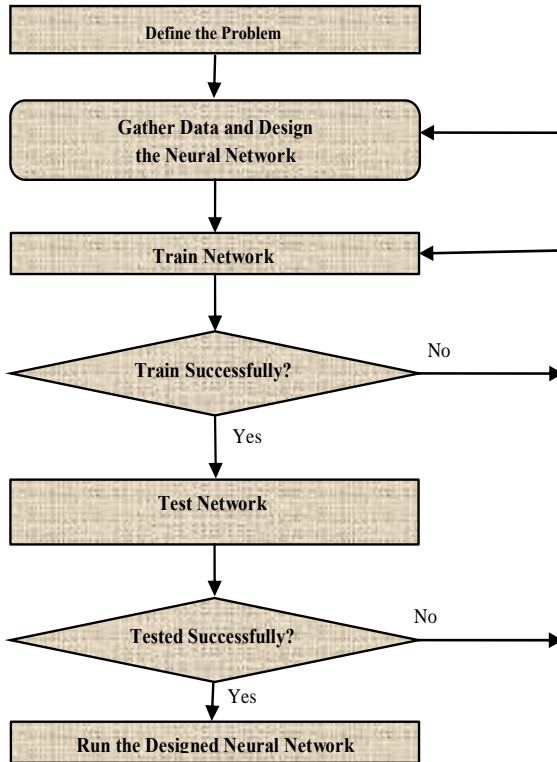


Fig. 3. Neural Network Design. [8]

### 6.1. Design of the Neural Network Model

Through this step, the following sequences were followed:

#### i. Neural Network Simulation Software Selection

Many design software are used for creating neural network models. As stated earlier in the previous studies phase, many researchers used Neural Network Software in construction management in general. In this research, the N-Connection Professional Software Version 2.0 was used to develop the Neural Network Model.

This application software is very easy to use and its predicting accuracy is very high compared to other software program. It is compatible with Microsoft Windows. The N-Connection uses the back propagation algorithm in its engine. The past researches proved that the back-propagation rule is a suitable learning rule for most problems. It is the most commonly used technique for solving estimation and prediction problems [16].

Firstly, in order to design the neural network model the (N-Connection V2.0) guidelines will be used for assistance. Moreover, to verify this research work the trial and error process was used to obtain the best structure of the model. During this procedure if the network is not

trained satisfactory, adding or removing of hidden layers and hidden nodes will be performed until an acceptable model structure is reached, that can predict the percentage of site overhead cost with an acceptable error limit. The learning rate, training and testing tolerance are fixed by the N-Connection V 2.0 automatically [16].

#### ii. Determining the Best Network Architecture

There are two questions in neural network designing that have no precise answers because they are application-dependent: How much data do you need to train a network? **And**, how many hidden layers and nodes are the best numbers to use? In general, the more facts and the fewer hidden layers and hidden nodes that you can use, is the better [16]. There is a subtle relationship between the number of facts and the number of hidden layers/nodes. Having too few facts or too many hidden layers/nodes can cause the network to "Memorize". When this happens, it performs well during training but tests poorly [16]. The network architecture refers to the number of hidden layers and the number of nodes within each hidden layer [16]. The two guidelines that are discussed in the following section can be used in answering the last two questions [8].

#### iii. Determining the Number of Hidden Layers/Nodes

Hidden layer is a layer of neurons in an artificial neural network that does not connect to the outside world but connects to other layers of neurons [16].

Hegazy et al. (1995), stated that one hidden layer with a number of hidden neurons as one-half of the total input and output neurons is suitable for most applications, but due to the ease of changing the network architecture during training, an attempt will be performed to verify this research work, through finding the network structure that generates the minimum RMS value for the given problem output parameters [9].

Before starting to build, train and validate the network model, there are two parameters that should be well defined to have a good training manner. These parameters are:

#### 1. Training and Testing Tolerance

Training and testing tolerance is a value that specifies how accurate the neural network's output must be considered correct during training and testing. The most meaningful tolerance is specified as a percentage of the output range, rather than the output value [16].

A tolerance of 0.1 means that the output value must be within 10% of the range of the output to be considered correct. Selecting a tolerance that is too loose (large) or too tight (small) can have an impact on the network's ability to make predictions. It is important that the selected tolerance will give responses close enough to the pattern



to be useful. However, it is not always possible for Neural Connection V2.0 to train if it begins with a very small tolerance. In this study the tolerance is set by the program to (0.1).

## 2. Learning Rate

The learning rate specifies how large an adjustment Neural Connection will make to the connection strengths when it gets a fact wrong. Reducing the learning rate may make it possible to train the network to a smaller tolerance. The learning rate pattern is automatically set by the Neural Connection 2.0 Software program in a way that maximizes the performance of the program to achieve the best results.

### iv. Training the Network

Training the network is a process that uses one of several learning methods to modify weight, or connection strengths. All trial models experimented in this study was trained in a supervised mode by a back-propagation learning algorithm. A training data set is presented to the network as inputs, and the outputs are calculated. The differences between the calculated outputs and the actual target output are then evaluated and used to adjust the network's weights in order to reduce the differences. As the training proceeds, the network's weights are continuously adjusted until the error in the calculated outputs converges to an acceptable level. The back-propagation algorithm involves the gradual reduction of the error between model output and the target output. Hence, it develops the input to output mapping by

- Equation (2):

$$\text{Absolute Difference} = \left( \frac{\text{Real Life Target Outcome} - \text{Predicted Model Outcome}}{\text{Real Life Target Outcome}} \times 100 \right)$$

An absolute difference of 10 means that there is a 10 percent difference between the models predicted outcome value and the actual real life outcome value for that given project. This difference can be positive or negative difference (i.e. absolute difference range =  $\pm 10$ ) and that must be clearly stated when testing phase is completed for it represents one of the main features of the constructed Neural Network Model characteristics [16].

### vi. Creating Data File for Neural Connection

N-Connection 2.0 is a tool that allows creating definition, training fact, and testing facts. The database that feeds into the Excel file consists of 47 examples of building construction site overhead costs percentage for projects constructed during the period 2002 to 2009 in Egypt, and 5 examples will be set aside for the final best model

minimizing a root mean square error (RMS) that is expressed in the equation (1) [16]:

- Equation (1):

$$\text{RMS} = \sqrt{\sum_{i=1}^n (O_i - P_i)^2 / n}$$

Where  $n$  is the number of samples to be evaluated in the training phase,  $O_i$  is the actual output related to the sample  $i$  ( $i=1\dots n$ ), and  $P_i$  is the predicted output. The training process should be stopped when the mean error remains unchanged. The training file has (90%) of the collected facts, i.e. has 47 facts (Projects). These facts are used to train and validate the network [11].

### v. Testing the Network

Testing the network is essentially the same as training it, except that the network is shown facts it has never seen before, and no corrections are made. When the network is wrong, it is important to evaluate the performance of the network after the training process. If the results are good, the network will be ready to use. If not, this means that it needs more or better data or even re-designs the network. A part of the collected facts (data) around (10%), i.e. 5 facts (projects) is set aside randomly from the set of training facts (projects) [11]. Then these facts are used to test the ability of the network to predict a new output where the absolute difference is calculated for each test project outcome by the equation (2) [16]:

testing. The Neural Connection 2.0 program will need around 34 (73%) of the facts for training, which are the calculated minimum needed number of facts for the program to train properly, which leaves 13 of the facts for validation [11].

### vii. Determining the Best Structure for the Model

The characteristics of the model learning rule, training and testing tolerance is set automatically by the program. The variables that the program requires setting during the design stage are [16]:

1. Number of Hidden Layers (the program accepts up to two Hidden Layers);
2. Number of Hidden Nodes in each Layer; and
3. Type of Transfer Function in each layer.



The program is generated through the following sequence of alterations and selecting the model structure that provides the minimum RMS value [11]:

1. One Hidden Layer with Sigmoid Transfer Function; (Table 2A)

2. One Hidden Layer with Tangent Transfer Function; (Table 2B)

3. Two Hidden Layers with Sigmoid Transfer Function in each; (Table 2C)

4. Two Hidden Layers with Tangent Transfer Function in each; (Table 2D)

Table 2A: Experiments for Determining the Best Model

Model No.	Input Nodes	Output Node	No. of Hidden Layers	No. of Hidden Nodes		Absolute Difference %	RMS
				In 1 <sup>st</sup> Layer	In 2 <sup>nd</sup> Layer		
1	10	1	1	3	0	7.589891	0.900969
2	10	1	1	4	0	5.491507	0.602400
3	10	1	1	5	0	8.939657	1.046902
4	10	1	1	6	0	7.766429	0.932707
5	10	1	1	7	0	4.979286	0.535812
6	10	1	1	8	0	5.818345	0.647476
7	10	1	1	9	0	4.947838	0.579932
8	10	1	1	10	0	8.887463	1.039825
9	10	1	1	11	0	4.858645	0.507183
10	10	1	1	12	0	5.352388	0.651948
11	10	1	1	13	0	2.476118	0.276479
12	10	1	1	14	0	2.857856	0.428663
13	10	1	1	15	0	4.074554	0.478028
14	10	1	1	20	0	8.065637	1.050137

i.e. Model trials from 1 to 14 has a Sigmoid transfer function.

The first fourteen model trails illustrated that the RMS and Absolute Difference values changed as the number of hidden nodes in the single hidden layer increased in a nonlinear relationship, where the lowest RMS value of 0.276479 and a corresponding Absolute Difference value of 2.476118 were achieved in the eleventh trial where there were thirteen hidden nodes in the single hidden layer with a sigmoid transfer function. On the other side highest

RMS value of 1.050137 and the corresponding Absolute Difference value of 8.065637 were achieved in the fourteenth trial when there was twenty hidden nodes in the single hidden layer with a sigmoid transfer function. For the remaining twelve model trails the RMS and Absolute Difference values changed consecutively within the above mentioned ranges for each model trial.

Table 2B: Experiments for Determining the Best Model

Model No.	Input Nodes	Output Node	No. of Hidden Layers	No. of Hidden Nodes		Absolute Difference %	RMS
				In 1 <sup>st</sup> Layer	In 2 <sup>nd</sup> Layer		
15	10	1	1	3	0	3.809793	0.490956
16	10	1	1	4	0	5.666974	0.703804
17	10	1	1	5	0	3.813867	0.425128
18	10	1	1	6	0	5.709665	0.709344
19	10	1	1	7	0	5.792984	0.634338
20	10	1	1	8	0	2.952316	0.343715
21	10	1	1	9	0	5.629162	0.655106
22	10	1	1	10	0	3.544173	0.387283
23	10	1	1	11	0	5.578666	0.686378
24	10	1	1	12	0	5.772656	0.701365
25	10	1	1	13	0	3.582526	0.380564
26	10	1	1	14	0	4.614612	0.515275
27	10	1	1	15	0	4.806596	0.641098
28	10	1	1	20	0	7.005237	0.826699

i.e. Model trials from 15 to 28 has a Tangent transfer function.

The model trails from 15 to 28 where there is one hidden layer, illustrated that the RMS and Absolute Difference values changed as the number of hidden nodes/hidden layer changed in a nonlinear relationship, where the lowest RMS value of 0.343715 and a corresponding Absolute

Difference value of 2.952316 were achieved in the twentieth model trial when there was eight (8) hidden nodes in the single hidden layer. On the other side, with a tangent transfer function, the highest RMS value of 0.826699 and the corresponding Absolute Difference

value of 7.005237 were achieved in the twenty eighth model trial when there were twenty hidden nodes in the single hidden layer. The remaining values changed

consecutively within the above mentioned ranges for each model trial.

Table 2C: Experiments for Determining the Best Model

Model No.	Input Nodes	Output Node	No. of Hidden Layers	No. of Hidden Nodes		Absolute Difference %	RMS
				In 1 <sup>st</sup> Layer	In 2 <sup>nd</sup> Layer		
29	10	1	2	2	1	9.919941	1.519966
30	10	1	2	2	2	5.170748	0.581215
31	10	1	2	3	1	10.374248	1.413138
32	10	1	2	3	2	11.167767	1.687072
33	10	1	2	3	3	8.013460	1.140512
34	10	1	2	4	1	5.679721	0.643957
35	10	1	2	4	2	5.577789	0.617385
36	10	1	2	4	3	5.448696	0.598400
37	10	1	2	4	4	4.079718	0.492011
38	10	1	2	5	3	4.191063	0.574500
39	10	1	2	5	4	6.024062	0.723419
40	10	1	2	5	5	5.322466	0.654373
41	10	1	2	6	4	7.257790	0.804202
42	10	1	2	6	5	5.158298	0.567479
43	10	1	2	6	6	5.270355	0.545017

i.e. Model trials from 29 to 43 has a Sigmoid transfer function for both hidden layers.

The model trails from 29 to 43 illustrated that the RMS and Absolute Difference values changed as the number of hidden nodes per each hidden layer increased in a nonlinear relationship, where the lowest RMS value of 0.492011 and a corresponding Absolute Difference value of 4.079718 were achieved in the model trial number (37) when there were two hidden layers with four hidden nodes in each layer and having a sigmoid transfer function. Contrarily, the highest RMS value of 1.687072 and the

corresponding Absolute Difference value of 11.167767 were achieved in the model trial number (32) when there were two hidden layers with three hidden nodes in the first layer and two hidden nodes in the second hidden layer and having a sigmoid transfer function. For the remaining thirteen model trails the RMS and Absolute Difference values changed consecutively within the above mentioned ranges for each model trial having a sigmoid function in each layer.

Table 2D: Experiments for Determining the Best Model

Model No.	Input Nodes	Output Node	No. of Hidden Layers	No. of Hidden Nodes		Absolute Difference %	RMS
				In 1 <sup>st</sup> Layer	In 2 <sup>nd</sup> Layer		
44	10	1	2	2	1	4.364562	0.499933
45	10	1	2	2	2	3.551318	0.380629
46	10	1	2	3	1	4.787220	0.493240
47	10	1	2	3	2	6.267891	0.852399
48	10	1	2	3	3	6.515138	0.829739
49	10	1	2	4	1	3.458081	0.481580
50	10	1	2	4	2	9.249286	1.158613
51	10	1	2	4	3	4.735680	0.552350
52	10	1	2	4	4	7.445228	0.991062
53	10	1	2	5	3	7.729862	1.105441
54	10	1	2	5	4	9.807989	1.180131
55	10	1	2	5	5	6.060798	0.657344
56	10	1	2	6	4	3.213154	0.355932
57	10	1	2	6	5	4.381631	0.490479
58	10	1	2	6	6	4.731568	0.502131

i.e. Model trials from 44 to 58 has a Tangent transfer function for both hidden layers.

The model trails from 44 to 58 illustrated that the RMS and Absolute Difference values changed as the number of hidden nodes per each hidden layer increased in a nonlinear relationship, where the lowest RMS value of 0.355932 and a corresponding Absolute Difference value of 3.213154 were achieved in the model trial number (56),

when there was two hidden layers with six hidden nodes in the first hidden layer and four hidden nodes in the second hidden layer and with a tangent transfer function in each layer. On the other side, the highest RMS value of 1.180131 and the corresponding Absolute Difference value of 9.807989 were achieved in the model trial

number (54) when there was two hidden layers with five hidden nodes in the first layer and four hidden nodes in the second hidden layer and with a tangent transfer function in each layer. For the remaining thirteen model trails the RMS and Absolute Difference values changed consecutively within the above mentioned ranges for each and with a sigmoid function in each layer [11].

The recommend model for this prediction problem is that with the least RMS value from all the fifty-eight trails and error process [16]. This is trial number eleven [11].

As a result, from training phase the characteristics of the satisfactory Neural Network Model that was obtained

through the trail and error process are presented in (Table 3) and (Fig. 4).

Model Trial Number Eleven with the following Eight Design Parameters, which are [11]:

1. Input layer with **10** Neurons (nodes);
2. One hidden layer with **13** Neurons (nodes);
3. Output layer with **1** Neuron (node);
4. With a Sigmoid Transfer Function;
5. Learning rate automatically adjusted by the program;
6. Training Tolerance = **0.10** (Adjusted by Program);
7. Root Mean Square Error = **0.276479**;
8. Absolute Mean Difference % = **2.476118**.

Table 3: Characteristics of the Best Model

Model	No. of input nodes	No. of hidden layers	No. of nodes/ hidden layer	LR	TF	No. of output nodes	RMS
11	10	1	13	Back propagation	Sigmoid function	1	0.276479

LR: Learning Rule; TF: Transfer Function; RMS: Root Mean Square Error.

#### viii. Testing the Validity of the Model

To evaluate the predictive performance of the network, the five projects that were previously randomly selected and reserved for testing from the total collected projects are introduced to the best model without the percentage of their site overhead cost for testing the prediction ability of the designed ANN-program.

The model will predict the percentage of building construction projects site overhead costs for projects constructed in Egypt. The predicted percentage will be compared to the real life projects percentage (stored outside the program). The difference between them will be calculated if it is equal or under the value of the designed model's Absolute Difference, then it is considered to be a correct prediction attempt. If it exceeds the value of the designed model Absolute Difference, then it is considered to be a wrong prediction attempt, (Table 4) presents the actual and predicted percentages for the test sample.

The model correctly predicted four from the five testing projects sample which is 80% of the test sample. The wrongly predicted project had a positive difference between the value of predicted percentage from the model output and the real life percentage for the same project equal to (+) 4.620294427%. This means that the predicted outcome is greater than the actual real life project value by this percentage value [11]. Such percentage is found to be acceptable; program user's manual, because the difference between the predicted program outcome for this project and the real life outcome for the same project is less than five percent (5%) which is found by the program to be very small (under 10%) and acceptable. The program (user's manual) clearly dictates to regard small differences and accept any sample difference that small to be a correct sample [16]. But even if the model's correct predicted outcome is taken to be (80%) that will still be considered as a very high and the model is accepted [8].

Table 4: Actual and Predicted Percentage of Building Site Overhead for the Test Sample.

Project No.	Actual real life percentage	Network output (predicted percentage)	Absolute difference %	Comments
1	8.13	8.32294	(-) 2.373185732	Correct
2	9.51	9.07061	(+) 4.620294427	Wrong
3	10.86	10.59704	(+) 2.421362799	Correct
4	10.84	11.11394	(-) 2.427121771	Correct
5	11.43	11.3421	(+) 0.769028871	Correct

As it is clear the correct predicted model outputs of the percentage of site overhead costs differ from the actual real life project percentage of site overhead costs value with a value under ( $\pm 2.476\%$ ) which is the designed model absolute difference%, which is assumed to be acceptable.

This demonstrates a very high accuracy for the proposed model and the viability of the neural network as a powerful tool for modeling the assessment of the building construction site overhead cost percentage for projects constructed in Egypt [11].

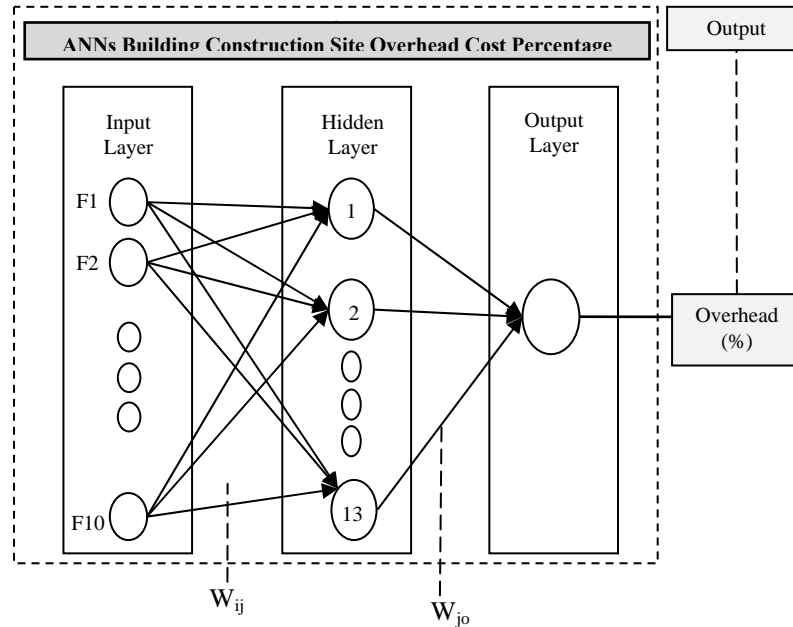


Fig. 4. Structure of the Best Model. [11]

## 7. Summary

Construction firms should carefully examine contract conditions and perform all the necessary precautions to make sure that project site overhead costs factors are properly anticipated for and covered within the total tender price. The study conducted a survey that investigated the factors affecting project's site overhead cost for building construction projects in the first and second categories of construction companies. An ANN model was developed to predict the percentage of site overhead cost for building construction projects in Egypt during the tendering process. A sample of building projects was selected as a test sample for this study. The impacts of different factors on the site overhead costs were deeply investigated. The survey results illustrated that site overhead costs are greatly affected by many factors. Among these factors come project type, size, location, site conditions and the construction technology. All of these factors make the detailed estimation of such overhead costs a more difficult task.

Hence, it is expected that a lump-sum assessment for such cost items will be a more convenience, easy, highly accurate, and quick approach. Such approach should take into consideration the different factors that affect site overhead cost. It was found that an ANN-Based Model would be a suitable tool for site overhead cost assessment.

## 8. CONCLUSIONS

The following conclusions are drawn from this research:

1. Through literature review potential factors that influence the percentage of site overhead costs for building construction projects were identified. Ten factors were identified;
2. The analysis of the collected data gathered from fifty-two real-life building construction projects from Egypt illustrated that project's duration, total contract value, projects type, special site preparation needs and project's location are identified as the top five factors that affect the value of the percentage of site overhead costs for building construction projects in Egypt;
3. Nature of the client, type of the contract and contractor-joint venture are the lowest affecting factors in the percentage of site overhead costs for building construction projects in Egypt;
4. A satisfactory Neural Network model was developed through fifty-eight experiments for predicting the percentage of site overhead costs for building construction projects in Egypt for the future projects. This model consists of one input layer with ten neurons (nodes), one hidden layer having thirteen hidden nodes with a sigmoid transfer function and one output layer. The learning rate of this model is set automatically by the N-Connection V2.0 while the training and testing tolerance are set to 0.1;

5. The results of testing for the best model indicated a testing root mean square error (RMS) value of 0.276479; and
6. Testing was carried out on five new facts (Projects) that were still unseen by the network. The results of the testing indicated an accuracy of (80%). As the model wrongly predicted the percentage of site overhead costs for only one project (20%) from the testing sample.

## 9. References

- [1] Ahuja and Campbell, Estimating from concept to completion. Prentice Hall, Englewood Cliffs, N.J, (1988).
- [2] Akintoye A. (2000). "Analysis of Factors Influencing Project Cost Estimating Practice." Construction Management and Economics. 18(1), 77-89.
- [3] Alcabes, J. (AACE, 1988), "Organizational concept for a coordinated estimating, cost control, and scheduling division".
- [4] Assaf Sadi, Abdulaziz Bubshait, Solaiman Atiyah, and Mohammed AL-Shahri, "The management of construction company overhead costs", International Journal of Project Management, Vol.19, No.5, 2001.
- [5] Bannes Lorry T., Fee analysis: A contractor's approach, Transactions of the American Association of Cost Engineers, Morgantown, WV, USA, 1994.
- [6] Becica Matt, Scott Eugene R. and Willett Andrew B., Evaluating responsibility for schedule delays on utility construction projects, Proceedings of the American Power Conference, Illinois Institute of Technology, Chicago, IL, USA, (1991).
- [7] Clough, R., and Sears, G. (1991), Construction project management. Wiley, New York.
- [8] Hatem A. A. (2009), "Developing a Neural Networks Model for Supporting Contractors In Bidding Decision In Egypt", A thesis submitted to Zagazig University in partial fulfillment to the requirement for the Master of Science Degree.
- [9] Hegazy T. and Moselhi O. (1995). "Elements of Cost Estimation: A Survey in Canada and the United States." Cost Engineering. 37(5), 27-31.
- [10] Holland, N. and Hobson, D. (1999). "Indirect cost categorization and allocation by construction contractors." Journal of Architectural Engineering, ASCE, 5(2) 49-56.
- [11] Ismaail Y. El-Sawy (2010), "Assessment of Overhead Cost for Building Construction Projects", A Thesis Submitted to Arab Academy for Science, Technology and Maritime Transport in partial fulfillment of the requirements for Master of Science Degree.
- [12] Ismaail Y. El-Sawy, Mohammed Abdel Razek, and Hossam E. Hosny (2010). "Factors Affecting Site Overhead Cost for Building Construction Projects" Journal of Al Azhar University Engineering Sector, *JAUES*, Issue 3/2010, May 2010, Cairo, Egypt.
- [13] Jones Walter B., Spreadsheet Checklist to Analyze and Estimate Prime Contractor Overhead, Cost

Engineering (Morgantown, W. Virginia), Vol.38, No.8, August 1996.

- [14] Kim In Ho, A study on the methodology of rational planning and decision of military facility construction cost, Journal of Architectural Institute of Korea, Vol.10, No.6, Ko-Korean, 1994.
- [15] Neil, Construction cost estimating for project control. Prentice-Hall, Englewood Cliffs, N.J, (1981).
- [16] N-Connection V2.0 Professional Software User Guide and Reference Manual (1997), California Scientific Software.
- [17] Peurifoy and Oberlender, Estimating construction costs, 4<sup>th</sup> Ed., McGraw Hill, New York, (1989).
- [18] Sadi Assaf, Abdulaziz Bubshait, Solaiman Atiyah, and Mohammed AL-Shahri, "Project Overhead Costs in Saudi Arabia", Cost Engineering Journal, Vol. 41, No. 4, (1999).
- [19] Yong-Woo Kim, Glenn Ballard, Case Study-Overhead Costs Analysis, Proceedings IGLC-10, Gramado, Brazil, (August, 2002).



**Ismaail Yehia Aly ElSawy**, has received his M.Sc. and B.Sc. degrees in Construction and Building Engineering, College of Engineering and Technology, from Arab Academy for Science, Technology and Maritime Transport, Alexandria, Egypt, 2010 and 2002. He joined in December (2004) the Egyptian Ministry of Electricity and Power as a Research Engineer in the Ministries National Research Center. He then joined the academic field in September (2008), as a Demonstrator (B.Sc.) then Assistant lecturer (M.Sc.) at the Civil Engineering Department, Thebes Higher Institute of Engineering. He has published more than 6 research papers in International/National Journals and Refereed International Conferences. He is interested in the implementation of Artificial Intelligence in Construction Project Management, and Construction Projects Financial Management.



# Time of Matching Reduction and Improvement of Sub-Optimal Image Segmentation for Iris Recognition

R. M. Farouk<sup>1</sup>, G. F. Elhadi<sup>2</sup>

<sup>1</sup> Department of Mathematics, Faculty of Science, Zagazig University,  
Zagazig, Egypt.

<sup>2</sup> Computer Science Department, Faculty of Computers and Information's,  
Menofia University, Menofia, Egypt.

## Abstract

In this paper, a new matching scheme based on the scalar product (SP) between two templates is used in the matching process. We also introduced the active contour technique to detect the inner boundary of the iris which is not often a circle and the circular Hough transform to determine the outer boundary of the iris. The active contour technique takes into consideration that the actual pupil boundary is near-circular contour rather than a perfect circle, which localize the inner boundary of the iris perfectly. The 1-D log-Gabor filter is used to extract real valued template for the normalized iris. We apply our system on two publicly available databases (CASIA and UBIRIS) and the numerical results show that, perfectly matching process and also the matching time is reduced. We also compare our results with previous results and find out that, the matching with SP is faster than the matching with other techniques.

**Keywords:** Biometric, Iris Recognition, Segmentation, Active Contour, Normalization, Feature Extraction, Matching, Scalar Product.

## 1. Introduction

The developments in science and technology have made it possible to use biometrics in applications where it is required to establish or confirm the identity of individuals. Applications such as passenger control in airports, access control in restricted areas, border control, database access and financial services are some of the examples where the biometric technology has been applied for more reliable identification and verification. Biometrics is inherently a more reliable and capable technique to identity human's authentication by his or her own physiological or behavioral characteristics. The features used for personnel identification by current

biometric applications include facial features, fingerprints, iris, palm-prints, retina, handwriting signature, DNA, gait, etc [17, 23]. The human iris is an annular part between pupil and sclera and its complex pattern contains many distinctive features such as arching ligaments, furrows, ridges, crypts, corona, and freckles Figure. 1. At the same time the iris is protected from the external environment behind the cornea and the eyelids. No subject to deleterious effects of aging, the small-scale radial features of the iris remain stable and fixed from about one year of age throughout one's life. The reader's two eyes, directed at this page, have identical genetics; they will likely have the same color and may well show some large scale pattern similarities; nevertheless, they have quite different iris pattern details.

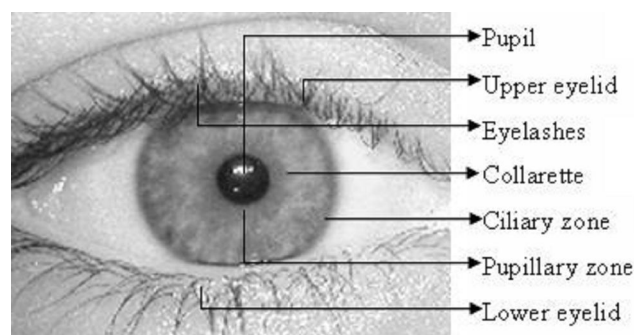


Fig 1: The image (Img 141 1 1) from the UBIRIS database

All these advantages let the iris recognition be a promising topic of biometrics and get more and more attention [7, 8, 26]. Even though iris is seen as the most reliable biometric measure, it is still not in everyday use because of the complexity of the systems. In an iris recognition system, iris location is an essential step that spends nearly more than half of the entire processing time [36]. The correctness of iris location is required for the

latter processes such as normalization, feature extraction and pattern matching. For those reasons, to improve the speed and accuracy of iris location becomes nontrivial. The algorithm proposed in this work is improvement of the matching process in the algorithms proposed by Daugman [8, 9]. The United Arab Emirates Expellees Tracking and Border Control System [22] is an outstanding example of the technology.

In general, the process of iris recognition system consists of: (i) image acquisition, (ii) Preprocessing the iris image including iris localization, image normalization and polar transformation, (iii) iris Feature extraction and (iv) iris matching.

### 1.1 Related Work

The research in the area of iris recognition has been receiving considerable attention and a number of techniques and algorithms have been proposed over the last few years. Flom and Safir first proposed the concept of automated iris recognition in [18]. The approach presented by Wildes [26] combines the method of edge detection with Hough transform for iris location. However, the parameters need to be precisely set and lengthy location time is required. Daugman's method is developed first using the integro-differential operator [10] for localizing iris regions along with removing possible eyelid noises. In the past few years, some methods made certain improvement based on the Daugman's method [8, 9]. Bowyer et al. [17] recently presented an excellent review of these methods. However, at this time, essentially all of the large scale implementations of iris recognition are based on the Daugman iris recognition algorithms [8]. The difference between a pair of iris codes was measured by their Hamming distance. Sanchez-Reillo and Sanchez-Avila [27] provided a partial implementation of the algorithm by Daugman. Boles and Boashash [34] calculated a zero-crossing representation of one-dimensional wavelet transform at various resolution levels of a concentric circle on an iris image to characterize the texture of the iris. Iris matching was based on two dissimilarity functions. [29] Decomposed an iris image into four levels using 2-D Haar wavelet transform and quantized the fourth-level high-frequency information to form an 87-bit code. A modified competitive learning neural network was adopted for classification. Tisse et al. [5] analyzed the iris characteristics using the analytic image constructed by the original image and its Hilbert transform. Emergent frequency functions for feature extraction were in essence samples of the phase gradient fields of the analytic image's dominant components [17, 31].

Similar to the matching scheme of Daugman, they sampled binary emergent frequency functions to form a feature vector and used Hamming distance for matching. Kumar et

al. [3] utilized correlation filters to measure the consistency of iris images from the same eye. The correlation filter of each class was designed using the two-dimensional Fourier transforms of training images. If the correlation output (the inverse Fourier transform of the product of the input images Fourier transform and the correlation filter) exhibited a sharp peak, the input image was determined to be from an authorized subject, otherwise an impostor one. Bae et al. [16] projected the iris signals onto a bank of basis vectors derived by independent component analysis and quantized the resulting projection coefficients as features. In another approach by Ma et al. [19] and Even Symmetry Gabor filters [10] are used to capture local texture information of the iris, which are used to construct a fixed length feature vector.

In the last year only, the iris takes the attention of many researchers and different ideas are formulated and published. For example, in [1] a bi-orthogonal wavelet based iris recognition system, is modified and demonstrated to perform  $\theta$ -angle iris recognition. An efficient and robust segmentation of noisy iris images for non-cooperative iris recognition is described in [32]. Iris image segmentation and sub-optimal images is discussed in

[13]. Comparison and combination of iris matchers for reliable personal authentication are introduced in [2]. Noisy iris segmentation, with boundary regularization and reflections removal, is discussed in [28].

### 1.2 Outline

In this paper, we first present the active contour models for iris preprocessing (segmentation step) which is a crucial step to the success of any iris recognition system, since data that is falsely represented as iris pattern data will corrupt the biometric templates generated, thus resulting in poor recognition rates. Once the iris region is successfully segmented from an eye image, the next stage is to transform the iris region so that it has fixed dimensions (normalization) in order to allow comparisons using Daugman rubber sheet model. After that the 1-D log-Gabor filter is used to extract real valued template for the normalized iris.

## 2. Iris Localization Techniques

It is the stage of locating the iris region in an eye image, whereas mentioned the iris region is the annular part

between pupil and sclera, see Figure 1. The iris segmentation has achieved by the following three main steps. The first step locates the center and radius of the iris in the input image by using the circular hough transform. Then a set of points is taken as pupil initialization from the nearby points to the iris center. The last step locates the pupil boundary points by using the region-based active contours.

## 2.1 Hough Transform

The Hough transform is a standard computer vision algorithm that can be used to determine the parameters of simple geometric objects, such as lines and circles, present in an image. The circular Hough transform can be employed to deduce the radius and center coordinates of the pupil and iris regions. For instance, recognition of a circle can be achieved by considering the strong edges in an image as the local patterns and searching for the maximum value of the circular Hough transform. An automatic segmentation algorithm based on the circular Hough transform is employed by Wildes et al. [26], and Tisse et al. [5].

The localization method, similar to Daugman's method, is also based on the first derivative of the image. In the proposed method by Wildes, an edge map of the image is first obtained by thresholding the magnitude of the image intensity gradient:

$$|\nabla G(x, y) * I(x, y)|, \quad (1)$$

$$G(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(x-x_0)^2 + (y-y_0)^2}{2\sigma^2}\right) \quad (2)$$

Where  $G(x, y)$  is a Gaussian smoothing function with scaling parameter  $\sigma$  to select the proper scale of edge analysis. Firstly, an edge map is generated by calculating the first derivatives of intensity values in an eye image and then thresholding the result. From the edge map, votes are cast in Hough space to maximize the defined Hough transform for the desired contour. Considering the obtained edge points as for the parameters of circles passing through each edge points as  $(x_i, y_i), i=1,2,3,\dots,n$ . These parameters are the center coordinates  $x_c$  and  $y_c$ , and the radius  $r$ , which are able to define any circle according to the equation:

$$x_c^2 + y_c^2 = r^2 \quad (3)$$

A Hough transform can be written as:

$$H(x_c, y_c, r) = \sum_1^n h(x_i, y_i, x_c, y_c, r) \quad (4)$$

$$h(x_i, y_i, x_c, y_c, r) = \begin{cases} 1 & \text{if } g(x_i, y_i, x_c, y_c, r) = 0 \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

Where the parametric function

$$g(x_i, y_i, x_c, y_c, r) = (x_i - x_c)^2 + (y_i - y_c)^2 - r^2.$$

Assuming a circle with the center  $(x_c, y_c)$  and radius  $r$ , the edge points that are located over the circle result in a zero value of the function  $g$ . The value of  $g$  is then transformed to 1 by the  $h$  function, which represents the local pattern of the contour. The local patterns are then used in a voting procedure using the Hough transform,  $H$ , in order to locate the proper pupil and limbus boundaries. In order to detect limbus, only vertical edge information is used. The upper and lower parts, which have the horizontal edge information, are usually covered by the two eyelids. The horizontal edge information is used for detecting the upper and lower eyelids, which are modeled as parabolic arcs.

## 2.2 Active Contour Models

Ritter et al. [24] make use of active contour models for localizing the pupil in eye images. Active contours respond to pre-set internal and external forces by deforming internally or moving across an image until equilibrium is reached. The contour contains a number of vertices, whose positions are changed by two opposing forces, an internal force, which is

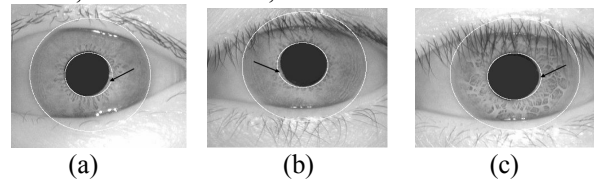


Figure 2: Errors in pupil localization by using the circular Hough transform.

dependent on the desired characteristics, and an external force, which is dependent on the image. Each vertex is moved between time  $t$  and  $t + 1$  by:

$$V_i(t+1) = V_i(t) + F_{int,i}(t) + F_{ext,i}(t) \quad (6)$$

Where  $F_{int,i}$  is the internal force,  $F_{ext,i}$  is the external force and  $V_i$  is the position of vertex  $i$ . For localization of the pupil region, the internal forces are calibrated so that the contour forms a globally expanding discrete circle. The external forces are usually found using the edge information.

In order to improve accuracy Ritter et al. use the variance image, rather than the edge image. A point interior to the pupil is located from a variance image and then a discrete circular active contour (DCAC) is created with this point as its center. The DCAC is then moved under the influence of internal and external forces until it reaches equilibrium, and the pupil is localized.

### 2.3 Discrete Circular Active Contour

Ritter (2003) et al. [25] proposed a model which detects pupil and limbus by activating and controlling the active contour using two defined forces: internal and external forces.

The internal forces are responsible to expand the contour into a perfect polygon with a radius  $\sigma$  larger than the contour average radius. The internal force  $F_{int,i}$  applied to each vertex,  $V_i$ , is defined as

$$F_{int,i} = \bar{V}_i - V_i \quad (7)$$

where  $\bar{V}_i$  is the expected position of the vertex in the perfect polygon. The position of  $\bar{V}_i$  can be obtained with respect to  $C_r$ , the average radius of the current contour, and the contour center,  $C = (C_x, C_y)$ . The center of a contour which is the average position of all contour vertices is defined as

$$C = (C_x, C_y) = \frac{1}{n} \sum_{i=1}^n V_i \quad (8)$$

The average radius of the contour is the average distance of all the vertices from the defined center point  $C$  is as the following equations

$$C_r = \frac{1}{n} \sum_{i=1}^n \|V_i - C\| \quad (9)$$

Then the position of the vertices of the expected perfect polygon is obtained as

$$\begin{aligned} \bar{V}_i = & (C_x + (C_r + \delta) \cos(2\pi i / n), \\ & C_y + (C_r + \delta) \cos(2\pi i / n)) \end{aligned} \quad (10)$$

where  $n$  is the total number of vertices.

The internal forces are designed to expand the contour and keep it circular. The force model assumes that pupil and limbus are globally circular, rather than locally, to minimize the undesired deformations due to specular reflections and dark patches near the pupil boundary. The contour detection process of the model is based on the equilibrium of the defined internal forces with the external forces. The external forces are obtained from the grey level intensity values of the image and are designed

to push the vertices inward. The magnitude of the external forces is defined as:

$$\|F_{ext,i}\| = I(V_i) - I(V_i + \hat{F}_{ext,i}) \quad (11)$$

where  $I(V_i)$  is the grey level value of the nearest neighbor to  $V_i$ .  $\hat{F}_{ext,i}$  is the direction of the external force for each vertex and it is defined as a unit vector given by:

$$\hat{F}_{ext,i} = \frac{C - V_i}{\|C - V_i\|} \quad (12)$$

Therefore, the external force over each vertex can be written as:

$$F_{ext,i} = \|F_{ext,i}\| \hat{F}_{ext,i} \quad (13)$$

The movement of the contour is based on the composition of the internal and external forces over the contour vertices. Replacement of each vertex is obtained iteratively by:

$$V_i(t+1) = V_i(t) + \beta F_{int,i}(t) + (1 - \beta) F_{ext,i}(t) \quad (14)$$

Where  $\beta$  is a defined weight that controls the pace of the contour movement and sets the equilibrium condition of internal and external forces. The final equilibrium is achieved when the average radius and center of the contour becomes the same for the first time in  $m$  iterations ago. The discrete circular active contour is applied on the three images in Figure 3.

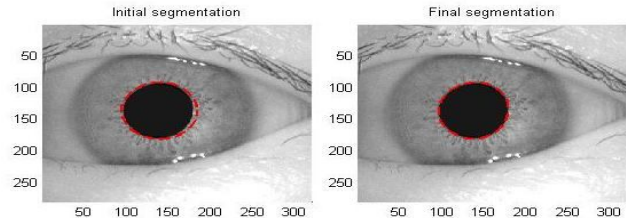


Figure 3: The segmentation of the DCA

### 2.4 Detecting Eyelids, Eyelashes and Noise Regions

The eyelids are detected by first fitting a line to the upper and lower eyelid using the linear Hough transform. A horizontal line is then drawn which intersects with the first line at the iris edge that is closest to the pupil. A second horizontal line allows the maximum isolation of eyelid regions.

Detecting eyelashes requires proper choice of features and classification procedure due to complexity and randomness of the patterns. The proposed eyelash detection by Kong et

al. consider eyelashes as two groups of separable eyelashes, which are isolated in the image, and multiple eyelashes, which are bunched together and overlap in the eye and applies two different feature extraction methods to detect eyelashes [35]. Separable eyelashes are detected



using 1-D Gabor filter, since the convolution of a separable eyelash with the Gaussian smoothing function results in a low output value.

Thus, if a resultant point is smaller than a threshold, it is noted that this point belongs to an eyelash. Multiple eyelashes are detected using the variance of intensity. If the

variance of intensity values in a small window is lower than a threshold, the center of the window is considered as a point in an eyelash. The two features combined with a

connectivity criterion would lead to the decision of presence of eyelashes. In addition, an eyelash detection method is also proposed by Huang et al. that uses the edge information obtained by phase congruency of a bank of Log-Gabor filters. The edge information is also infused with the region information to localize the noise regions [15], as in Figure 4.

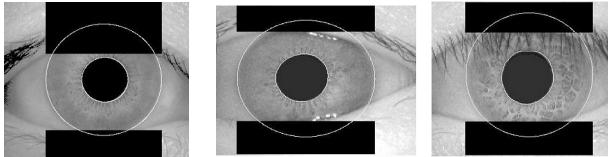


Figure 4: illustrates the perfect iris localization, where black regions denote detected eyelids and eyelashes regions.

### 3. Normalization

Once the iris region is successfully segmented from an eye image, the next stage is to transform the iris region so that it has fixed dimensions in order to eliminate dimensional inconsistencies between iris regions, and to allow comparisons. The dimensional inconsistencies between eye images are mainly due to the stretching of the iris caused by pupil dilation from varying levels of illumination. Other sources of inconsistency include, varying imaging distance, rotation of the camera, head tilt, and rotation of the eye within the eye socket. The normalization process will produce iris regions, which have the same constant dimensions, so that two images of the same iris under different conditions will have the same characteristic features at the same spatial location. A proper normalization technique is expected to transform the iris image to compensate these variations. Most normalization techniques are based on transforming iris into polar coordinates, known as unwrapping process. Pupil boundary and limbus boundary are generally two non-concentric contours. The non-concentric condition leads to different choices of reference points for transforming an iris into polar coordinates. Proper choice of reference point is very important where the radial and angular information would be defined with respect to this point. Unwrapping iris using pupil center is proposed by Boles and Boashash [34] and Lim et al. [14]. Another

reference point proposed by Arvacheh [6], which is the virtual center of a pupil with radius equal to zero (linearly-guessed center). The experiments demonstrate that the linearly-guessed center provides much better recognition accuracy. The linearly-guessed center is equivalent to the technique used by Joung et al. [4].

In addition, most normalization approaches based on Cartesian to polar transformation unwrap the iris texture into a fixed-size rectangular block. For example, in Lim et al. method, after finding the center of pupil and the inner and outer boundaries of iris, the texture is transformed into polar coordinates with a fixed resolution. In the radial direction, the texture is normalized from the inner boundary to the outer boundary into 60 pixels. The angular resolution is also fixed to a 0.8o over the 360o, which produces 450 pixels in the angular direction. Other researchers such as Boles and Boashash, Tisse et al. [5]. And Ma et al. [20] also use the fixed size polar transformation model.

However, the circular shape of an iris implies that there are different number of pixels over each radius. Transforming information of different radii into same resolution results in different amount of interpolations, and sometimes loss of information, which may degrade the performance of the system.

#### 3.1 Daugman's Rubber Sheet Model

It transforms a localized iris texture from Cartesian to polar coordinates. It is capable of compensating the unwanted variations due to distance of eye from camera (scale) and its position with respect to the camera (translation). The Cartesian to polar transformation is defined as

$$I((x(r, \theta), y(r, \theta)) \rightarrow I(r, \theta) \quad (15)$$

where

$$\begin{aligned} x(r, \theta) &= (1 - r) \times x_p(\theta) + r \times x_i(\theta), \\ y(r, \theta) &= (1 - r) \times y_p(\theta) + r \times y_i(\theta), \end{aligned}$$

and

$$\begin{aligned} x_p(\theta) &= x_{p0}(\theta) + r_p \times \cos(\theta), \\ y_p(\theta) &= y_{p0}(\theta) + r_p \times \sin(\theta), \\ x_i(\theta) &= x_{i0}(\theta) + r_i \times \cos(\theta), \\ y_i(\theta) &= y_{i0}(\theta) + r_i \times \sin(\theta), \end{aligned}$$

where  $I(x; y)$  is the iris region image,  $(x; y)$  are the original Cartesian coordinates,  $(r, \theta)$  are the corresponding normalized polar coordinates, and  $(x_p; y_p)$  and  $(x_i; y_i)$  are the coordinates of the pupil and iris boundaries along the  $\theta$  direction. The process is inherently dimensionless in the angular direction. In the radial direction, the texture is



assumed to change linearly, which is known as the rubber sheet model, as shown in Figure 5.

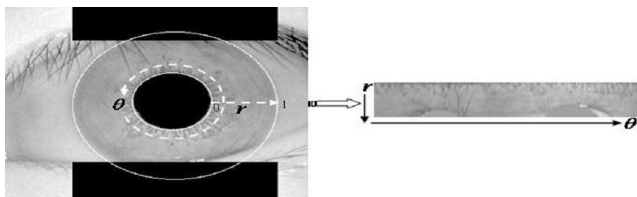


Figure 5: The rubber sheet model for normalizing the segmented irises.

The rubber sheet model [30] linearly maps the iris texture in the radial direction from pupil border to limbus border into the interval  $[0 \ 1]$  and creates a dimensionless transformation in the radial direction as well. It takes into account pupil dilation and size inconsistencies in order to produce a normalized representation of constant dimensions. In this way the iris region is modeled as a flexible rubber sheet anchored at the iris boundary with the pupil center as the reference point.

Although the normalization method compensates variations due to scale, translation and pupil dilation, it is not inherently invariant to the rotation of iris. Rotation of an iris in the Cartesian coordinates is equivalent to a shift in the polar coordinates. In order to compensate the rotation of iris textures, a best of  $n$  test of agreement technique is proposed by Daugman in the matching process. In this method, iris templates are shifted and compared in  $n$  different directions to compensate the rotational effects. The rubber sheet model is applied on 4 different iris images, as shown in Figure 6.

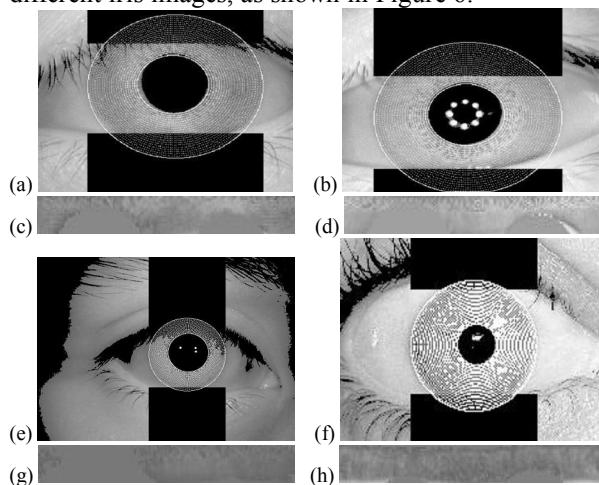


Figure 6: The normalized iris image and its polar form to four different iris images by using rubber sheet model, where (a) is an image from the CASIA-Iris V. 1, (b) is an image from the CASIA-Iris V. 3-Interval, (e) is an image from the CASIA-Iris V. 3-Lamp, and (f) is an image from the UBIRIS database. (c), (d), (g) and (h) are the polar form to the four iris images respectively.

## 4. Feature extraction

In order to provide accurate recognition of individuals, the most discriminating information present in an iris pattern must be extracted. Only the significant features of the iris must be encoded so that comparisons between templates can be made. Most iris recognition systems make use of a band pass decomposition of the iris image to create a biometric template.

The first step after the iris normalization is to extract the features from the normalized iris image. The Gabor wavelet method with log-polar transformation was designed by Daugman in 1993 and is widely used in commercialized iris recognition systems [10]. The log-Gabor wavelet method with Polar transformation was designed by Masek and Kovsesi and implemented in Matlab

[21]. Wavelets can be used to decompose the data in the iris region into components that appear at different resolutions. Wavelets have the advantage over traditional Fourier transform in that the frequency data is localized, allowing features which occur at the same position and resolution to be matched up. A number of wavelet filters, also called a bank of wavelets, are applied to the 2-D iris region, one for each resolution with each wavelet a scaled version of some basis function. The output of applying the wavelets is then encoded in order to provide a compact and discriminating representation of the iris pattern.

Some works have used multi-resolution techniques for iris feature extraction [8, 26, 34] and have proven a high recognition accuracy. At the same time, however, it has been observed that each multi-resolution technique has its specification and situation in which it is suitable; for example, a Gabor filter bank has been shown to be most known multi-resolution method used for iris feature extraction and Daugman [8] in his proposed iris recognition system demonstrated the highest accuracy by using Gabor filters.

### 4.1 The 1-D Log-Gabor Filter

The 1-D log-Gabor band pass filter is used to extract the features in an iris [35, 36], it is defined as

$$G(w) = \exp\left(\frac{-\log(w/w_0)^2}{2 \log(\sigma^2)}\right)$$

(16)

where,  $\sigma$  is used to control the filter bandwidth and  $w_0$  is the filter's center frequency, which is derived from the filter's wavelength,  $\lambda$ . The 1-D log-Gabor filter does not have a spatial domain format. Each row of the iris image, in the log-polar coordinates, is first transformed to the frequency domain using fast Fourier transform (FFT). This frequency domain row signal is then filtered with the

1-D log-Gabor filter (i.e. multiplied with the 1-D log-Gabor filter in the frequency domain).

The filtered row signal is transferred back to the spatial domain via inverse fast Fourier transform (IFFT). The spatial domain signal is then transferred to a filtered image in the spatial domain, and hence the biometric code (template) is obtained from the filtered image.

Figure 7 shows the step-by-step process of the 1-D log Gabor filter feature extraction.

## 5. Matching

Once an iris image relevant texture information extracted, the resulting feature vector (iris template) is compared with enrolled iris templates. The template generated needs a corresponding matching metric, which gives a measure of similarity between two iris templates. This metric should give one range of values when comparing templates generated from the same eye, known as intra-class comparisons, and another range of values when comparing templates created from different irises, known as extra-class comparisons.

These two cases should give distinct and separate values, so that a decision can be made with high confidence as to whether two templates are from the same iris, or from two different irises. The following subsections introduce some famous matching metrics, and finally the scalar product (SP) method.

### 5.1 The Normalized Hamming Distance

The Hamming distance (HD) gives a measure of how many bits are the same between two bit patterns, especially if the template is composed of binary values. Using the HD of two bit patterns, a decision can be made as to whether the two patterns were generated from different irises or from the same iris. For example, comparing the bit patterns P and Q, the HD is defined as the sum of disagreeing bits (sum of the exclusive-OR between P and Q) over N, the total number of bits in each bit pattern. It is known as the normalized HD, and is defined as:

$$HD = \frac{1}{N} \sum_{i=1}^N P_i \otimes Q_i \quad (17)$$

Since an individual iris region contains features with high degrees of freedom, each iris region will produce a bit-pattern which is independent to that produced by another iris, on the other hand, two iris codes produced from the same iris will be highly correlated.

In case of two completely independent bit patterns, such as iris templates generated from different irises, the HD

between the two patterns should equal 0.5. This occurs because independence implies that, the two bit patterns will be totally random, so there is 0.5 chance of setting any bit to 1, and also to zero. Therefore, half of the bits will agree and half will disagree between the two patterns. If two patterns are derived from the same iris, the HD between them will be close to 0.0, since they are highly correlated and the bits should agree between the two iris codes.

Daugman [8] uses this matching metric as following, the simple Boolean Exclusive-OR operator (XOR) applied to the 2048 bit phase vectors that encode any two iris patterns, masked (AND'ed) by both of their corresponding mask bit vectors to prevent noniris artifacts from influencing iris comparisons. The XOR operator  $\otimes$  detects disagreement between any corresponding pair of bits, while the AND operator  $\cap$  ensures that the compared bits are both deemed to have been uncorrupted by eyelashes, eyelids, specular reflections, or other noise. The norms  $\| \cdot \|$  of the resultant bit vector and of the AND'ed mask vectors are then measured in order to compute the fractional HD (Equation 5.18), as the measure of dissimilarity between any two irises, whose two phase code bit vectors are denoted  $codeP$ ;  $codeQ$  and whose mask bit vectors are denoted  $maskP$ ;  $maskQ$ :

$$HD = \frac{\| (codeP \otimes codeQ) \cap maskP \cap maskQ \|}{\| maskP \cap maskQ \|} \quad (18)$$

The denominator tallies the total number of phase bits that mattered in iris comparisons after artifacts such as eyelashes, eyelids, and specular reflections were discounted, so the resulting HD is a fractional measure of dissimilarity; 0.0 would represent a perfect match.

### 5.2 The Weighted Euclidean Distance

The weighted Euclidean distance (WED) can be used to compare two templates, especially if the template is composed of integer values. It gives a measure of how similar a collection of values are between two templates. This metric is employed by Zhu et al. [37] and is defined as:

$$WED (P) = \sum_{i=1}^N \frac{(f_i - f_i^p)^2}{(\delta_i^p)^2} \quad (19)$$

where  $f_i$  is the  $i^{th}$  feature of the unknown iris, and  $f_i^p$  is the  $i^{th}$  feature of iris template  $k$ , and  $\delta_i^p$  is the standard deviation of the  $i^{th}$  feature in iris template  $k$ . The unknown iris template is found to match iris template  $k$ , when the WED is a minimum at  $k$ .

### 5.3 The Normalized Correlation

Wildes et al. [26] make use of Normalized correlation (NC) between the acquired and database representation for goodness of match. This is represented as:

$$NC = \frac{\sum_{i=1}^m \sum_{j=1}^n (p_1[i, j] - \mu_1)(p_2[i, j] - \mu_2)}{m * n * \sigma_1 * \sigma_2} \quad (20)$$

where  $p_1$  and  $p_2$  are two images of size  $m \times n$ ,  $\mu_1$  and  $\sigma_1$  are the mean and standard deviation of  $p_1$ , and  $\mu_2$  and  $\sigma_2$  are the mean and standard deviation of  $p_2$ . Normalized correlation is advantageous over standard correlation, since it is able to account for local variations in image intensity that corrupt the standard correlation calculation.

### 5.4 The Scalar Product

The Scalar product method (SP) can be used to compare two templates, especially if the template is composed of real values. It considers the two templates as two vectors and gives the  $\cos(\theta)$  between the two templates. The  $\cos(\theta)$  between any two templates is between -1 and 1. If  $\cos(\theta)$  is close to 1, the two templates are for the same iris, but if it was close to zero, the templates are for different irises. For example suppose that we have two templates P and Q, the scalar product is defined as:

$$P \cdot Q = \|P\| \|Q\| \cos(\theta) \quad (21)$$

The localization of the iris and the coordinate system desc-

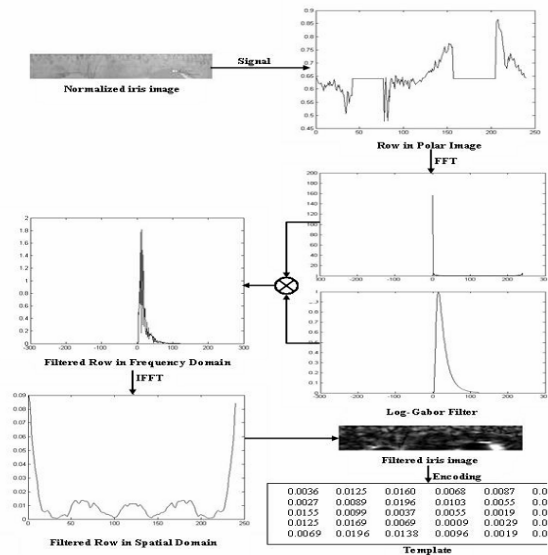


Figure 7: The step-by-step process of a row signal feature extraction by using the 1-D log-Gabor filter

Iter. No	Templates	$\cos(\theta)$
1	P = [2 6 2 1 2 8 5 10 2 3], Q = [1 2 3 4 5 6 7 8 9 10]	0:7881
2	P = [2 6 2 1 2 8 5 10 2 3], Q = [2 3 4 5 6 7 8 9 10 1]	0:8235
3	P = [2 6 2 1 2 8 5 10 2 3], Q = [3 4 5 6 7 8 9 10 1 2]	0:8911
4	P = [2 6 2 1 2 8 5 10 2 3], Q = [4 5 6 7 8 9 10 1 2 3]	0:7013
5	P = [2 6 2 1 2 8 5 10 2 3], Q = [5 6 7 8 9 10 1 2 3 4]	0:6723
6	P = [2 6 2 1 2 8 5 10 2 3], Q = [6 7 8 9 10 1 2 3 4 5]	0:5469
7	P = [2 6 2 1 2 8 5 10 2 3], Q = [7 8 9 10 1 2 3 4 5 6]	0:6144
8	P = [2 6 2 1 2 8 5 10 2 3], Q = [8 9 10 1 2 3 4 5 6 7]	0:7141
9	P = [2 6 2 1 2 8 5 10 2 3], Q = [9 10 1 2 3 4 5 6 7 8]	0:7817
10	P = [2 6 2 1 2 8 5 10 2 3], Q = [10 1 2 3 4 5 6 7 8 9]	0:7206

Table 1: This table indicates that, the maximum  $\cos(\theta) = 0:8911$ , thus  $\theta = 26:988$  which is the smallest  $\theta$  between the two templates. i.e., there is no match between the two templates for ever.

The previous table is for a simple example, but for iris the algorithm will perform 4800 iterations for comparing every two templates, because each template consists of 4800 elements.

## 6. Results

The actual iris image was first segmented using the gradient-based Hough transform to detect the outer iris boundary, and the DCAC for the inner iris boundary to avoid the errors of Hough transform, and then the eyelids, eyelashes, and noise regions are detected. Secondly the detected iris image is normalized using Daugman's rubber sheet model. After that the relevant texture information is extracted using the 1-D Log-Gabor filter, hence we have a real valued template of  $20 \times 240$  elements which will be converted to a vector of  $1 \times 4800$  elements. Finally these templates are stored to comprise a database of templates which will be used in the matching process by using the scalar product method.

This database of templates has two categories, the CASIA which consists of 996 templates and UBIRIS which consists of 723 templates. The SP method was tested by using 915 and 448iris images from CASIA and UBIRIS d-

ribed above achieve invariance to the 2-D position and size of the iris, and to the dilation of the pupil within the iris. However, it would not be invariant to the orientation of the iris within the image plane. The most efficient way to achieve iris recognition with orientation invariance is not to rotate the image itself using the Euler matrix, but rather to compute the iris phase code in a single canonical orientation and then to compare this very compact representation at many discrete orientations by cyclic scrolling of its angular variable. Thus for example to apply the SP method on two different templates P = [2 6 2 1 2 8 5 10 2 3], reference template and Q = [1 2 3 4 5 6 7 8 9

10], template from the database of 10 elements, it will work as shown in Table 1.

database respectively, and was found to give good correct recognition rates compared to other matching methods as shown in Table 2.

Matching measure	Correct recognition rate (CRR)%
WED	98.73
SP	98.26
HD	98.22

Table 2: The correct recognition rates achieved by three matching measures using the CASIA and UBIRIS database.

In our experimental results the false match rate (FMR), the rate which non-authorized people are falsely recognized during the feature comparison which contrasts the false accept rate (FAR) and the false non-match rate (FNMR), the rate that authorized people are falsely not recognized during feature comparison which contrasts the false reject rate (FRR) are estimated. Figure 8, illustrates the receiver operating characteristic (ROC) curves for the CASIA database after applying the SP matching method. Where 100-FNMR is plotted vs. the FMR.

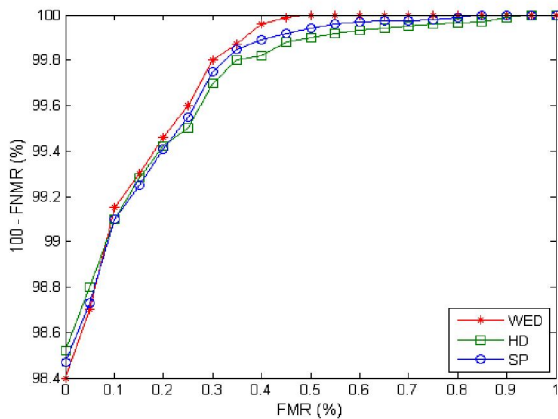


Figure 8: The obtained ROC curves to three different matching measures using the CASIA database.

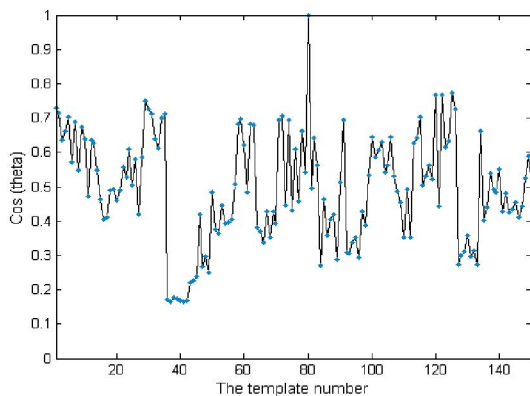


Figure 9: The matching of (012 1 3) iris image from (CASIA-Iris V. 1) with the template number 80 from 150 templates, where as shown  $\cos(\theta) = 1$

between the compared iris template and the template number 80, hence the two are templates for the same iris image.

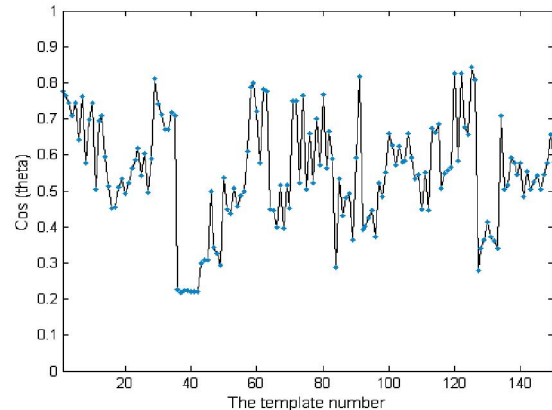


Figure 10: There is no match of (050 1 3) iris image (CASIA-Iris V. 1) with any template from 150 templates, where the maximum  $\cos(\theta) = 0.83$  is between the compared iris template and the template number 124, hence the two templates are very similar but they are not templates for the same iris image.

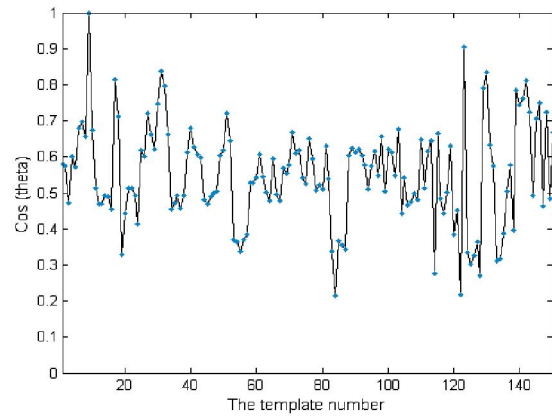


Figure 11: The matching of (Img 2 1 4) iris image from (UBIRIS database) with the template number 9 from 150 templates, where as shown  $\cos(\theta) = 1$  between the compared iris template and the template number 9, hence the two are templates for the same iris image.

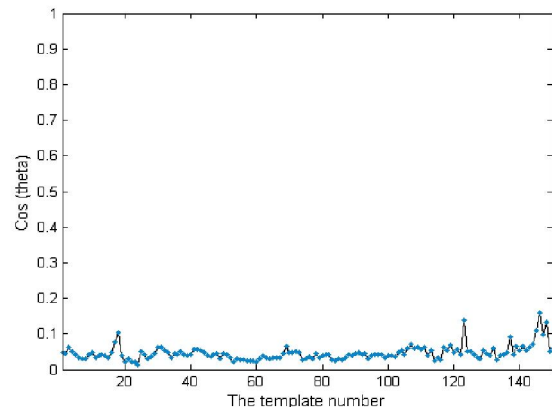


Figure 12: There is no match of (Img 235 1 5) iris image (UBIRIS database) with any template from 150 templates, where the maximum



$\cos(\theta) = 0.18$  is between the compared iris template and the template number 145, hence the two templates are not so similar and also they are not templates for the same iris image.

## 7. Conclusion

Here we have presented an active contour model, in order to compensate for the iris detection error caused by two circular edge detection operations. After perfect iris localization, the segmented iris region is normalized (transformed into polar coordinates) to eliminate dimensional inconsistencies between iris regions. This was achieved by using Daugman's rubber sheet model, where the iris is modeled as a flexible rubber sheet, which is unwrapped into a rectangular block with constant polar dimensions  $(20 \times 240)$  elements.

The next stage is to extract the features of the iris from the normalized iris region. This was done by the convolution of the 1-D Log-Gabor filters with the normalized iris region. After that the convoluted iris region is reshaped to be a template of  $(1\_4800)$  real valued elements.

Finally the scalar product matching scheme is used, which give the  $\cos(\theta)$  between two templates. If  $\cos(\theta) = 1$  between two templates P and Q this means that, the two templates were deemed to have been generated from the same iris, otherwise they have been generated from different irises.

## References

- [1] A. Abhyankara, S. Schuckers, A novel biorthogonal wavelet network system for o\_angle iris recognition, *Pattern Recognition*, 43 (2010), 987-1007.
- [2] A. Kumar, A. Passi, Comparison and combination of iris matchers for reliable personal authentication, *Pattern Recognition*, 43 (2010), 1016-1026.
- [3] B. Kumar, C. Xie, J. Thornton, A. Bovik, Iris verification using correlation filters, *Proceedings of 4th International Conference on Audio- and Video- Based Biometric Person Authentication*, (2003), 697-705.
- [4] B. J. Joung and C. H. Chung and K. S. Lee and W. Y. Yim and S. H. Lee, On Improvement for Normalizing Iris Region for a Ubiquitous Computing, *Proceedings of International Conference on Computational Science and Its Applications ICCSA, Singapore*, (2005), 1213-1219.
- [5] C. Tisse, L. Martin, L. Torres, M. Robert, Person Identification Technique Using Human Iris Recognition, *Proc. Vision Interface*, (2002), 294-299.
- [6] E. M. Arvacheh, A Study of Segmentation and Normalization for Iris Recognition Systems, University of Waterloo, Waterloo, Ontario, Canada, (2006).
- [7] J. Daugman, Demodulation by Complex-valued Wavelets for Stochastic Pattern Recognition, *International Journal of Wavelets, Multiresolution and Information Processing*, 1 (1) (2003), 1-17.
- [8] J. Daugman, How Iris Recognition Works, *IEEE Transactions on Circuits and Systems for Video Technology*, 14 (1) (2004), 21-30.
- [9] J. Daugman, The Importance of Being Random: Statistical Principles of Iris Recognition, *Pattern Recognition*, 36 (2) (2003), 279-291.
- [10] J. Daugman, High Confidence Visual Recognition of Persons by a Test of Statistical Independence, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15 (11) (1993), 1148-1160.
- [11] J. Daugman, Statistical Richness of Visual Phase Information: Update on Recognizing Persons by Iris Patterns, *Int. J. Computer Vision*, 45 (1) (2001), 25-38
- [12] J. Havlicek, D. Harding, A. Bovik, The multi-component AM-FM image representation, *IEEE Trans. Image Process*, 5 (1996), 1094-1100.
- [13] J. R. Matey, R. Broussard, L. Kennell, Iris image segmentation and sub-optimal images, *Image and Vision Computing*, 28 (2010), 215-222.
- [14] J. Huang, Y. Wang, T. Tan, J. Cui, A new iris segmentation method for recognition, *Proceedings of the 17th International Conference on Pattern Recognition, ICPR*, 3 (2004), 554-557.
- [15] J. Huang, Y. Wang, T. Tan, J. Cui, A new iris segmentation method for recognition, *Proceedings of the 17th International Conference on Pattern Recognition, ICPR*, 3 (2004), 554-557.
- [16] K. Bae, S. Noh, J. Kim, Iris feature extraction using independent component analysis, *Proceedings of 4th International Conference on Audio- and Video-Based Biometric Person Authentication*, (2003), 838-844.
- [17] K. W. Boweyer, K. Hollingsworth, Patrick J. Flynn, Image understanding for iris biometrics: A survey, *Computer Vision and Image Understanding*, 110 (2008), 281-307.
- [18] L. Flom, A. Safir, Iris recognition system, U.S. Patent, 4 (1987), 394-641.
- [19] L. Ma, T. Tan, Y. Wang, D. Zhang, Efficient Iris Recognition by Characterizing Key Local Variations, *IEEE Transactions on Image Processing*, 13 (2004), 739-750.
- [20] L. Ma and T. Tan and Y. Wang and D. Zhang, Personal identification based on iris texture analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25 (12) (2003), 1519-1533.
- [21] L. Masek, P. Kovsi, MATLAB Source Code for a Biometric Identification System Based on Iris Patterns, The University of Western Australia, (2003).
- [22] M. Almualla, The UAE Iris Expellees Tracking and Border Control System in: *Biometrics Consortium September*, Crystal City, VA, (2005).
- [23] N. Duta, A survey of biometric technology based on hand shape, *Pattern Recognition*, 42 (2009), 2797-2806.
- [24] N. Ritter, Location of The Pupil-iris Border in Slit-lamp Images of The Cornea, *Proceedings of the International Conference on Image Analysis and Processing*, (1999).
- [25] N. Ritter, J. R. Cooper, Locating the iris: A first step to registration and identification, *Proceedings of the 9th*



- IASTED International Conference on Signal and Image Processing, (2003), 507-512.
- [26] R. P. Wildes, Iris Recognition: An Emerging Biometric Technology, Proceedings of the IEEE, 85 (9) (1997), 1348-1363.
- [27] R. Sanchez-Reillo, C. Sanchez-Avila, Iris recognition with low template size, Proceedings of International Conference on Audio- and Video-Based Biometric Person Authentication, (2001), 324-329.
- [28] R. D. Labati, F. Scotti, Noisy iris segmentation with boundary regularization and reflections removal, Image and Vision Computing, 28 (2010), 270-277.
- [29] S. Lim, K. Lee, O. Byeon, T. Kim, Efficient iris recognition through improvement of feature vector and classifier, ETRI J. 23, 2 (2001), 1-70.
- [30] S. Sanderson and J. H. Erbetta, Authentication for secure environments based on iris scanning technology, IEEE Colloquium on Visual Biometrics, (2000), 1-8.
- [31] T. Tangsukson, J. Havlicek, AM-FM image segmentation, Proceedings of IEEE International Conference on Image Processing, (2000), 104-107.
- [32] T. Tan, Z. He, Z. Sun, Efficient and robust segmentation of noisy iris images for non-cooperative iris recognition, Image and Vision Computing, 28 (2010), 223-230.
- [33] V. A. Pozdin, Y. Du, Performance analysis and parameter optimization for iris recognition using log-Gabor wavelet, SPIE Electronic Imaging 6491, (2007), 1-11.
- [34] W. Boles, B. Boashash, A Human Identification Technique Using Images of The Iris and Wavelet Transform, IEEE Transactions on Signal Processing, 46 (1998), 1185-1188.
- [35] W. Kong, D. Zhang, Eyelash detection model for accurate iris segmentation, Proceeding of ISCA 16th International Conference on Computers and their Applications, (2001), 204-207.
- [36] X. Ye, Z. Zhuang, Y. Zhuang, A New and Fast Algorithm of Iris Location, Computer Engineering and Applications, 30 (2003), 54-56.
- [37] Y. Zhu, T. Tan, Y. Wang, Biometric personal identification based on iris patterns, Proceedings of the 15<sup>th</sup> International Conference on Pattern Recognition, Spain, 2 (2000).

# Recurrent Neural Networks Design by Means of Multi-Objective Genetic Algorithm

## Case study : Phoneme Recognition

Hanen Chihi<sup>1</sup> and Najet Arous<sup>2</sup>

Institut Supérieur d'Informatique, ISI  
Département Génie Logiciels et Systèmes d'Information, GLSI  
Université Tunis El Manar, Tunis Tunisie

### Abstract

Evolutionary algorithms are considered more efficient for optimal system design because they can provide higher opportunity for obtaining the global optimal solution. This paper introduces a method for construct and train Recurrent Neural Networks (RNN) by means of Multi-Objective Genetic Algorithms (MOGA). The use of a multi-objective evolutionary algorithm allows the definition of many objectives in a natural way. The case study of the proposed model is the phoneme recognition. We have shown that the proposed model is able to achieve good results in recognition tasks.

**Keywords:** *Recurrent neural network, Genetic algorithm, Phonemes recognition, Multi-objective optimization.*

## 1. Introduction

Recurrent Neural Networks (RNN) represent a large and varied class of computational models that are designed by more or less detailed analogy with biological brain modules. In this paper we focus on the use a particular network : Elman-type recurrent networks in witch the hidden layer is returned to the input layer [7].

In recent years, gradient-based RNN solved many tasks [26]. The Back-propagation, however, has two major limitations: a very long training process, with problems such as local minima and network design. The back-propagation algorithm adjusts exclusively the connection weights for particular network architecture, but the algorithm does not adjust the network architecture to define the optimum Neural Network (NN) for a particular problem [3], [25]. To overcome these restrictions, various methods for auto-design NN have been proposed [2], [10].

Genetic Algorithms (GA) are a search heuristic that mimics the process of natural evolution. They maintain a population of solution candidates and evaluate the fitness of each solution according to a specific fitness function. Even though, GA are not guaranteed to find the global

optimum, they can find an acceptable solution relatively in a wide range of problems [4].

Various combinations of GA and NN have been investigated [3], [10], [24]. Much research concentrates on the acquisition of parameters for a fixed network architecture [6], [9]. Other work allows a variable topology, but disassociates structure acquisition from acquisition of weight values by interweaving a GA search for network topology with a traditional parametric training algorithm over weights [2], [10]. Some studies attempt to co-evolve both the topology and weight values within a GA framework, but the network architectures are restricted [15].

Many researches exist, describing multitude applications for GA [4]. A substantial proportion of these applications involve the evolution of solutions to problems with more than one objective [13], [22], [27]. More specifically, such problems consist of several separate objectives, with the required solution being one where some or all of these objectives are satisfied to a greater or lesser degree.

Multi-objective genetic algorithms (MOGA) have been widely used for the evolution of NN. Dehuri and Cho [11] propose a multi-criterion pareto GA used to train NN for classification problems. Delgado and Pegalajar [12] propose a MOGA for obtaining the optimal size of RNN for grammatical inference.

In this study, we combine RNN with MOGA to provide an alternative way for optimizing both RNN structure and weights. An important aspect of our work is the use of multi-objective optimization to evaluate the ability of new RNN [20]. The use of different objectives for each network allows a more accurate estimation of the goodness of a network.

This paper is organized as follows. Section 2 explains the application of multi-objective optimization to the problem

of fitness estimation. Section 3 describes the proposed constructive multi-objective RNN. Section 4 presents the experimental results obtained on the classification of the TIMIT vowels.

## 2. Multi-objective optimization

In this section, we briefly present the formulation of a multi-objective optimization problem (MOO) such as some required notions about Pareto based multi-objective optimization and some concepts relating to Pareto optimality [2], [12].

The scenario considered in this paper involves an arbitrary optimization problem with  $k$  objectives, which are, without loss of generality, all to be minimized and all equally important, i.e., no additional knowledge about the problem is available. We assume that a solution to this problem can be described in terms of a decision vector denoted by:

$$x = (x_1, x_2, \dots, x_n) \quad (1)$$

where  $x_1, x_2, \dots, x_n$  are the variables of the problem.

Mathematically, the multi-objective optimization problem is stated by :

$$MOO : \begin{cases} \min F(x) = (f_1(x), f_2(x), \dots, f_k(x)), \\ s.c. x \in C. \end{cases} \quad (2)$$

where  $f_i$  are the decision criteria and  $k$  is the number of objective function.

An optimization problem searches the action  $x^*$  where the constraints  $C$  are satisfied and the objective function  $F(x)$  is optimized.

In practical applications, there is no solution that can minimize all of the  $k$  objectives. As a result, MOO problems tend to be characterized by a family of alternatives solutions.

The approach most used is to weight and sum the separate fitness values in order to produce just a single fitness value for every solution, thus allowing the GA to determine which solutions are fittest as usual. However, as noted by Goldberg [14], the separate objectives may be difficult or impossible to manually weight because of unknowns in the problem. Additionally, weighting and summing could have a detrimental effect upon the evolution of acceptable solutions by the GA (just a single incorrect weight can cause convergence to an unacceptable solution).

The concept of Pareto-optimality helps to overcome this problem of comparing solutions with multiple fitness values. A solution is Pareto optimal if it is not dominated by any other solutions. A Pareto optimal solution is defined as follows: a decision vector  $x$  is said to dominate a decision vector  $y$  if and only if  $\forall i \in \{1, \dots, k\} : f_i(x) \leq f_i(y) \wedge \exists j \in \{1, \dots, k\} : f_j(x) < f_j(y)$ . The decision vector  $x$  is Pareto optimal if and only if  $x$  is non-dominated [5].

The Pareto approach is based on two aspects: the ranking and the selection. The ranking methods are the following:

- NDS (Non Dominated Sorting) : In this method, the rank of an individual is the number of solutions dominating this individual plus one [12].
- WAR (Weighted Average Ranking) : In this method, population members are ranked separately according to each objective function. Fitness equal to the sum of the ranks in each objective is assigned [2].
- NSGA (Non-dominated Sorting Genetic Algorithm) [21]: In this method, all non-dominated individuals of the population have rank 1. Then, these individuals are removed and the next set of non-dominated individuals are identified and assigned next rank [21].

Several methods of selection based on the concept of dominance are:

- Tournament based selection [2]: at each tournament, two individuals  $A$  and  $B$  fall in competition against a set of  $t_{dom}$  individuals in the population. If the competitor  $A$  dominates all individuals and all the other competitor  $B$  is dominated by at least one individual, then individual  $A$  is selected.
- Pareto reservation strategy [5]: in this method, the non-dominated individuals are always saved to the next generation.
- Ranking method [2]: the cost associated with a new individual is determined by the relative distance in objective space with respect to individuals not dominated of the current population.

## 3. Recurrent neural networks design by means of multi-objective genetic algorithm

We shall now tackle the problem of finding RNN having the smallest recognition error and the least number of hidden units. For this reason, we formulate the problem as optimisation algorithm, more specifically, as a matter of MOO. In order to solve it we shall use an algorithm based on Pareto optimality that

his goal is to optimize three objectives. A performance goal minimizes the recognition error (to maximize the successes in the testing set). Two goals of diversity to increase diversity in the population : mutual information and internal diversity.

In this paper we propose a model called Recurrent Neural Networks Design by means of Multi-Objective Genetic Algorithm (RNND-MOGA). It reflects the types of networks that arise from a RNN performing both structural and weight learning. The general architecture of RNND-MOGA is straightforward. Input and output units are considered to be provided by the task and they are immutable by the algorithm; thus each network for a given task always has  $m_{in}$  input units and  $m_{out}$  output units. The number of hidden units and bias varies from 0 to a user supplied maximum  $h_{max}$ .

The proposed hybrid learning process is the following (see Fig. 1). In each generation, networks are first evaluated using Pareto optimisation algorithm. The best  $P\%$  RNN are selected for the next generation; all other networks are discarded and replaced by mutated copies of networks selected by proportional selection. Generating an offspring is done using only two types of mutation operators : the parametric mutation and the structural mutation. The parametric mutation alters the value of parameters (link weights) currently in the network, whereas structural mutation alters the number of the hidden units, thus altering the space of parameters.

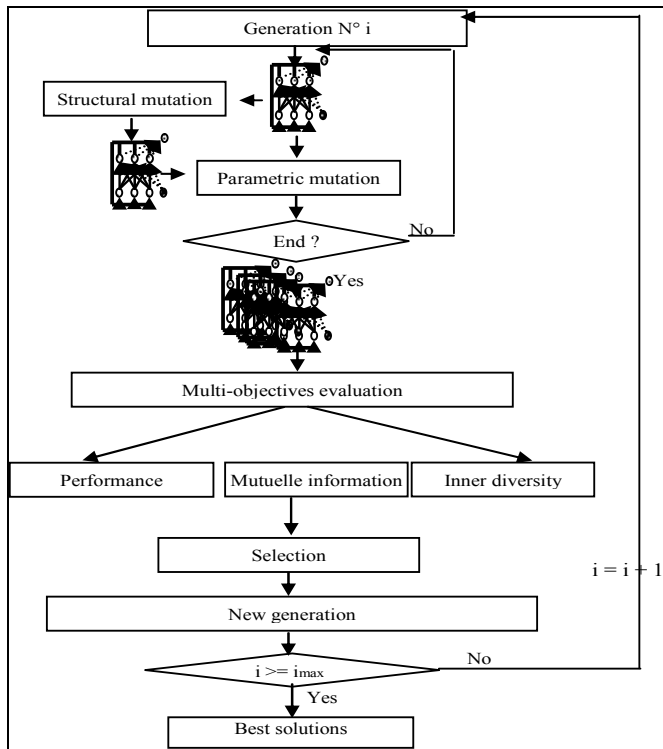


Fig. 1. Proposed evolution strategy

### 3.1 Encoding

The proposed chromosome representation is a structure encoding the learning parameters, the weights and the bias. The chromosome structure is a record composed of these attributes :

- IW: matrix of the RNN input weights ;
- LW: matrix of the RNN connection weights ;
- b1: vector of RNN bias ;
- trainPrm: save learning rate and epochs number ;
- learnFcn: save the learning function of an RNN.

### 3.2 Initialization

The proposed algorithm initializes the population with randomly generated RNN. The number of hidden units for each one is chosen from a uniform distribution in a user defined range ( $0 \leq h \leq h_{max}$  |  $h_{max}$  is the maximum number of hidden units in the network). Once a topology has been chosen, all links are assigned weights random initialized.

### 3.3 Genetic operators

GA used here is a modified algorithm. The main differences compared to the standard GA are that there is no crossover and structural mutations are added. Both of the mutation operators will be described in detail below. The crossover operator, which combines genes from two individuals, is rarely useful when evolving NN and is therefore not used here.

The parametric mutation changes the weights of a network without changing its topology. In this work, we use the back-propagation algorithm as a parametric mutation operator. It is run using a low learning rate for few epochs. In our model, this epochs number is randomly chosen within a user defined rang. The network is allowed to draw lessons from the training set, but it is also prevented from being too similar to the rest of networks. The parametric mutation is always performed after the structural one, because it does not alter the structure of a network and it is used to adapt mated networks.

Fig. 2 describe two types of structural mutation :

- *Add hidden units:* Generating an offspring using structural mutation involves three steps: copying the parent, determining the severity of the mutations to be performed, and finally mutating the copy. The severity of a mutation of a given parent is dictated by its score. It defines the number of hidden units to be added. Networks with a low score suffer a severe mutation, and those with a high score are undergoing a slight transformation. Equations (3) and (4) calculate, respectively, the

severity of mutation and the number of hidden units to add.

$$T(i) = 1 - \frac{Score(i)}{\sum_{k=1}^N Score(k)} \quad (3)$$

$$HU(i) = \lceil \Delta_{min} + \alpha T(i)(\Delta_{max} - \Delta_{min}) \rceil \quad (4)$$

where  $Score(i)$  represents the score of the  $i^{th}$  individual,  $\Delta_{min}$  and  $\Delta_{max}$  are respectively the minimum and maximum number of hidden units to be add,  $\alpha$  is a random value between 0 and 1.

Once the number of units to be added is determined, we modify the network structure under the new constraints and the connections' weights of these units are randomly initialized.

- *Remove hidden units:* This type of mutation is used to remove the hidden units that do not contribute to improve recognition of the network. The process of deleting a hidden unit occurs as follows. In the first step, we seek the inactive unit among hidden units of the network. This is done by calculating the score of each hidden unit using equation (5). It calculates the difference in score of RNN with and without the hidden unit. The unit having the lowest fitness is eliminated.

$$S_u(i) = Score(i) - Score_u(i) \quad (5)$$

where  $Score(i)$  is the generalisation rate of the  $i^{th}$  RNN,  $Score_u(i)$  is the generalization rate of the  $i^{th}$  RNN without the  $u^{th}$  unit.

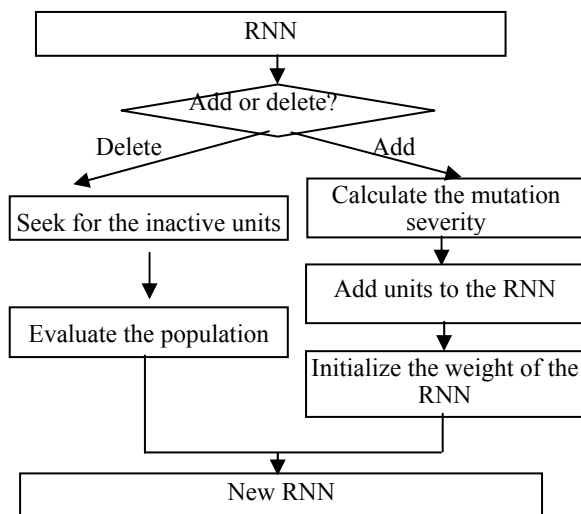


Fig. 2. Structural mutation strategy

### 3.4 Multi-objective optimization

A promising approach for performing optimization problems is the MOGA aiming at producing Pareto optimal solutions [11]. The key concept here is dominance. However, the success of a Pareto optimal GA depends largely on its ability to maintain diversity. Usually, this is achieved by employing niching techniques such as fitness sharing [5] and the inclusion of some useful measures applied to other models, such as negative correlation or mutual information [17]. The MOGA employed in this work can be described as a niched Pareto GA with NSGA [21] and tournament selection [2]. The algorithm uses a specialised tournament selection approach, based on the concept of dominance.

The proposed algorithm is based on the concept of Pareto optimality [19]. We consider a population of networks where the  $i^{th}$  individual characterised by a vector of objectives values. In fact, the population has  $N$  individuals and  $M$  objectives are considered. In our study, tree objectives are considered.

In this paper, we define the following four objectives:

- *Objective of performance:* The performance of RNN is given by its generalization rate.
- *Mutual information:* The mutual information between RNN  $f_i$  and  $f_j$  is given by equation (6) :

$$O_{MI}(f_i, f_j) = -\frac{1}{2} \log(1 - \rho_{ij}^2) \quad (6)$$

where  $\rho_{ij}$  is the correlation coefficient between the networks. The objective is the average of mutual information between each pair of networks [18].

- *Internal diversity:* The internal diversity of a RNN measures the difference between the outputs of the networks [16]. The internal diversity of the  $k^{th}$  RNN is given by equation (7) :

$$O_{ID}(i) = \frac{1}{N-1} \sum_{j=1, j \neq i}^N O_{FD}(f_i, f_j) \quad (7)$$

where  $N$  is the size of the RNN population,  $P$  is the number of training vectors and  $O_{FD}$  is the functional diversity between the  $i^{th}$  and the  $j^{th}$  network :

$$O_{FD}(f_i, f_j) = \frac{1}{P} \sum_{k=1}^P \|f_i(x_k) - f_j(x_k)\| \quad (8)$$

where  $x_k$  is the  $k^{th}$  training vectors.



## 4. Experimental results

In this section, we evaluate and compare the described and the proposed evolutionary constructive RNN for continuous speech recognition on the maco-class of vowels of TIMIT speech corpus [1].

### 4.1 Database description

The third component is a phoneme recognition module. The speech database used is the DARPA TIMIT acoustic-phonetic continuous speech corpus which contains: /iy/, /ih/, /eh/, /ey/, /ae/, /aa/, /aw/, /ay/, /ah/, /ao/, /oy/, /ow/, /uh/, /uw/, /ux/, /er/, /ax/, /ix/, /axr/ and /ax-h/. The corpus contains 13 699 phonetic unit for training and 4041 phonemes for testing.

Speech utterance was sampled at a sampling rate of 16 KHz using 16 bits quantization. Speech frames are filtered by a first order filter. After the pre-emphasis, speech data consists of a large amount of samples that present the original utterance. Windowing is introduced to effectively process these samples. This is done by regrouping speech data into several frames. A 256 sample window that could capture 16 ms of speech information is used. To prevent information lost during the process, an overlapping factor of 50% is introduced between adjacent frames.

Thereafter, mel frequency cepstral analysis was applied to extract 12 mel cepstrum coefficients (MFCC) [8].

Among all parameterization methods, the cepstrum has been shown to be favourable in speech recognition and is widely used in many automatic speech recognition systems [23]. The cepstrum is defined as the inverse Fourier transform of the logarithm of the short-term power spectrum of the signal. The use of a logarithmic function permits us to deconvolve the vocal tract transfer function and the voice source. Consequently, the pulse sequence originating from the periodic voice source reappears in the cepstrum as a strong peak in the 'quefreny' domain. The derived cepstral coefficients are commonly used to describe the short-term spectral envelope of a speech signal. The advantage of using such coefficients is that they induce a data compression of each speech spectral vector while maintaining the pertinent information it contains. The mel-scale is a mapping from a linear to a nonlinear frequency scale based on human auditory perception. It is proved that such a scale increases significantly the performance of speech recognition systems in comparison with the traditional linear scale. The computation of MFCC requires the selection of M critical bandpass filters. To obtain the MFCC, a discrete cosine transform, is applied to the output of M filters. These filters are triangular and cover the 156 – 6844 Hz frequency range; they are spaced on the mel-frequency scale. This scale is logarithmic above 1 kHz and linear

below this frequency. These filters are applied to the log of the magnitude spectrum of the signal, which is estimated on a short-time basis.

### 4.2 Discussion

In the experiments below, the number of hidden units for networks of the initial population was selected uniformly between 1 and 5. Each network has 12 input units representing the 12 MFCC coefficients and 20 output units representing the TIMIT vowels. Table 1 represents the parameter setting.

In this section, results produced by the proposed model will be presented and compared with results produced by the Elman model using 30 hidden units the GA and the Elman model using 16 hidden units (the best topology given by the proposed model).

Table 1: Learning parameters of the proposed model

Parameter name	Value
Learning rate for the training of the Elman model	0.5
Epochs number for the training of the Elman model	100
Mutation rate for the standard GA	0.8
Crossover rate for the standard GA	0.4
Structural mutation rate	0.2
Parametric mutation rate	0.3
Generation number of the population of networks	20

The learning process of the GA used for comparison is the following. First, a population of chromosomes is created and initialised randomly. Then, a roulette selection is used to select individuals to be reproduced. Thereafter, a one-point crossover operator is used to produce new individuals. During crossover process, pairs of genomes are mated by taking a randomly selected string of bits from one and inserting it into the corresponding place in the other, and vice versa. After that, a classic mutation operator is used to mate these individuals. The classic mutation operator exchanges a random selected gene with a random value within the range of the gene's minimum value and the gene's maximum value. 40% of the best individuals are guaranteed a place in the new generation. This process is repeated for 100 generations.

The best structure of RNN provided by the proposed model is composed of 16 hidden units. We use the back-propagation algorithm to train a RNN using this structure. We note that, using this network, recognition rates and run time are greatly improved than those given by the RNN using 30 hidden units (see tables 2 and 3). We conclude that the proposed constructive evolutionary process improves the objective of defining the best structure of a RNN.

Tables 2 and 3 present a comparison of training rates, generalization rates and run time of the studied models. The Elman model using 30 hidden neurons provides the lowest recognition rate and the greater runtime of about 10 hours. GA gives best recognition rates than those given by the Elman model using 30 hidden units and it requires only 3 hours 30 minutes.

Furthermore, we note that the proposed model provides the best training rate of about 58.79% and the best generalisation rate of about 58.38%. In addition, it ameliorates the recognition rate of most of the phonemes such as /ey/ having 18% rather than 2% and /ay/ having 39% rather than 8%. We conclude, then, that the proposed multi-objective constructive model improves the objective of training of RNN. Furthermore, it should be noted that the proposed model takes 7 hours for training. This is justified by the fact that we use several objectives.

Table 2: Training rates of the Elman model using 30 hidden units, the GA, the Elman model using 16 hidden units and the RNND-MOGA model

Vowels	Samples	Elman (30 hidden units)	GA	Elman (16 hidden units)	RNND-MOGA
iy	1552	77.83	<b>85.5</b>	77.19	84.99
ih	1103	11.6	18.04	17.32	<b>41.52</b>
eh	946	28.43	25.58	28.65	<b>57.19</b>
ey	572	2.27	1.40	0.35	17.83
ae	1038	77.84	<b>86.71</b>	74.95	84.49
aa	762	71.39	72.57	66.01	<b>80.18</b>
aw	180	0.00	0.56	0.00	<b>5.00</b>
ay	600	7.67	17.33	1	<b>38.83</b>
ah	580	7.07	7.41	<b>23.79</b>	12.41
ao	665	64.36	72.03	62.86	<b>83.16</b>
oy	192	0.00	0.00	0.00	0.00
ow	549	14.39	29.87	<b>41.71</b>	28.05
uh	141	0.00	0.00	0.00	0.00
uw	198	47.98	20.71	50.51	<b>66.67</b>
ux	400	2.25	1.00	2.00	<b>11.25</b>
er	392	8.42	16.58	8.67	<b>37.24</b>
ax	871	38.35	47.19	38.12	<b>57.41</b>
ix	2103	71.85	70.28	66.14	<b>84.31</b>
axr	739	52.23	63.46	54.26	<b>64.68</b>
axh	86	37.21	34.88	<b>62.79</b>	38.37
Global rate	13966	43.63	47.68	44.29	<b>58.79</b>
Runtime		10h20mn	<b>3h30mn</b>	4h	7h

## 5. Conclusion

In this paper, we have presented a model based on multi-objective genetic algorithms in order to train and to design

recurrent neural networks. This algorithm is able to reach a wider set of possible RNN structures. We have shown that this model is able to achieve good performance in the recognition of TIMIT vowels, outperforming other studied methods.

The main results are as follows:

- The best RNN structure produced by the proposed model gives a better recognition rate at a lower runtime.
- The proposed model improves the recognition rate of the TIMIT vowels macro-classes of about 15% compared with the Elman model.

We suggest extending the constructive method to determine the optimal number of hidden layer and the number of hidden units in each one.

Table 3: Generalization rates of the Elman model using 30 hidden units, the GA, the Elman model using 16 hidden units and the RNND-MOGA model

Vowels	Samples	Elman (30 hidden units)	GA	Elman (16 hidden units)	RNND-MOGA
iy	522	72.22	83.33	72.41	<b>86.02</b>
ih	327	8.26	12.23	16.51	<b>34.86</b>
ch	279	30.83	22.94	24.01	<b>63.44</b>
ey	162	1.24	1.85	0.00	<b>20.99</b>
ae	237	73.1	<b>86.92</b>	73	81.43
aa	237	62.87	59.49	54.85	<b>74.68</b>
aw	30	0.00	0.00	0.00	0.00
ay	168	2.38	17.86	0.00	<b>41.07</b>
ah	183	8.74	9.84	<b>21.86</b>	12.02
ao	222	59.91	64.41	54.96	<b>82.88</b>
oy	51	0.00	0.00	0.00	0.00
ow	171	9.94	26.32	<b>35.09</b>	22.81
uh	59	0.00	0.00	0.00	0.00
uw	51	31.37	11.76	23.53	<b>39.22</b>
ux	104	2	2.88	2.88	<b>8.65</b>
er	141	3.55	16.31	4.96	<b>36.88</b>
ax	249	50.2	61.85	47.39	<b>67.07</b>
ix	610	67.21	69.18	60.98	<b>81.97</b>
axr	210	55.72	65.71	46.19	<b>69.05</b>
axh	28	32.14	39.29	39.29	28.57
Global rate	4042	41.28	46.57	40.68	<b>58.38</b>

## Acknowledgments

The authors are grateful to the anonymous reviewers for their valuable comments which improved the presentation and contents of this paper considerably.

## References

- [1] [http://www ldc.upenn.edu/Catalog/readme\\_files/timit.readme.html](http://www ldc.upenn.edu/Catalog/readme_files/timit.readme.html)
- [2] P.J. Angeline, G.M. Saunders, and J.B. Pollack, *An evolutionary algorithm that constructs recurrent neural networks*, IEEE Transactions on Neural Networks (1993).
- [3] N. Arous, *Hybridation des cartes de Kohonen par les algorithmes génétiques pour la classification phonémique*, Ph.D. thesis, Thèse de doctorat, ENIT, 2003.
- [4] H. Azzag, F. Picarougne, C. Guinot, and G. Venturini, *Un survol des algorithmes biomimétiques pour la classification*, Revue des nouvelles technologies de l'information (RNTI) (2004), 13–24.
- [5] D. Beasley and R. Martin, *A sequential niche technique for multimodel function operation*, Conference on evolutionary computation **1** (1993), 101–125.
- [6] P.A. Castillo, J.J. Merelo, M.G. Arenas, and G. Romero, *Comparing evolutionary hybrid systems for design and optimization of multilayer perceptron structure along training parameters*, Information Sciences **177** (2007), 2884–2905.
- [7] R. Chandra, M. Freaan and M. Zhang, *Building Subcomponents in the Cooperative Coevolution Framework for Training Recurrent Neural Networks*, School of Engineering and Computer Science, Victoria University of Wellington, Wellington, New Zealand, 2009.
- [8] M. Chetouani, B. GAS, and J.L. Zarader, *Une architecture modulaire pour l'extraction de caractéristiques en reconnaissance de phonèmes*, International conference on information processing (ICONIP'02) (2002).
- [9] H. Chihi and N. Arous, *Adapted evolutionary recurrent neural network*, JTEA (2010).
- [10] D. Dasgupta and D. R. McGregor, *Designing application-specific neural networks using the structured genetic algorithm*.
- [11] S. Dehuri and S.-B. Cho, *Multi-criterion pareto based particle swarm optimized polynomial neural network for classification : A review and state-of-the-art*, Computer Science Review **3** (2009), 19–40.
- [12] M. Delgado and M.C. Pegalajar, *A multiobjective genetic algorithm for obtaining the optimal size of a recurrent neural network for grammatical inference*, Pattern Recognition **38** (September 2005), 1444–1456.
- [13] N. Garcia and C.J. Hervas, *Multi-objective cooperative coevolution of artificial neural networks*, Neural Networks **15** (2002), 1259–1278.
- [14] D.E. Goldberg, *Algorithmes génétiques exploration optimisation et apprentissage automatique*, Kluwer Academic Publisher, 19 janvier 1996.
- [15] J.R. Koza and J.P. Rice, *Genetic generation of both the weight and architecture for a neural network*, Proceedings of the International Joint Conference on Neural Networks (1991), 397–404.
- [16] L. Kuncheva and C.J. Whitaker, *Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy*, Machine Learning **51** (2003), 181–207 51.
- [17] Y. Liu and X. Yao, *Ensemble learning via negative correlation*, Neural Networks **12** (1999), 1399–1404.
- [18] Y. Liu, X. Yao, Q. Zhao and T. Higuchi, *Evolving a cooperative population of neural networks by minimizing mutual information*, In Proceedings of the 2001 IEEE Congress on Evolutionary Computation (2001), 384–389.
- [19] K. Maneeratana, K. Boonlong and N. Chaiyaratana, *Multi-objective Optimisation by Co-operative Co-evolution*, PPSN VIII : parallel problem solving from nature, 772-781, (2004 ).
- [20] R.T. Marler and J.S. Arora, *Survey of multi-objective optimization methods for engineering*, Struct Multidisc Optim **26**, 369–395 (2004).
- [21] N. Srinivas and K. Deb, *Multi-objective function optimization using non-dominated sorting genetic algorithms*, Evolution. Comput. **2** (1994), 221–248.
- [22] E.G. Talbi, *Metaheuristiques pour l'optimisation combinatoire multi-objectif : Etat de l'art*, PM2O'1999 (1999).
- [23] L. Tcheeko, *Un réseau de neurones pour la classification et la reconnaissance de la parole*, Ecole nationale supérieure polytechnique (1994), 277–280.
- [24] R. Tlemsani, N.R. Tlemsani, N. Neggaz, and A. Benyettou, *Amélioration de l'apprentissage des réseaux neuronaux par les algorithmes évolutionnaires : application à la classification phonétique*, SETIT (2005).
- [25] S. Kazarlis V.Petridis and A. Papaikonomou, *A genetic algorithm for training recurrent networks*, Proceedings of IJCNN .93 (1993), 2706–2709.
- [26] M. Zhang and V. Ciesielski, *Using back propagation algorithm and genetic algorithms to train and refine neural networks for object detection*, Database and expert systems applications. International conference No10 1677 (1999), 626–635.
- [27] A. Zinflou, *Système interactif d'aide à la décision basé sur des algorithmes génétiques pour l'optimisation multi-objectifs*, Master's thesis, UNIVERSITÉ DU QUEBEC, 2004.

**Hanan Chihi** received computer science engineering degree from Institut Supérieur d'Informatique (ISI), Tunis, Tunisia, the MS degree Software Engineering (Intelligent Imaging Systems and Artificial Vision) from ISI Tunisia. She is currently working towards the Ph.D degree, Tunisia. Her research interests include optimization, pattern classification and evolutionary neural networks.

**Najet Arous** received computer science engineering degree from Ecole Nationale des Sciences d'Informatique, Tunis, Tunisia, the MS degree in electrical engineering (signal processing) from Ecole Nationale d'IngTenieurs de Tunis (ENIT), Tunisia, the Ph.D. degree in electrical engineering (signal processing) from ENIT. She is currently a computer science assisting master in the computer science department at FSM, Tunisia. Her research interests include scheduling optimization, speech recognition and evolutionary neural networks.

# Selective Acknowledgement Scheme to Mitigate Routing Misbehavior in Mobile Ad Hoc Network

Nimitr Suanmali<sup>1</sup>, Kamalrulnizam Abu Bakar<sup>2</sup> and Suardinata<sup>3</sup>

<sup>1</sup>Department of Computer System and Communication, Faculty of Computer Science & Information Systems, University Teknologi Malaysia, Johor Bahru, Malaysia  
Suan Dusit Rajabhat University, Bangkok, Thailand

<sup>2</sup>Department of Computer System and Communication, Faculty of Computer Science & Information Systems, University Teknologi Malaysia, Johor Bahru, Malaysia

<sup>3</sup>Department of Computer System and Communication, Faculty of Computer Science & Information Systems, University Teknologi Malaysia, Johor Bahru, Malaysia  
STMIK Indonesia Padang, Padang Indonesia

## Abstract

Mobile Ad Hoc Networks (MANETs) rely on the preliminary hypothesis that all co-operating nodes completely cooperate in an infrastructureless wireless network. Each node helps each other to perform network functions in a self-organization way. However, some nodes in a network may oppose to cooperating with others to avoid consuming their battery power and other resources. Recently, the routing misbehavior has been an interesting topic in this research field. In this paper, we propose selective acknowledgement (SACK), an end-to-end network-layer acknowledgement scheme, which can be easily attached on top of all source routing protocol. Dissimilar all previous research attempts made to tolerate routing misbehavior, this study discloses the malicious action and then recognizes compromised node or malicious nodes in the network. The malicious node will be prevented in the future routing process to improve the performance of the network throughput. Additional information of SACK scheme and preliminary evaluation are presented in this paper.

**Keywords:** Mobile Ad hoc Network, MANET, Selfish, Routing Misbehavior, Malicious, Non-cooperation, Reputation.

## 1. Introduction

The growth of wireless computer networks plays increasingly vital roles in modern society. Self organization, lacks of infrastructure, and dynamic change of nodes are the main characteristic of Mobile Ad Hoc Network (MANET). A MANET is a collection of wireless mobile nodes performing a temporary network without any established

infrastructure or centralized authority[1]. Such network does not rely on fixed architecture and pre-determined connectivity. Nodes transmit information directly to another in a range of their wireless signal. The transmission range depends not only on the power level used for the transmission, but also on the terrain, obstacles and the specific scheme used for transmitting the information[2]. Nodes in MANET are dynamically changed, which means that the topology of such networks may change rapidly and unpredictably over time. A MANET consists of devices that are autonomously self-organized into networks. A self-organizing capability makes MANET completely different from any other network. MANET is one of the most innovative and challenging areas of wireless networks. It is a key step in the evolution of wireless networks. The network is a self-organization which means that all network activity including discovering the topology and delivering messages must be executed by themselves, i.e., routing functionality will be incorporated into mobile nodes. An extensive description about the ad-hoc networks and the interrelated research topics can be found in [16][17][18][19][20]. The main challenge of MANET is the vulnerability to security attacks. The security challenge has become a primary concern to provide secure communication.

In MANETs, routing misbehavior can seriously downgrade the performance at the routing layer. Particularly, nodes may take part in the route discovery process and maintenance processes but deny to forward data packets. How do we disclose a misbehavior activity? How can we perform such a detection processes more effective, with low routing control overhead, and more accurate, with less false detection rate and false alarm?

In this paper, we concentrate on routing misbehavior that is a severe threat to Mobile Ad hoc networks. Although many research attempts have been proposed to secure routing protocols, but it is not adequately addressed for the routing

misbehavior. We have studied routing misbehavior in which a malicious node kindly forward a routing message but intentionally drops the data packets they received, unlike all previous research efforts made to tolerate routing misbehavior, our work detected the malicious activity and then identified the compromised nodes or malicious behavior nodes in the network. We propose a scheme called Selective Acknowledgement (SACK) to detect misbehaving nodes, which can be implemented on network layer of any source routing protocol. The source node validates that the packet forwarded is received completely by neighbor nodes on the source route by a specific type of acknowledgement packets, called SACK packets. SACK packets have a related operation as the SACK packets on the TCP layer, but the SACK packets in TCP are used for reliable communication and flow-control. A neighbor node noticed the arriving of data packet by reply back to the source node with a SACK packet. The neighbor node will suspect to be a malicious node, if the source node does not accept a SACK packet interrelated to a specific data packet that was replied back. The malicious node will be avoided in the future routing process, so the throughput performance of overall network will be enhanced.

The rest of the paper is organized as follows. In section 2, various approaches mitigated routing misbehavior are summarized. In section 3, the details of routing misbehavior are given. The information of the SACK system and interrelated discussion are presented in section 4. We conclude the work in section 5.

## 2. Related Work

The fundamental technique for the most of an intrusion detection system that found in this section is Watchdog. Sergio Marti et al. [3] proposed an intrusion detection technique called Watchdog and constructed on a Dynamic Source Routing (DSR) protocol [4]. The authors proposed two techniques to improve a throughput ratio in the situation that compromised nodes willing to forward routing packets but reject to forward data packets. The first technique is Watchdog, which recognizes misbehaving nodes while the second technique, the Pathrater, which is similar to an intrusion detection system that helps routing protocols to eliminate these misbehaving nodes from the active route. When a node forwards a packet, the node's Watchdog verifies that the next node in the path also forwards the packet by listening continuously in a promiscuous mode to the neighbor node's transmissions. If the neighbor node does not forward the packet, it was decided as a misbehaving node. The Watchdog increases the misbehaving counter every time a node misses to forward the packet. If the misbehaving counter reaches a particular threshold, it recognized that the node is misbehaving node, then this node is prevented using the Pathrater. The drawbacks of watchdog are that it might not detect a misbehaving node in the presence of receiver collision, ambiguous collision, false misbehavior reporting, limited transmission power, partial dropping and collusion.

Hasswa et al. proposed an intrusion detection and response system for mobile ad hoc network called Routeguard[5]. This technique is a combination of two techniques, Watchdog and Pathrater, proposed by Marti et al. [3], to categorize each neighbor node into 4 categories: fresh, member, unstable, suspect. The Watchdog classified each node based on the ratings acquired from its behavior. Moreover, each category has a various trust level as trusted and untrusted. The trusted member lets the node to take part in the network. On the other hand, the untrusted member corresponds to a node that is absolutely untrusted and not allowed from using the network resources. Routeguard is a similar process to the Pathrater which performs by every node in the network and takes over a rating for all neighbors nodes in it wireless signal range. A Routeguard enhances Pathrater performance by distributing ratings to all participant nodes and measuring a path metric. Therefore, it demonstrates a more detailed and standard classification system that rates every node in the network.

Nasser and Chen [6] proposed an improved intrusion detection system for detecting malicious nodes in MANETs named ExWatchdog based on the Watchdog technique proposed by Marti et al. [3]. The researchers focus on the false misbehaving of the Watchdog technique, where a malicious node which is the actual intruder incorrectly reports another node as misbehaving. In ExWatchdog, a table is looked after by every node to record the quantity of packets the node forwards, receives or sends respectively. The source node will discover another path, when it obtains information of the misbehaving node, to enquire the destination node related to the number of received packets. The actual malicious node reports another node as misbehaving will be suspected, If the source node found that it is the same packets that it has sent. Otherwise, nodes being broadcasted information about a malicious node do false detection. However, there is still a drawback, it is impracticable to approve and confirm the number of packets with the destination node if the actual misbehaving node exists in all active paths from source to destination.

Parker et al in [7] proposed an improvement to an original the Watchdog technique which not only suitable for DSR protocol but also suitable to all routing protocols used in MANETs. In differentiating to the Watchdog, the nodes overhear all the other nodes in their neighborhoods and not only the next forward node on the path. The authors also proposed two response mechanisms, passive response and active response. The passive response mode performs freely, and eventually the intrusive node will be prevented from using all network resources. The second mechanism is the active response mode where the decision making is done by a cluster head which starting a voting procedure. If the majority decides that the suspected node is the intruder, and the intruder node will be prevented from using network resources. After all, an alert will be broadcasted throughout the network.

Animesh and Amitabh [8] proposed a method to improve performance of Watchdog technique by focus to the problem of collusion attack, which means a malicious



behavior from a collaboration of many nodes. The researchers assumed that the few nodes established the network are trusted nodes and the others that would join the network later are ordinary nodes. The Watchdog nodes are chosen from the trusted nodes to prevent the problem of inaccurate reporting. The two thresholds are maintained in every Watchdog, for all its neighbors that are not trusted nodes called `SUSPECT_THRESHOLD` and `ACCEPTANCE_THRESHOLD` respectively. The `SUSPECT_THRESHOLD` used for measure a node's misbehaving, and the `ACCEPTANCE_THRESHOLD` used for measure a node's good behavior. The Watchdog node will distinguish the neighboring nodes as a malicious or trusted node based on these thresholds.

Sonja Buchegger and Jean Boudec proposed another reputation mechanism called "CONFIDANT", which means for Cooperation Of Nodes: Fairness In Dynamic Ad-hoc NeTworks [9]. The CONFIDANT has four main components, a reputation system, a monitor, a trust manager and a path manager. Each node implemented these components to monitor its neighbors by hearing to the transmission of the next node or by watching routing protocol behavior. A trust manager will be broadcasted alarm messages to all nodes in the network by when a misbehaving node is detected. The reputation system is used to measure nodes' reputation in a network. A path manager is responsible to rank a path according to a security metric. Furthermore, a path manager will punish a selfish node by denying it all services. The simulation result of the performance of protocol in a scenario when a third of nodes behave selfishly showed that the throughput given by CONFIDANT is quite similar to the throughput of a usual network condition without selfish nodes. Since the CONFIDENT protocol relying on the Watchdog mechanism, it receives many of the Watchdog problems.

Michiardi et al. [10] proposed the other protocol that also uses a Watchdog mechanism called CORE, a COLlaborative REputation mechanism. However, it is complemented by a complex reputation mechanism that differentiates from subjective reputation. This protocol includes of three main components, functional reputation, observations and indirect reputation that use positive reports by others. These three components are weighted for a combined reputation value that is used to take decisions about cooperation or gradual isolation of a node. Each node takes part in the IDS has reputation table and Watchdog mechanism. The reputation table keeps track of reputation values of other nodes in the network. Since a misbehaving node can accuse a good node, only positive rating factors can be distributed in CORE. This protocol also depends on the use of the Watchdog mechanism that inherited its disadvantages and problems.

### 3. Problem of routing misbehavior

In this section, we give elaborate more detail the problem caused by routing misbehavior. The design of routing protocols used in Wireless Ad Hoc networks such as DSR, AODV [21] and DSDV [22] are highly vulnerable to routing misbehavior due to faulty or compromised nodes. A selfish

node operates normally in the Route Discovery and the Route Maintenance phases of the DSR protocol, but it does not intend to perform the packet forwarding function for data packets unrelated to it. The source node may be confused since such misbehaving nodes participate in the Route Discovery phase, they may be included in the routes chosen to forward the data packets from the source, but the misbehaving nodes refuse to forward the data packets from the source. In TCP, the source node may either choose an alternate route from its route cache or initiate a new Route Discovery process. The alternate route may again contain misbehaving nodes and the data transmission may fail again. However, the new Route Discovery phase will return a similar set of the same routes which including the misbehaving nodes. Eventually, the source node may conclude that routes are unavailable to deliver the data packets. This cause the network fails to provide reliable communication for the source node even though such routes are available. In UDP, the source simply sends out data packets to the next-hop node, which forwards them on. The existence of a misbehaving node on the route will cut off the data traffic flow. The source has no knowledge of this at all. Node's misbehavior can be classified [11] into 3 categories as follow:

- Malfunctioning nodes: This behavior happen when nodes suffer from hardware failures or software errors.
- Selfish nodes: In this group, nodes refuse to forward or drop data packet and can be defined into three types [12] (i.e. SN1, SN2 and SN3). SN1 nodes take participation in the route discovery and route maintenance phases but refuse to forward data packets to save its resources. SN2 nodes neither participate in the route discovery phase nor in data-forwarding phase. Instead they use their resource only for transmissions of their own packets. SN3 nodes behave properly if its energy level lies between full energy-level  $E$  and certain threshold  $T1$ . They behave like node of type SN2 if an energy level lies between threshold  $T1$  and another threshold  $T2$  and if an energy level falls below  $T2$ , they behave like node of type SN1.
- Malicious: These nodes use their resource and aims to weaken other nodes or whole network by trying to participate in all established routes thereby forcing other nodes to use a malicious route which is under their control. After being selected in the requested route, they cause serious attacks either by dropping all received packets as in case of Black Hole attack [13], or selectively dropping packets in case of Gray Hole attack [14]. For convenience such malicious nodes are referred as MN nodes. SN2 type nodes do not pose significant threat therefore can simply be ignored by the routing protocol. On the other hand SN1, SN3 and MN nodes are much more dangerous to routing protocols. These nodes interrupt the data flow by either by dropping or refusing to forward the data packets thus forcing routing protocol to restart the route-discovery or to select an alternative route if it is available which in turn may again include some malicious nodes, therefore the new route will also fail. This process form a loop which enforce source to conclude that data cannot be further transferred.

### 4. Proposed Scheme

In this section, we elaborate more details of our solution to address the routing misbehavior. Our solution has two main processes. We detect the malicious activity in the first effort and then identify the malicious or compromised nodes in the network. Our scheme can be integrated on top of any source

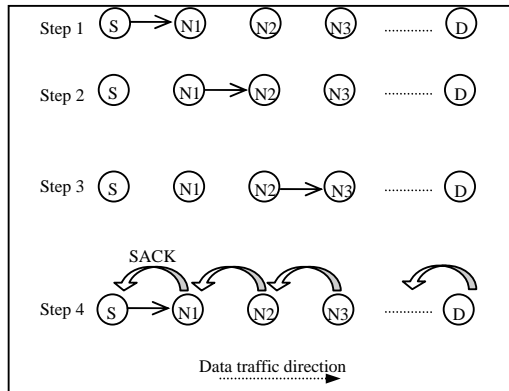


Fig.1 The SACK scheme

routing protocol such as DSR and AODV.

The Selective Acknowledgement (SACK) is a network layer acknowledgment-based scheme that considered as an enhancement system of an end-to end acknowledgment scheme (ACK). It aims to improve the performance of ACK scheme. It reduces the routing overhead of ACK while maintaining better performance and increases its detection efficiency by applying node detection instead of link detection. It is built on top of DSR routing protocol because it needs a source route protocol.

Figure 1 illustrates the operational detail of SACK scheme. Assume that the process of Routing Discovery has already established a source route from a source node S through N1,N2,N3 to a destination node D. In the SACK scheme, instead of sending back an acknowledge packets all the time when a data packet is received, a node wait until a certain amount of data packets of the same source node arrive, then it send back one SACK packet acknowledge for multiple data packets that have been received. When a source node S send out any packet to a destination node D through its neighbor nodes N1, N2, N3, all these node add a packet ID in to a list of receive data packet as shown in figure 2. In stead of sending back an acknowledgement every time when a data packet is received, a node waits until a certain number of data packets of the same source node arrive. Then the node sends back one SACK packet acknowledging multiple data packets that have been received. If the source node receives a SACK packet from the destination that means there are no misbehaving nodes along the path.

Nid	Mcount	ID_List
Neighbor ID	Misbehavior counter	List of data packet IDs Awaiting SACK

Fig.2. Data Structure of Misbehavior Detection List

Figure 2 illustrates the data structure of the Misbehavior Detection List. To detect misbehavior nodes, the sender of a data packet maintains a list of data packet IDs that receive a SACK packet from neighbor nodes. Each node maintains its unique list for each neighbor node. When a node, N1, sends or forwards a data packet to its neighbor node, N2, it adds the packet ID to its Misbehavior Detection List corresponding to N2. When it receives a SACK packet, it updates the node N2, and then removes the corresponding packet ID from the list. The node N2 will be suspected if its data packet ID stays on the list longer than a certain period of time, *time\_out*. The misbehavior counter, *Mcount*, is increased by one when misbehavior is suspected. When *Mcount* reaches certain of threshold level, *threshold*, a node declares its neighbor node, N2, as a misbehaving node and broadcasts an RERR message to report a source node and all its neighbor nodes about this misbehavior node. All nodes in the same network update its misbehaving list and avoid this misbehaving node in the next routing process.

## 5. Conclusion

This paper presents a frame work in detecting misbehaving nodes and isolating such nodes from routing process in MANETs. This scheme can be combined on top of any source routing protocol such as DSR. A comprehensive analysis of routing misbehavior was made to develop a security module that would meet the network security goal. Currently we are working on its simulation in ns-2 simulator [15] to show the results and effectiveness of our solution on DSR routing protocol. Similar approaches can also be integrated to these source routing algorithms to address other attacks like black hole and gray hole attacks in MANETs.

## Acknowledgement

We would like to thank Suan Dusit Rajabhat University and Universiti Teknologi Malaysia for supporting us.

## References

- [1] S. Alampalayam, A. Kumar, and S. Srinivasan, "Mobile ad hoc network security-a taxonomy," Advanced Communication Technology, ICACT 2005., pp. 839-844, 2005.
- [2] F. Anjum and P. Mouchtaris, Security for Wireless Ad-hoc Networks: John Wiley & Sons, 2006.
- [3] S. Marti, T. J. Giuli, K. Lai, and M. Baker, "Mitigating Routing Misbehavior in Mobile Ad-hoc Networks," Proceedings of the 6<sup>th</sup> Annual International Conference on Mobile Computing and Networking (MobiCom'00), PP. 255-265, August 2000.
- [4] D. B. Johnson, D. A. Maltz, and Y.-C. Hu, "The dynamic source routing protocol for mobile ad hoc networks (dsr)," Published Online, IETF MANET Working Group, INTERNET-DRAFT, July 2004, expiration: January 2005. [Online]. Available: <http://www.ietf.org/internetdrafts/draft-ietf-manet-dsr-10.txt>
- [5] Hasswa, A.; Zulkernine, M.; Hassanein, H., "Routeguard: an intrusion detection and response system for mobile ad-hoc networks," Wireless And Mobile Computing, Networking And Communications, 2005. (WiMob'2005), IEEE International Conference on , vol.3, no., pp. 336- 343 Vol. 3, 22-24 Aug. 2005

- [6] Nasser, N.; Chen, Y., "Enhanced Intrusion Detection System for Discovering Malicious Nodes in Mobile Ad-hoc Networks," Communications, 2007. ICC '07. IEEE
- [7] Parker, J.; Undercoffer, J.; Pinkston, J.; Joshi, A., "On intrusion detection and response for mobile ad-hoc networks," Performance, Computing, and Communications, 2004 IEEE International Conference on , vol., no., pp. 747-752, 2004
- [8] Patcha, A.; Mishra, A., "Collaborative security architecture for black hole attack prevention in mobile ad-hoc networks," Radio and Wireless Conference, 2003. RAWCON '03. Proceedings, vol., no., pp. 75-78, 10-13 Aug. 2003
- [9] S. Buchegger and J.-Y. L. Boudec, "Performance analysis of the confidant protocol (cooperation of nodes: Fairness in dynamic ad-hoc networks)," in MOBIHOC'02, 2002.
- [10] P. Michiardi and R. Molva, "CORE: a collaborative reputation mechanism to enforce node cooperation in mobile ad hoc networks," in CMS'2002, Communication and Multimedia Security 2002 Conference, September 26-27, 2002.
- [11] A. S. A. Ukey and M. Chawla, "Detection of Packet Dropping Attack Using Improved Acknowledgement Based Scheme in MANET," IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 4, No 1, July 2010, pp. 12-17
- [12] Abdelaziz Babakhouya, Yacine Challal, and Abdelmadjid Bouabdallah, "A Simulation Analysis of Routing Misbehaviour in Mobile Ad Hoc Networks," in Proc. of the Second International Conference on Next Generation Mobile Applications, Services, and Technologies, September 2008, pp. 592-597.
- [13] Mohammad Al-Shurman, Seong-Moo Yoo, and Seungjin Park, "Black hole attack in mobile Ad Hoc networks," in Proc. of the 42nd annual Southeast regional conference, ACM Southeast Regional Conference, April 2004, pp. 96-97.
- [14] J. Sen, M.G. Chandra, S.G. Hariharan, H. Reddy, and P. Balamuralidhar, "A mechanism for detection of gray hole attack in mobile Ad Hoc networks," in Proc. of the 6<sup>th</sup> International Conference on Information, Communications & Signal Processing, December 2007, pp. 1-5.
- [15] The Vint Project, "The ns-2 network simulator," <http://www.isi.edu/nanam/ns>
- [16] C. E. Perkins, Ad-hoc Networking. Addison Wesley Professional, December 2000.
- [17] M. Ilyas, ed., The Handbook of Ad-hoc Wireless Networks. CRC Press, December 2002.
- [18] R. Hekmat, Ad-hoc Networks: Fundamental Properties and Network Topologies, Springer, 2006.
- [19] M. Barbeau, E. Kranakis, Principles of Ad-hoc Networking. Wiley, 2007.
- [20] S. K. Sarkar, T. G. Basavaraju, C. Puttamadappa, Ad Hoc Mobile Wireless Networks. Auerbach Publications, 2008.
- [21] Charles E. Perkins and Elizabeth M. Royer, "Ad hoc on demand distance vector (AODV) routing (Internet-Draft)", Aug- 1998.
- [22] C. Perkins, P. Bhagwat, "Highly dynamic destination-sequenced distance-vector routing (DSDV) for mobile computers", ACM SIGCOMM Computer Communication Review 1994; 24(4):234-244.



**Nimitr Suanmali**, is a Ph.D. student in the Dept. of Computer System and Communication, Faculty of Computer Science and Information System, University Teknologi Malaysia, Johor Bahru Malaysia. He received his B.Sc. degree in computer science from Suan Dusit Rajabhat University, Thailand in 1998, M.Sc. degree in Information Technology from King Mongkut's University of Technology Thonburi, Thailand in 2003. Since 2003, he has been working as lecturer at Suan Dusit Rajabhat University, Bangkok Thailand. His research interests include Network Security, Intrusion Detection and Intrusion Prevention, Wireless Ad-Hoc Networks, and Distributed Systems.



**Kamalrulnizam bin Abu Bakar** obtained his Ph.D degree in Computer Science (Network Security) from Aston University (Birmingham, UK) in 2004, B.S 1996 in Computer Science, Universiti Teknologi Malaysia and M.S. in Computer Communication & Networks, Leeds Metropolitan University, UK. in 1998. Currently, he is an Associate Professor in

Computer Science at Universiti Teknologi Malaysia (Malaysia) and member of the "Pervasive Computing" research group. He involves in several research projects and is the referee for many scientific journals and conferences. His specialization includes mobile and wireless computing, information security and grid computing.



**Suardinata**, he is received the Diploma III 1999 in Information Management at AMIK Riau, Indonesia, Bachelor Degree in Information Engineering from STMIK Riau, Indonesia, and Master Degree in Information Technology from Universitas Putra Indonesia, Padang, Indonesia. Currently he is a Ph.D. student in the Dept. of Computer System and Communication, Faculty of Computer Science and Information System, University Teknologi Malaysia, Johor Bahru Malaysia. He has been working as Lecturer at STMIK Indonesia Padang from 2005 in the Dept. of Computer Science and Information Systems, STMIK Indonesia Padang. His research interests include Multimedia and Voice over IP network, Network Security, Traffic Engineering and Quality of Service issues in IP networks, Wireless Ad-Hoc Networks, and Distributed Systems.

# An Analytical Framework for Multi-Document Summarization

Jayabharathy<sup>1</sup>, Kanmani<sup>2</sup> and Buvana<sup>3</sup>

<sup>1</sup> Pondicherry Engineering College  
Pondicherry-605014

<sup>2</sup> Pondicherry Engineering College  
Pondicherry-605014

<sup>3</sup> Pondicherry Engineering College  
Pondicherry-605014

## Abstract

Growth of information in the web leads to drastic increase in field of information retrieval. Information retrieval is the process of searching and extracting the required information from the web. The main purpose of the automated information retrieval system is to reduce the overload of document retrieval. Today's retrieval system presents vast information, which suffers from redundancy and irrelevance. There arises a need to provide high quality summary in order to allow the user to quickly locate the desired and concise information ase number of documents available on user's desktops and internet increases. This paper provides the complete survey, which gives a comparative study about the existing multi-Document summarization techniques. This study gives an overall view about the current research issues, recent methods for summarization, data set and metrics suitable for summarization. This frame work also investigates about the performance competence of the existing techniques.

**Keywords:** *Multi-Document Summarization, Generic Summary, Query Based Summary.*

## 1. Introduction

Document Summarization is an automated technique, which reduces the size of the documents and gives the outline and concise information about the given document. That is the summarization process extracts the most important content from the document. In general, the summaries are created in two ways. They are generic summary and query based summary. The generic summary refines overall content of the input document given by the user whereas the query based one retrieves the information that more relevant to the user query. Document summarizations are of two types, they are single document summarization and Multi-document summarization. The

summary that is extracted and created from a single document is known as Single Document Summarization, whereas Multi-document Summarization is an automatic procedure for the extraction of information from multiple sources.

The purpose of a brief summary is to shorten the information search and to minimize the time by spotting the most relevant source documents. Widespread multi-document summary itself hold the required information, hence limiting the need for accessing original files to some cases when refinement is required. Automated summaries give the extracted information from multiple sources algorithmically.

The remainder of this paper is organized as follows: Section 2 provides the Classification of various summarization techniques and describes about the related works in field of generic based and query based summary generation. The general framework for extracting summary from documents sources and steps involved in this process of summary extraction are described in section 3. Section 4 gives the detailed discussion about the framework for analyzing existing summarization techniques. The paper is concluded with a brief discussion in section 5.

## 2. Classification of Summarization Techniques

This chapter gives an overview about various summarization techniques. The summarization techniques are classified into two major groups Generic and Query based summary creation. The generic summary refines overall content of the input document given by the user whereas the query based one retrieves the information that

is more relevant to the user query. The classification of multi-document summarization is shown in the figure 1. The brief description about each technique is stated below.

## 2.1 Generic Summary Extraction Techniques

The RANDOM based technique [9] is the simplest technique, which randomly selects lines from the input source documents. Depending upon the compression rate i.e. the size of the summary, the randomly selected lines will be included to the summary. In this technique, a random value between 0 and 1 is assigned to each sentence of the document. A threshold value for length of the sentence is provided in general. The score of 0 to 1 is assigned to all sentences that do not meet assigned length cut-off. Finally, required sentences are chosen according to assigned highest score for desired summary.

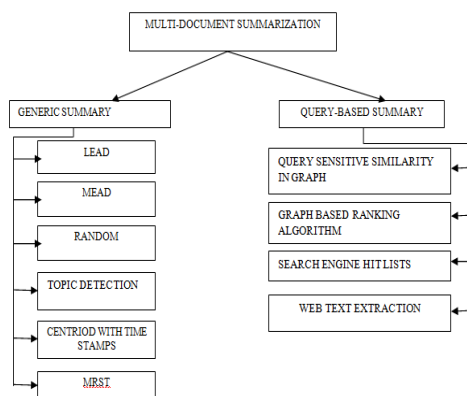


Fig.1 Classification of summarization techniques

LEAD based technique is one where first or first and last sentence of the paragraph are chosen depending upon the compression rate (CR) and it is suitable for news articles. It can be reasonable that n% sentences are chosen from beginning of the text e.g. selecting the first sentence in all the document, then the second sentence of each, etc. until the desired summary is constructed. This method is called LEAD [9] based method for summarization. In this technique a score of  $1/n$  to each sentence is assigned, where n is the sentence number in the corresponding document file. This means that the first sentence in each document will have the same scores; the second sentence in each document will have the same scores, and so on. The length value is also provided as a threshold. The sentences with less length than the specified threshold value are thrown out.

MEAD is a commonly used technique which can perform many different summarization tasks. It can also summarize individual documents or clusters of related documents.

MEAD is the combination of two baseline summarizers: lead-based and random based. Lead-based summaries are produced by selecting the first sentence of each document, then the second sentence of each, etc. until the desired summary size is met. A random summary consists of enough randomly selected sentences (from the cluster) to produce a summary of the desired size. MEAD is a centroid-based extractive summarizer that scores sentences based on sentence-level and inter-sentence features that indicate the quality of the sentence as a summary sentence. It then chooses the top-ranked sentences for inclusion in the output summary. MEAD extractive summaries score the sentences according to certain sentence features – Centriod [9], Position [9], and Length [9].

Dragomir R. Radev [1] et al proposed a multi-document text summarizer, called MEAD. The proposed system creates the summary based on cluster centroids. Centroid is the set of words that are most important to the cluster. In addition to the Centroid, position and first sentence overlap values are involved in the score calculation. Two new techniques namely cluster based relative utility and cross sentence information subsumption were applied to the evaluation of both single and multiple document summaries. Cluster base relative utility refers to the degree of relevance of a particular sentence to the general topic of the cluster. Summarization evaluation methods used could be divided into two categories: intrinsic and extrinsic. Intrinsic evaluation method measures the quality of multi-document summaries in a direct manner. Extrinsic evaluation methods measure how successfully the summaries help in performing a particular task. The extrinsic evaluation in terms called task-based evaluation. The new utility-based technique called CBSU was used for the evaluation of MEAD and of summarizers in general. It was found that MEAD produces summaries that are similar in quality to the ones produced by humans. MEAD's performance was compared to an alternative method, multi-document lead and showed how MEAD's sentence scoring weights can be modified to produce summaries significantly better than the alternatives.

Afnan Ullah Khan [3] et al proposed a new technique for information summarization, which is the combination of the rhetorical structure theory and MEAD summarizer. In general MEAD summarizer is totally based on mathematical calculation and lack a knowledge base. Rhetorical structure theory is used to overcome this weakness. The new summarizer system is evaluated against the original MEAD summarizer system. The proposed summarizer tool was exploited mainly in two areas of information that are Financial Articles and PubMed abstracts. The experimental results show that MEAD produces successful summaries 75% time for both



short and long documents whereas MRST produces successful summaries for short documents 70% of the time and long documents summaries 65% of the time, as the size of the document increases the performance of MRST deteriorates.

The two-stage sentence selection approach proposed by Zhang Shu [4] et al is based on deleting sentences in a candidate sentence set to generate summary. The two stages are (1) acquisition of a candidate sentence set and (2) the optimum selection of sentence. The candidate sentence set is obtained by redundancy-based sentence selection approach at the first stage where as in the second stage, optimum selection of sentences technique is used to delete sentences in the candidate sentence set according to its contribution to the whole set until desired summary length is met. With a test corpus, the ROUGE value obtained for the proposed approach proves its validity, compared to the traditional method of sentence selection. The influence of the chosen token in the two-stage sentence selection approach on the quality of the generated summaries is analysed. It differs from the traditional method of adding sentences to create summary by deleting the sentences in a set of candidate sentences to create the summary. With the test corpus used in DUC 2004, and compared to the redundancy based sentence selection, the experiments show that the two-stage sentence selection approach increases the ROUGE value of the summaries, which proves the validity of the proposed approach.

Dingding Wang [7] et al proposed a summarization system which is mainly based on sentence-level semantic analysis and non-negative matrix factorization. The sentence-sentence similarity is calculated by using the semantic analysis and the similarity matrix is constructed. Then the symmetric matrix factorization process is used to group the similar documents into clusters. The experimental result on DUC2005 and DUC2006 datasets achieves the higher performance.

Ben Hachey [8] proposed a generic relation extraction based summarization system. A GRE system builds the systems for relation identification and characterization which can be transferred across domains and tasks without any modification in model parameters. Relation identification is the extraction of relation forming entity mention pairs whereas relation characterization is the assignment of types of relation mentions. An experimental result shows that the proposed system's performance is slightly superior when compared to the existing system.

Md. Mohsin Ali [9] et al proposed two techniques for both single and multi document text summarization. The first technique is adding a new feature called SimWithFirst (Similarity with First Sentence) with MEAD (Combination of Centroid, Position, and Length Features)

called CPSL and second is the combination of LEAD and CPSL called LESM. In general LEAD is the summarization technique in which first or first and last sentence of the paragraph are chosen depending upon the compression rate (CR). The results of proposed techniques are compared with conventional methods called MEAD with respect to some evaluation techniques. The results demonstrate that CPSL shows better performance for short summarization than MEAD and for remaining cases it is almost similar to MEAD and LESM also shows better performance for short summarization than MEAD but for remaining cases it does not show better performance than MEAD.

Shu Gong [11] et al proposed a Subtopic-based Multi-documents Summarization (SubTMS) method. This method adopts probabilistic topic model to find out the subtopic information inside each and every sentence and uses a hierarchical subtopic structure to explain both the whole documents collection and all sentences inside it. here the sentences represented as subtopic vectors, it assess the semantic distances of sentences from the documents collection's main subtopics and selects sentences which have short distance as the final summary. They have found that, training a topic's documents collection with some other topics' documents collections as background knowledge, this approach achieves fairly better ROUGH scores compared to other peer systems in the experimental results on DUC2007 dataset.

A.Kogilavani [12] et al proposed an approach to cluster multiple documents by using document clustering approach and to produce cluster wise summary based on feature profile oriented sentence extraction strategy. Most similar documents are grouped into same cluster using document clustering algorithm. Feature profile is generated which mainly includes the word weight, sentence position, sentence length, and sentence centrality, proper nouns in the sentence and numerical data in the sentence. Based on this feature profile sentence score is calculated for each and every sentence in the cluster of similar documents. According to different compression ratio sentences are extracted from each cluster and ranked. Then the sentences are extracted and included in the summary. Extracted sentences are arranged in chronological order as in input documents and with the help of this, cluster wise summary will be generated. An experimental result shows that the proposed clustering algorithm is efficient and feature profile is used to extract most important sentences from multiple documents. The summary generated using the proposed method is compared with human summary created manually and its performance has been evaluated and the result shows that the machine generated summary coincides with the human intuition for the selected dataset of documents.

## 2.2 Query Based Summary Techniques

Dragomir R. Radev [2] et al designed a prototype system called SNS, which is pronounced as “essence”. This mainly integrates natural language processing and information retrieval techniques in order to perform automatic customized summarization of search engine results. The proposed system actually retrieves documents related to an unrestricted user query and summarizes a subset of them as selected by the user Task-based extrinsic evaluation showed that the system is of reasonably high quality.

Furu [5] et al proposed a graph based query oriented summarization based on query sensitive similarity measure. For the evaluation of sentence-sentence edges the similarity measure incorporates the query influence technique. Graph modeling and graph based ranking algorithm is used for finding the similarity between the sentences. Then sentences which are more similar to the user query will be retrieved. The experimental results on DUC 2005 shows that it improves ROUGH score.

Xiao [6] et al designed and proposed a system to automate the multi-document summarization. The proposed system retrieves the documents related to the query given by the user. The sentence score is calculated based on relevant value and in-formativeness value. These values are realized by word sentence overlap and semantic graph techniques. Then the sentences with the highest score are included to the summary. The investigational result shows that the proposed system achieves better quality.

Lei Huang [10] et al considers document summarization as a multi-objective optimization problem involving four objective functions, namely information coverage, significance, redundancy and text coherence. These functions measure the possible summaries based on the identified core terms and main topics (i.e. a cluster of semantically or statistically related core terms). The datasets namely DUC 2005 and 2006 have been chosen for query-oriented summarization tasks to test the proposed model. The experimental results indicate that the multi-objective optimization based framework for document summarization is truly a promising research direction. It is valuable to note that a real optimization based summarization method is different from the existing non-optimization based methods in two noteworthy aspects. First, it ranks summaries instead of ranking individual sentences. Second, though ignored in the previous literature, the approach to rank summaries should not directly rely on the approach to rank sentences. Otherwise, the optimization solutions will degenerate to the traditional non-optimization based (e.g. MMR like) methods.

## 3. GENERAL PROCEDURE FOR DOCUMENT SUMMARIZATION

Usually document sources are of unstructured format, transforming these unstructured documents to structured format requires some pre-processing steps. Fig.1 presents the sequence of steps involved in document Summarization. Some commonly used pre-processing steps are

**Sentence Decomposition:** The given input document is decomposed into sentences.

**Stop words removal:** Stop words are typical frequently occurring words that have little or no discriminating power, such as \a", \about", \all", etc., or other domain-dependent words. Stop words are often removed.

**Stemming:** Removes the affixes in the words and produces the root word known as the stem [13]. Typically, the stemming process is performed so that the words are transformed into their root form. For example connected, connecting and connection would be transformed into ‘connect’. Most widely used stemming algorithms are Porter [17], Paice stemmer [16], Lovins [15], S-removal [14]

**Feature Vector Construction:** Feature vector is constructed based on term frequency (TF-DF) and inverse document frequency (TF-IDF).

After applying the preprocessing techniques, the processed documents are clustered using a clustering algorithm in order to group the similar documents. Cluster analysis or clustering is the assignments of a set of observations into subsets (called clusters) so that observations of same cluster are similar in some sense. Some of the famous types of clustering are described below.

Hierarchical algorithms find consecutive clusters using previous clusters. They are of two types namely agglomerative ("bottom-up") and divisive ("top-down"). The first type begins with each element as a individual cluster and merge them into larger clusters. Divisive algorithms start with the whole document set and divide it into smaller clusters.

Partitional algorithms typically resolve all clusters at once, but can be used as divisive algorithms in the hierarchical clustering.

After the clustering process the summary is created for the clustered documents. We have discussed the variety of summary creation techniques in the previous section.

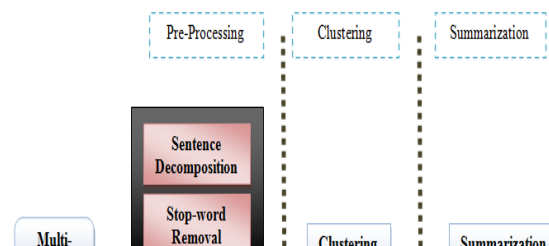


Fig. 2 General Procedure for Document Summarization

#### 4. A FRAMEWORK FOR ANALYZING DOCUMENT SUMMARIZATION

This study mainly highlights the recent research work in the field of multi-document summarization. This paper primarily focuses about the proposed framework for comparing various multi-document summarization techniques. The comparative study is based on the survey, which is made by analyzing the existing algorithms, considering the characteristic factors like Document summarization technique, Data set used for experiments, Performance metrics and detail about the performance of the proposed technique. Column one in this table presents the title of the related papers. The Framework, algorithm and techniques which are discussed in the existing papers are stated in column two. The third column gives the details of the data set which are considered for conducting the experiments. Metrics considered by the authors for performance evaluation are given in column four. The concise details about the performance of the proposed techniques are listed in column five.

Table 1: Comparison Of Existing Summarization Techniques

Paper Title	Algorithm/Technique	DataSet	Evaluation tool/Metric	Performance
Centriod based summarization of multiple documents 2003	Mead extraction algorithm	News articles	Utility based evaluation, User studies and System evaluation	Utility is very high

Automatic summarization of search engine hit lists	Centriod, Position and First sentence overlap	Global E-commerce Framework	Time, Reliability	Better Speedup in reading time, Better Reliability
MRST: a new technique FOR Information Summarization, 2005	MRST	Financial Articles and PubMed abstracts	Coherence, Correctness, Compression, Overall	Existing technique such as Mead comes out more successful when compared to MRST
Two stage sentence selection approach for multi document summarization-2008	Redundancy based sentence selection	DUC2004	ROUGH	Increased ROUGH score, Proves Validity
A Query-Sensitive Graph-Based Sentence Ranking Algorithm for Query-Oriented	Graph Modeling, Graph-Based Ranking Algorithm	DUC2005	ROUGH-1, ROUGH-2, ROUGH-SU4	4.9% improvement in ROUGH-2
Multi-Document Summarization via Sentence – Level Semantic Analysis and	Semantic similarity matrix construction, Symmetric Non-negative Matrix Factorization and kernel K-means clustering	DUC2005, DUC2006	ROUGH-1, ROUGH-2, ROUGH-N(n-gram recall) ROUGH-L, ROUGH-W(ROUGH-SU(ski	Better ROUGH Scores

Symmetric Matrix Facorrization-2008			p- bigram plus unigra m)	
Multi-Document Summarization Using generic Relation Extraction	GRE	DUC2001	ROUG H-SU4	Maximum ROUGH Score of 0.396 is obtained
Multi-document Text Summarization : SimWithFirst Based Features and Sentence Co-selection Based Evaluation-2009	CPSL, LESM	DUC2004	Precisi on, Recall, Kappa Coeffi cient, Cross Judge Utility agreem ent	Better Performance for short Summaries
Modeling Document Summarization as Multi-objective Optimizati on-2010	Summary Ranking	DUC200 5 &2006	cover age, signif icanc e, redun danc y and text coher ence	Produces Optimized summary
Subtopic-based Multi-	Subtopic vector construction and semantic distance calculation	DUC2007	ROU GH	Better ROUGH Scores

documents				
Summariz ation-2010				

This frame work precisely states the details about the algorithms, data sets, metrics and performance. From the analysis it is understood that majority of the researchers concentrate on multi- document summarization. During the earlier stage most of the researchers concentrate on single document summarization. Multi document summarization came in picture from 2000 onwards. At earlier days the position of the sentences are considered to be important and have included to the summary like including title sentences, sentences at the mid of the paragraph etc. This method is suitable for documents which are related to news documents. But recent researchers not only concentrate on position they also give importance to the semantics of the sentences and their significance are identified and then it is added to the summary. Most of the researchers compare their proposed work with human generated summaries and justifies their work. From the survey it is concluded that MEAD is the most popular tool for Document summarization. Precision, Recall, Kappa Coefficient, F-Measure, etc are metrics used for evaluating the generated summary.

Rough score gives the measurement of sentence relevance. The Rough score are used by majority of researchers in association with DUC dataset for evaluating the quality of generated summary. In addition to that some of the document summarization uses the news articles and financial articles as the dataset. Some summarization technique ranks the sentences according the factor like position, semantic, number of nouns, length etc are included to the summary. Compression rate is considered to be one more factor for summary generation. Generic summary generation draws the attention of many researchers.

## 5. CONCLUSION

In this paper a frame work for analyzing existing document summarization algorithms was proposed. This framework gives the brief overview of recent research work on various algorithms in document summarization technology. Some inferences from the analytical frame work were also discussed. This gives the clear idea about the ongoing field of research in summarization. Document Summarization still has a scope in summarization in Distributed Environment and in Dynamic Multi-Document Summarization or update summarization. Automatic evaluation methods for document summarization are still

an ongoing research process. Redundancy elimination in generated summary is also an attractive area of research.

## 6. REFERENCES

- [1] Dragomir R.Radev, Hongyan, Malgorzata Stys and Danial Tam, "Centriod- based summarization of multiple documents", Information Processing and Management, 2004.
- [2] D.R.Radev, Weiguo Fan, "Automatic summarization of search engine hit lists ", *University ofMichigan Business School*.
- [3] Afnan Ullah Khan, Shahzad khan and Waqar Mahmood, "MRST:A NewTechnique for Information Summarization" *World Academy of Science Engineering and Technology*, 2005.
- [4] Zhang Shu ,Zhao Tiejun, Zheng Dequan& Zhao Hua , "Two stage sentence selection approach for multi-Docment summarization", *Journal of electronics*, Vol.2, No.4, July 2008.
- [5] Furu Wei, YanXiang He ,Wenjie Li and Qin Lu, "A Query-Sensitive Graph-Based Sentence Ranking Algorithm for Query-Oriented Multi-Docment Summarization", *International Symposiums on Information Processing*, 2008.
- [6] Xiao-Peng Yang and Xiao-Rong Liu, "Personalized Multi-Docment Summarization in Information Retrieval", *Seventh International Conference on Machine Learning and Cybernetics, Kunming*, 12-15 July 2008.
- [7] Dingding Wang, Tao Li, Shenghou Zhu, Chris Ding, "Multi-Docment Summarization via Sentence -Level Semantic Analysis and Symmetric Matrix Factorization", *SIGIR Singapore*, July 20-24, 2008.
- [8] Ben Hachey, "Multi-Docment Summarization Using Generic Relation Extraction", *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 420-429, 2009.
- [9] Md. Mohsin Ali, Monotosh Kumar Ghosh, and Abdullah-Al-Mamun, "Multi- docment Text Summarization: SimWithFirst Based Features and Sentence Co-selection Based Evaluation", *International Conference on Future Computer and Communication*, 2009.
- [10] Lei Huang, Yanxiang He, Furu Wei, and Wenjie Li, "Modeling Document Summarization as Multi-objective Optimization", *Third International Symposium on Intelligent Information Technology and Security Informatics*, 2010.
- [11] Shu Gong, Youli Qu and Shengfeng Tian, "Subtopic-based Multi-docments Summarization", *Third International Joint Conference on Computational Science and Optimization*, 2010.
- [12] A.Kogilavani and Dr.P.Balasubramani, "Clustering and Feature Specific Sentence Extraction Based Summarization of Multiple Documents", *International Journal of computer science & information Technology (IJCSIT)* Vol.2, No.4, August 2010.
- [13] WB Frakes, CJ Fox, "Strength and Similarity of Affix Removal Stemming Algorithms", *ACM SIGIR Forum*, 2003.
- [14] Harman, D. "How Effective is Suffixing." *Journal of the American Society for Information Science* 42 (1), 1991, 7-15.
- [15] Lovins, J. B. "Development of a Stemming Algorithm",

*Mechanical Translation and Computational Linguistics*11, 1968, 22-31.

- [16] Paice, Chris D. "Another Stemmer.", *SIGIRForum* 24 (3), 1990, 56-61.
- [17] Porter, M. F. "An Algorithm for Suffix Stripping." *Program* 14, 1980, 130-137.
- [18] Fung B, Wnag K & Ester .M, "Hierarchical Document Clustering using Frequent itemsets", *SIAM International Conference on Data Mining, SDM '03*.2003. Pp 59-70

J.Jayabharathy received her M.Tech in 1999 from Department of Computer Science and Engineering , Pondicherry University, Puducherry. She has been working as a Assistant Professor in the Department of Computer Science and Engineering, Pondicherry Engineering College, Puducherry. Currently she is working towards the Ph.D degree in Document clustering. Her areas of interest are Distributed Computing, Grid Computing, Data Mining and Document Clustering.

Dr. S. Kanmani received her B.E and M.E in Computer Science and Engineering from Bharathiyar University and Ph.D in Anna University, Chennai. She had been the faculty of Department of Computer Science and Engineering, Pondicherry Engineering College from 1992 onwards. Presently she is working as Professor in the Department of Information Technology, Pondicherry Engineering College. Her research interests are Software Engineering, Software testing, Object oriented system, and Data Mining. She is the Member of Computer Society of India, ISTE and Institute of Engineers, India. She has published about 65 papers in various International conferences and journals.

Miss Buvana Received her B.Tech(2005) in Computer Science and Engineering from Pondicheery University and Currently Doing her M.tech in Pondicherry Engineering College.



# Improving Web Page Readability by Plain Language

Walayat Hussain<sup>1</sup>, Osama Sohaib<sup>2</sup> and Arif Ali<sup>3</sup>

<sup>1</sup> Department of Computer Science, Balochistan University of I.T. Engineering and Management Science,  
Quetta, Pakistan

Department of CS & IT, University of Balochistan  
Quetta, Pakistan

<sup>3</sup> Department of Information Technology, Balochistan University of I.T. Engineering and Management Science  
Quetta, Pakistan

## Abstract

In today's world anybody who wants to access any information the first choice is to use the web because it is the only source to provide easy and instant access to information. However web readers face many hurdles from web which includes load of web pages, text size, finding related information, spelling and grammar etc. However understanding of web pages written in English language creates great problems for non native readers who have basic knowledge of English. In this paper, we propose a plain language for a local language (Urdu) using English alphabets for web pages in Pakistan. For this purpose we developed two websites, one with a normal English fonts and other in a local language text scheme using English alphabets. We also conducted a questionnaire from 40 different users with a different level of English language fluency in Pakistan to gain the evidence of the practicality of our approach. The result shows that the proposed plain language text scheme using English alphabets improved the reading comprehension for non native English speakers in Pakistan.

**Keywords:** *Web readability, readability enhancement, text readability, lower literate people, typography, content usability, web accessibility.*

## 1. Introduction

In today's life World Wide Web has been considered as an instant and easy source to get any information. No doubt the World Wide Web contributes greatly in creation of an increasing global information database. Using web site anyone can promote its ideas business very easily. Almost all households have access to the internet and can reach the whole world in just few clicks. Beside the uses of internet there are lots of problems associated with the web like lost of web pages, web pages load, and text too small or too large, bad spelling or grammar, finding related

information, appropriate guidance for users to relevant information and when get some information then the understanding of these information [15]. Among all these problems one of the major problem is the understanding of the text and materials written in website. In today's web primary language for the websites are English language due to which non native readers whose primary language is not English faces great problems due to their limited knowledge about unfamiliar vocabulary, the grammar, composition and structure of sentences, self explanatory graphs, and use of abbreviations or intimidating content display [1], which creates lots of problems in web readability. Web readability can be defined as "a combination of reading comprehension, reading speed and user satisfaction in terms of reading comprehension, dictionary, thesaurus and existing online tools and browser add-ones". Readability may also be defined as "how easily a person can read and understand any written materials". Website readability is an indicator of overall difficulty level of a website [1][2].

Many researchers have discussed web readability issues and proposed various ideas to enhance web readability [1][2][3][4], this study also discuss web readability by addressing following research question.

- How to enhance web readability for users whose first language is not an English language?
- Which approach has been adopted to cope up readability issue while using English language and Local languages?

In this paper, we propose a plain language text using English alphabets for web pages. For this purpose we developed two websites, one with a normal English alphabet and other in a local language (Urdu) text scheme using English alphabets with a questionnaire to evaluate our proposal. We conducted a study on 40 different users

with a different level of English language fluency in Pakistan.

The work is organized as follows. Section 2 gives the overview and background of the internet users in Pakistan. Section 3 present some related studies. Section 4 provides the approach of our study. Section 5 and 6 shows interpretations. Finally the study is concluded and leaves some an open issue.

## 2. Background and Motivation

Plain Language is a writing approach that is effective to understand information easily the first time reader's reads it [14].

Pakistan is a multilingual country, it has two official languages: English and Urdu. Urdu is also the national language. Additionally, Pakistan has four major provincial languages Punjabi, Pashto, Sindhi, and Balochi, as well as two major regional languages: Saraiki and Kashmiri [10].

Table 1: Pakistani languages

Languages	Percentage of speakers
Punjabi	44.15
Pashto	15.42
Sindhi	14.10
Siraiki	10.53
Urdu	7.57
Balochi	3.57
Other	4.66

Internet access has been available in Pakistan since 1990s. The country has been following an aggressive IT policy, aimed at enhancing Pakistan's drive for economic modernization and creating an exportable software industry. There is no doubt that has been helping increase the popularity of the Internet. Table 2 shows the number of users within a country that access the Internet [11].

Table 2: Internet users in Pakistan

Year	Internet users	Rank	Percent Change	Date of Information
2003	1,200,000	47		2000
2004	1,500,000	48	25.00 %	2002
2005	1,500,000	49	0.00 %	2002
2006	10,500,000	23	600.00 %	2005
2007	10,500,000	24	0.00 %	2005
2008	17,500,000	17	66.67 %	2007
2009	17,500,000	17	0.00 %	2007
2010	18,500,000	20	5.71 %	2008

Where English is also an official language but it is not the most spoken language of Pakistan. Because English is so widely spoken, it has often been referred to as a world language [12]. That is why English is taught as foreign language in Pakistan, but still the percentage of English fluency is low among the people in Pakistan. The Literacy

rate of Pakistan is (56%). Sindh (58%) and Punjab (58%) are equally more literate as compared to NWFP (50%) and Balochistan (49%) provinces. The percentage of English speakers in the country is only 10.9% [13]. So the users when try to read the information on the web in English, they suffer with web readability. We try to solve this problem by using English alphabets written in local language (Urdu). Although the Google translation of the web pages from English to Urdu (national language of Pakistan) is available, but the main problem with that, it translate the sentence word by word, which does not make the Urdu sentence understandable

Major headings are to be column centered in a bold font without underline. They need be numbered. "2. Headings and Footnotes" at the top of this paragraph is a major heading.

## 3. Related Work

Web becomes more complex with the fast growth of information distributed through web pages especially that use a fashion-driven graphical design but readability of WebPages is not taken into consideration. The readability is an important criterion for measuring the web accessibility especially non-native readers encounter even more problems.

Readability crucial presentation attributes that web summarization algorithms consider while generating a query based web summary. Text on the web of a suitable level of difficulty for rapid retrieval but appropriate techniques needs to be work out for locating it. Readability measurement is widely used in educational field to assist instructors to prepare appropriate materials for students. However, traditional readability formulas are not fit to attract much attention from both the educational and commercial fields [1][2][5][6][7][8][9].

Miller and Hsiang Yu [1] propose a new transformation method, Jenga Format, to enhance web page readability. A user study on 30 Asian users with moderate English fluency is conducted and the results show that the proposed transformation method improved reading comprehension without negatively affecting reading speed. The authors have solved the problems of distraction elimination and content transformation. They have found two important factors, sentence separation and sentence spacing, affecting the reading.

Pang Lau and King [2] propose a bilingual readability assessment scheme for web site in English and Chinese languages. The Experimental results show that, for page readability apart from just indicating difficulty, the estimated score acts as a good heuristic to figure out pages with low textual content such as index and multimedia pages.

Gradišar et al [5] identifies the existence of factors that influence reading experience, the authors examined the readability of combination of 30 different text colors that are presented on the CRT display by measurement of speed of reading through Chapman-Cook Speed. The results show that there are no statistically major differences in readability between 30 color combinations but they have prove an existence of at least five factors, which simultaneously and differently affect readability of a colored text.

Uitdenbogerd [6] experimented and compare the range of difficulty of the text web that is found in traditional hard-copy texts for English as Second Language (ESL) learners using standard readability measures. The results suggest that an on-line text retrieval engine based on readability can be of use to language learners because of the ESL text readability range fall within the range for web text.

Xing et al [7] demonstrates a novel approach, in order to increase the accuracy of readability for measuring English readability applying techniques from natural language processing and information theory. The authors have found by applying the concept of entropy in information theory that the readability differences are not caused by the text itself but by the information gap between text and reader.

Gottron and Martin [8], describes the modern content extraction algorithms that help to estimate accurately the readability of a web document prior to index calculation. The authors observed the SMOG and the FRE index to be far more accurate in combination with CE in comparison to calculating them on the full document.

Kanungo and Orr [9] propose a machine learning methodology that first models the readability of abstracts using training data with human judgments, and then predicts the readability scores for previously unseen documents using gradient boosted decision trees. The performance of the model goes beyond that of other kinds of readability metrics such as Collins-Thompson-Callan, Fog or Flesch-Kincaid. The model can also be used in the automatic summarization algorithm to generate summaries that are more readable

## 4. The Approach

In order to understand the effect of content transformation and to analyze the difference and compare the readability between English language and local language written in English alphabet (plain language) we developed two websites with four web pages each and conducted a formal user study to investigate the effectiveness of both contents from end users point of view [Table 3].

Table 3: web pages information

	Page 1	Page 2	Page 3	Page 4
Website 1	Passage A (English)	Questionnaire	Passage B (Translated)	Questionnaire
Website 2	Page 1 Passage A (Translated)	Page 2 Questionnaire	Page 3 Passage B (English)	Page 4 Questionnaire

Websites contents are from standard IELTS test. In first website first page has small passages written in English language, and second page had another same size passage as first but this was translated into national language (Urdu) using English alphabets. In second website the first page had the same content as the first page of first website but this was translated passage, and the second page had the English version of the second page of first website. At the end of each passage there is a questionnaire which included few MCQ's related to that passage.

The website first register the users, then a reader started the test by reading a first passage after completing questions, at the end of a each test a reader moved to the next test with another passage translated into national language using English alphabet. For each test there were timers which take the duration of time spent on each test. After completing both tests there is another page for the feed back in order to comments about both tests.

## 5. Results

There are 40 users selected for this test: 12 females and 28 males. All the users are from Pakistan whose first language are not English. We have categorized our users into three categories [Table 4]:

- Undergraduate Students.
- Professionals having good knowledge of English language
- Other Workers having basic knowledge of English language

Table 4: user's categories

Category	Education	No. of Users
Professionals	Masters / Bachelors	10
Students	Undergraduate	18
Workers	Secondary School	12

At the end of each passage there are nine questions related to that passage. The result of each group is shown in Table 5 and Figure 1. The time taken in reading both passages of English and a plain language (Local language-Urdu) written in English alphabets is shown in Figure 2.

Table 5: Correct answers attempt by users with time taken

No. of correct answers attempted by Students				
No. of	English	Time/User	Translated	Time/User

cases	Pages		pages	
18	50%	9min	84%	11min
<b>No. of correct answers attempted by Workers, Lower literate user</b>				
No. of cases	English	Time/User	Translated	Time/User
12	9%	40min	86%	19min
<b>No. of correct answers attempted by Professionals</b>				
No. of cases	English	Time/User	Translated	Time/User
10	88%	8min	90%	10min

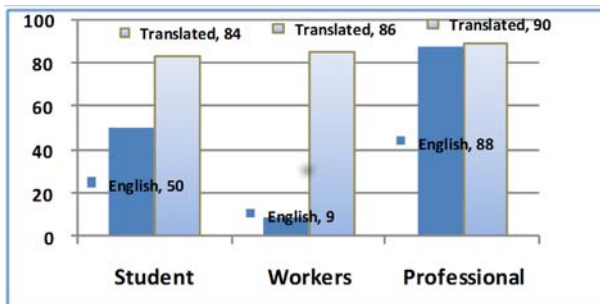


Figure 1: Percentage of correct answers of both tests

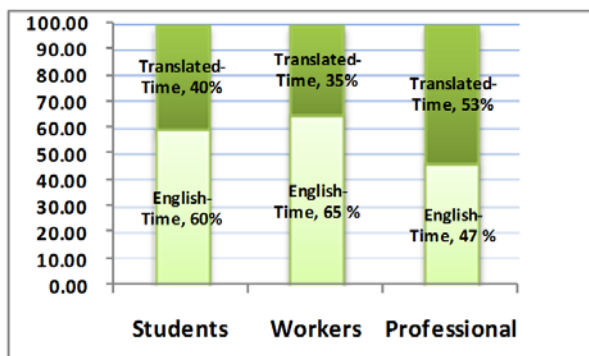


Figure 2: Percentage of Time taken for both test (English passages and Translated passages)

## 5. Findings

Based on the results of the study, we can say that:

1. The transformation of the text content enhances web readability for non native user i.e. whose first language is not English.
2. The translated version of an English text gives better result and the percentage of correct answers is more than English passage text.
3. The ratio of correct answers for translated version is very high in Worker, lower-literate user and in undergraduate student's category because they have only basic and moderate knowledge of English. In Student category there is slight difference while for professionals who have good

understanding of English and have high education is almost equal.

4. As in Pakistan the level for higher education is less than the population having basic education, so the translated version of English text is more readable than English text.

There are many interesting comments from users which they have given at the end of test. Like most of the people are not very much familiar with the translated version that's why it takes longer time for them and they preferred for English version of a text although they made more mistakes while solving the questions of English passage compared to translate version. Many users recommend the translated version of a passage because of easily understandable.

## 5. Conclusion and Future Work

In order to analyze the difference and to compare the readability between English language and a pain language (local language -Urdu) using English alphabets we develop two sites having few pages. We take two passages from the standard IELTS reading passage. At the end of each passage there is a questionnaire for investigating the effect of our approach.

We have observed that by changing the contents of web pages into local language by using English alphabets we get better result. The level of understanding content is very much high for all those who have basic knowledge of English. By using this approach we can enhance web readability to all those non native English speaker countries whose local language text is incompatible with the web pages.

We plan to extend our study in other Asian countries for the non native readers to investigate the usefulness of our approach to web readability.

## References

- [1] C-H. Yu and R. C. Miller, "Enhancing web page readability for non-native readers," Proc Users and attention on the web CHI 2010, Atlanta, GA, USA. April 10-15. Pp. 2523-2531.
- [2] T. P. Lu and I king, "Bilingual Web page and site readability assessment," Proc. WWW, 2006, pp. 993-994.
- [3] Y. Miyazaki and K. Norizuki, "Developing a Computerized readability estimation program with a Web-searching Function to Match Text Difficulty with Individual Learners reading ability," . In: Proceedings of WorldCALL 2008, Fukuoka, Japan, CALICO, 2008, d-111.
- [4] Klare, G. R. A second look at the validity of readability formulas. Journal of reading Behavior, 1976. pp 129-152.
- [5] M. Gradišar, I. Humar, and T. Turk, "Factors Affecting the Readability of Colored Text in Computer Displays," Proc. 28<sup>th</sup> international conference on information technology interfaces, 2006, pp 245 – 250.

- [6] A. L. Uitdenbogerd, "Web Readability and Computer-Assisted Language Learning," Proc. Australasian Language Technology Workshop (ALTW2006), 2006, pp.99-106.
- [7] F. Xing, D. Cheng, J. Pu, "New Approach to Readability Study Based on Information Computing," Proc. International Conference on Advanced Language Processing and Web Information Technology. IEEE press, 2008, pp.156-161.
- [8] T. Gottron, L. Martin, "Estimating Web Site Readability Using Content Extraction," Proc. WWW 2009 MADRID, Poster Sessions, ACM press, 2009, pp 1169-1170.
- [9] T. Kanungo, D. Orr, "Predicting the Readability of Short Web Summaries," Proc. WSDM '09, 2009.
- [10] "Languages of Pakistan," [Online] Available [http://en.wikipedia.org/wiki/Languages\\_of\\_Pakistan](http://en.wikipedia.org/wiki/Languages_of_Pakistan) [Accessed: November 2010].
- [11] "Pakistan internet users," [Online] Available [http://www.indexmundi.com/pakistan/internet\\_users.html](http://www.indexmundi.com/pakistan/internet_users.html) [Accessed: November 2010].
- [12] "English Language," [Online] Available [http://en.wikipedia.org/wiki/English\\_language](http://en.wikipedia.org/wiki/English_language) [Accessed: November 2010].
- [13] "List of countries by English speaking population," [Online] Available [http://en.wikipedia.org/wiki/List\\_of\\_countries\\_by\\_English-speaking\\_population](http://en.wikipedia.org/wiki/List_of_countries_by_English-speaking_population) [Accessed: November 2010].
- [14] "Plain Language," [Online] Available [http://en.wikipedia.org/wiki/Plain\\_language](http://en.wikipedia.org/wiki/Plain_language) [Accessed: November 2010].

**Walayat Hussain** received the BS (Software Development) Hons degree from Hamdard University Karachi, Post Graduation in Computer Science from AIT (Asian Institute of Technology) - Bangkok and MS (Computer Science) degree from BUIITEMS Quetta, Pakistan in 2004, 2008 and 2009 respectively. He is currently working as an Assistant Professor in Department of CS in Balochistan University of IT, Engineering and Management Sciences (BUIITEMS), Quetta Pakistan.

**Osama Sohaib** received the BS (Software Development) Hons degree from Hamdard University Karachi, Post Graduation in Information Management from AIT (Asian Institute of Technology) - Bangkok and MS (Software Engineering) degree from PAF-KIET (Karachi Institute of Economics and Technology), Karachi Pakistan, in 2005, 2008 and 2010 respectively. He is currently working as a Lecturer in Department of CS & IT in University of Balochistan, Quetta Pakistan.

**Arif Ali** received the BSc (Business Information Technology) Hons degree from University of Salford UK in 2006 and MSc Business Information Systems from University of Bolton UK in 2008. He is currently working as Lecturer at Department of IT at Balochistan University of IT, Engineering and Management Sciences (BUIITEMS), Quetta Pakistan since November 2009.



# 2-Jump DNA Search Multiple Pattern Matching Algorithm

Raju Bhukya<sup>1</sup>, DVLN Somayajulu<sup>2</sup>

<sup>1</sup>Dept of CSE, National Institute of Technology,  
Warangal, A.P, India. 506004.

<sup>2</sup>Dept of CSE, National Institute of Technology,  
Warangal, A.P, India. 506004.

## Abstract

Pattern matching in a DNA sequence or searching a pattern from a large data base is a major research area in computational biology. To extract pattern match from a large sequence it takes more time, in order to reduce searching time we have proposed an approach that reduces the search time with accurate retrieval of the matched pattern in the sequence. As performance plays a major role in extracting patterns from a given DNA sequence or from a database independent of the size of the sequence. When sequence databases grow, more efficient approaches to multiple matching are becoming more important. One of the major problems in genomic field is to perform pattern comparison on DNA and protein sequences. Executing pattern comparison on the DNA and protein data is a computationally intensive task. In the current approach we explore a new technique which avoids unnecessary comparisons in the DNA sequence called 2-jump DNA search multiple pattern matching algorithm for DNA sequences. The proposed technique gives very good performance related to DNA sequence analysis for querying of publicly available genome sequence data. By using this method the number of comparisons gradually decreases and comparison per character ratio of the proposed algorithm reduces accordingly when compared to the some of the existing popular methods. The experimental results show that there is considerable amount of performance improvement due to this the overall performance increases.

**Keywords-** Characters, matching, patterns, sequence.

## 1. Introduction

Bioinformatics is the application of computer technology for managing the biological information. Computers are used to gather, store, analyze and integrate biological and genetic information which can then be applied to gene based drug discovery and development. The problem of exact string matching is to find all occurrences of pattern 'P' of size 'm' in the text string 'T' of size 'n'. Researchers have been focused this sphere of research, various techniques and algorithms have been purposed and designed to solve this problem. Exact String matching algorithms are widely used in bibliographic search, question answering application, DNA pattern matching, text processing applications and information retrieval from databases. The pattern matching

problem has attracted a lot of interest throughout the history of computer science, particularly in the present day high performance computing and has used in various computer applications for several decades. These algorithms are applied in most of the operating systems, editors, search engines on the internet, retrieval of information (from text, image or sound) and searching nucleotide or amino acid sequence patterns in genome and protein sequence databases. Bioinformatics is a multi disciplinary science that uses methods and principle from mathematics and computer science and statistics for analyzing biological data where DNA pattern analysis plays a vital role, for various analyses like discrimination of cancer from the gene expression, mutations evolution, protein-protein interaction in cellular activities etc. Pattern matching plays a vital role in various applications in computational biology for data analysis like feature extraction, searching, disease analysis, structural analysis.

Pattern matching focuses on finding the occurrences of a particular pattern of in a text. The problem in pattern discovery is to determine how often a candidate pattern occurs, as well as possibly some information on its frequency distribution across the sequence/text. In general, a pattern will be a description of a set of strings, each string being a sequence of symbols. Hence, given a pattern, it is usual to ask for its frequency, as well as to examine its occurrences in a given sequence/text. Many algorithms have been developed each designed for a specific type of search. Although they all serve the same function but they vary in the way they process the search, and second in the methods they use to efficiently achieve the optimal processing time.

Every human has his/her unique genes. Genes are made up of DNA; therefore the DNA sequence of each human is unique. However, surprisingly, the DNA sequences of all humans are 99.9% identical, which means there is only 0.1% difference. DNA is contained in each living cell of an organism, and it is the carrier of that organism's genetic

code. The genetic code is a set of sequences, which define what proteins to build within the organism. Since organisms must replicate and reproduce tissue for continued life, there must be some means of encoding the unique genetic code for the proteins used in making that tissue. The genetic code is information, which will be needed for biological growth and reproductive inheritance.

DNA is the basic blue print of life and it can be viewed as a long sequence over the four alphabets A, C, G and T. DNA contains genetic instructions of an organism. It is mainly composed of nucleotides of four types. Adenine (A), Cytosine (C), Guanine (G), and Thymine (T). The amount of DNA extracted from the organism is increasing exponentially. So pattern matching techniques plays a vital role in various applications in computational biology for data analysis related to protein and gene in structural as well as the functional analysis. It focuses on finding the particular pattern in a given DNA sequence. The biologists often queries new discoveries against a collection of sequence databases such as GENBANK, EMBL and DDBJ to find the similarity sequences. As the size of the data grows it becomes more difficult for users to retrieve necessary information from the sequences. Hence more efficient and robust methods are needed for fast pattern matching techniques. It is one of the most important areas which have been studied in computer science. The string matching can be described as: given a specific strings  $P$  generally called pattern searching in a large sequence/text  $T$  to locate  $P$  in  $T$ . if  $P$  is in  $T$ , the matching is found and indicates the position of  $P$  in  $T$ , else pattern does not occurs in the given text. Pattern matching techniques has two categories and is generally divides into multiple pattern matching and single pattern matching algorithms.

- Single pattern matching
- Multiple pattern matching techniques

In a standard problem, we are required to find all occurrences of the pattern in the given input text, known as single pattern matching. Suppose, if more than one pattern are matched against the given input text simultaneously, then it is known as, multiple pattern matching. Whereas single pattern matching algorithm is widely used in network security environments. In network security the pattern is a string indicating a network intrusion, attack, virus, and snort, spam or dirty network information, etc. Multiple pattern matching can search multiple patterns in a text at the same time. It has a high performance and good practicability, and is more useful than the single pattern matching algorithms. To determine the function of specific genes, scientists have learned to read the sequence of nucleotides comprising a DNA sequence in a process called DNA sequencing. Comparison, pattern recognition, detecting similarity and phylogenetic trees constructing in genome sequences are the most popular tasks. The

process of sequence alignment allows the insertion, deletion and replacements of symbols that representing the nucleotides or amino acids sequences. From the biological point of view pattern comparison is motivated by the fact that all living organisms are related by evolution. That implies that the genes of species that are closer to each other should show signs of similarities at the DNA level. Moreover, those similarities also extend to gene function. Normally, when a new DNA or protein sequence is determined, it would be compared to all known sequences in the annotated databases such as GenBank, SwissProt and EMBL.

Let  $P = \{p_1, p_2, p_3, \dots, p_m\}$  be a set of patterns of  $m$  characters and  $T = \{t_1, t_2, t_3, \dots, t_n\}$  in a text of  $n$  characters which are strings of nucleotide sequence characters from a fixed alphabet set called  $\Sigma = \{A, C, G, T\}$ . Let  $T$  be a large text consisting of characters in  $\Sigma$ . In other words  $T$  is an element of  $\Sigma^*$ . The problem is to find all the occurrences of pattern  $P$  in text  $T$ . It is an important application widely used in data filtering to find selected patterns, in security applications, and is also used for DNA searching. Many existing pattern matching algorithms are reviewed and classified in two categories.

- Exact string matching algorithm
- Inexact/approximate string matching algorithms

Exact pattern matching algorithm will find that whether the probability will lead to either successful or unsuccessful search. The problem can be stated as: Given a pattern  $p$  of length  $m$  and a string/Text  $T$  of length  $n$  ( $m \leq n$ ). Find all the occurrences of  $p$  in  $T$ . The matching needs to be exact, which means that the exact word or pattern is found. Some exact matching algorithms are Naïve Brute force algorithm, Boyer-Moore algorithm [3], KMP Algorithm [7].

Inexact/Approximate pattern matching is sometimes referred as approximate pattern matching or matches with  $k$  mismatches/ differences. This problem in general can be stated as: Given a pattern  $P$  of length  $m$  and string/text  $T$  of length  $n$ . ( $m \leq n$ ). Find all the occurrences of sub string  $X$  in  $T$  that are similar to  $P$ , allowing a limited number, say  $k$  different characters in similar matches. The Edit/transformation operations are insertion, deletion and substitution. Inexact/Approximate string matching algorithms are classified into: Dynamic programming approach, Automata approach, Bit-parallelism approach, Filtering and Automation Algorithms. Inexact sequence data arises in various fields and applications such as computational biology, signal processing and text processing. Pattern matching algorithms have two main objectives.

- Reduce the number of character comparisons required in the worst and average case analysis.

- Reducing the time requirement in the worst and average case analysis.

In many cases most of the algorithm operates in two stages. Depending upon the algorithm some of the algorithm uses pre-processing phase and some algorithm will search without it. Many Pattern matching algorithms are available with their own merits and demerits based upon the pattern length and the technique they use. Some pattern matching algorithm concentrates on pattern itself. Other algorithm compare the corresponding characters of the patterns and text from the left to right and some other perform the character from the right to left. The performance of the algorithm can be measured based upon the specific order they are compared. Pattern matching algorithms has two different phases.

- Pre-processing phase or study of the pattern.
- Processing phase or searching phase.

The pre-processing phase collects the full information and is used to optimize the number of comparisons. Whereas searching phase finds the pattern by the information collected in pre-processing.

Bioinformatics has found its applications in many areas. It helps in providing practical tools to explore proteins and DNA in number of other ways. Bio-computing is useful in recognition techniques to detect similarity between sequences and hence to interrelate structures and functions. Another important application of bioinformatics is the direct prediction of protein 3-Dimensional structure from the linear amino acid sequence. It also simplifies the problem of understanding complex genomes by analyzing simple organisms and then applying the same principles to more complicated ones. This would result in identifying potential drug targets by checking homologies of essential microbial proteins. Bioinformatics is useful in designing drugs. Pattern matching in biology differs from its counterpart in computer science. DNA strings contain millions of symbols, and the pattern itself may not be exactly known, because it may involve inserted, deleted, or replacement of the symbols. Regular expressions are useful for specifying a multitude of patterns and are ubiquitous in bioinformatics. However, what biologists really need is to be able to infer these regular expressions from typical sequences and establish the likelihood of the patterns being detected in new sequences.

The sequence of DNA constitutes the heritable genetic information in nuclei, plasmids, mitochondria, and chloroplasts that forms the basis for the developmental programs of all living organisms. Determining the DNA sequence is therefore useful in basic research studying fundamental biological processes, as well as in applied

fields such as diagnostic or forensic research. Because DNA is key to all living organisms, knowledge of the DNA sequence may be useful in almost any biological subject area. For example, in medicine it can be used to identify, diagnose and potentially develop treatments for genetic diseases. Similarly, genetic research into plant or animal pathogens may lead to treatments of various diseases caused by these pathogens.

When we know a particular sequence is the cause for a disease, the trace of the sequence in the DNA and the number of occurrences of the sequence defines the intensity of the disease. As the DNA is a large database we need to go for efficient algorithms to find out a particular sequence in the given DNA. We have to find the number of repetitions and the start index and end index of the sequence, which can be used for the diagnosis of the disease and also the intensity of the disease by counting the number of pattern matching strings, occurred in a gene database.

Since children inherit their genes from their parents, they can also inherit any genetic defects. Children and siblings of a patient generally have a 50% chance of also being affected with the same disease. Genetic testing can identify those family members who carry the familial unusual mutation and should undergo annual tumor screening from an early age. Genetic testing can also identify family members who do not carry the familial unusual mutation and do not need to undergo the increased tumor surveillance recommended for patients with unusual mutations. The unusual pattern in the strand reflects in the split strand and hence increases in the unusual mutations increase in the cells. All familial cancer syndromes are caused by a defect in a gene that is important for preventing development of certain tumors. Everybody carries two copies of this gene in each cell, and tumor development only occurs if both gene copies become defective in certain susceptible cells. Genetic testing can help to diagnose by detecting defects in the unusual mutated gene.

The rest of the paper is organized as follows. We briefly present the background and related work in section 2. Section 3 deals with proposed model *i.e.*, 2-JUMP DNA search multiple pattern matching algorithm. Experimental results and discussion are presented in Section 4 and we make some concluding remarks in Section 5.

## 2. Background and Related Work

This section reviews some work related to DNA sequences. An alphabet set  $\Sigma = \{A, C, G, T\}$  is the set of characters for DNA sequence which are used in this algorithm.

The following notations are used in this paper:

DNA sequence characters  $\Sigma = \{A, C, G, T\}$ .

$\phi$  Denotes the empty string.

$|P|$  Denotes the length of the string  $P$ .  
 $S[n]$  Denotes that a text which is a string of length  $n$ .  
 $P[m]$  Denotes a pattern of length  $m$ .  
CPC-Character per comparison ratio.

String matching mainly deals with problem of finding all occurrences of a string in a given text. In most of the DNA applications it is necessary for the user and the developer to be able to locate the occurrences of specific pattern in a sequence. In Brute-force algorithm the first character of the pattern  $P$  is compared with the first character of the string  $T$ . If it matches, then pattern  $P$  and string  $T$  are matched character by character until a mismatch is found or the end of the pattern  $P$  is detected. If mismatch is found, the pattern  $P$  is shifted one character to the right and the process continues. The complexity of this algorithm is  $O(mn)$ . The Bayer-Moore algorithm [3] applies larger shift-increment for each mismatch detection. The main difference the Naive algorithm had is the matching of pattern  $P$  in string  $T$  is done from right to left *i.e.*, after aligning  $P$  and string  $T$  the last character of  $P$  will be matched to the first of  $T$ . If a mismatch is detected, say  $C$  in  $T$  is not in  $P$  then  $P$  is shifted right so that  $C$  is aligned with the right most occurrence of  $C$  in  $P$ . The worst case complexity of this algorithm is  $O(m+n)$  and the average case complexity is  $O(n/m)$ .

In IFBMPMA [12] the elements in the given patterns are matched one by one in the forward and backward until a mismatch occurs or a complete pattern matches. The KMP algorithm [7] is based on the finite state machine automation. The pattern  $P$  is pre-processed to create a finite state machine  $M$  that accepts the transition. The finite state machine is usually represented as the transition table. The complexity of the algorithm for the average and the worst case performance is  $O(m+n)$ .

In IBKMPM [13] algorithm we first choose the value of  $k$  (a fixed value), and divide both the string and pattern into number of substrings of length  $k$ , each substring is called as a partition. If  $k$  value is 3 we call it as 3-partition else if it is 4 then it is 4-partition algorithm. We compare all the first characters of all the partitions, if all the characters are matching while we are searching then we go for the second character match and the process continues till the mismatch occurs or total pattern is matched with the sequence. If all the characters match then the pattern occurs in the sequence and prints the starting index of the pattern or if any character mismatches then we will stop searching and then go to the next index stored in the index table of the same row which corresponds to the first character of the pattern  $P$ .

In approximate pattern matching method the oldest and most commonly used approach is dynamic programming. In 1996 Kurtz [8] proposed another way to reduce the space

requirements of almost  $O(mn)$ . The idea was to build only the states and transitions which are actually reached in the processing of the text. The automaton starts at just one state and transitions are built as they are needed. The transitions those were not necessary will not be built.

The Deviki-Paul algorithm [5] for multiple pattern matching requires a preprocessing of the given input text to prepare a table of the occurrences of the 256 member ASCII character set. This table is used to find the probability of having a match of the pattern in the given input text, which reduces the number of comparisons, improving the performance of the pattern matching algorithm. The probability of having a match of the pattern in the given text is mathematically proved.

In the MSMPMA [18] technique the algorithm scans the input file to find the all occurrences of the pattern based upon the skip technique. By using this index as the starting point of matching, it compares the file contents from the defined point with the pattern contents, and finds the skip value depending upon the match numbers (ranges from 1 to  $m-1$ ). Harspool [6] does not use the good suffix function, instead it uses the bad character shift with right most character. The time complexity of the algorithm is  $O(mn)$ .

Berry-Ravindran [2] calculates the shift value based on the bad character shift for two consecutive text characters in the text immediately to the right of the window. This will reduce the number of comparisons in the searching phase. The time complexity of the algorithm is  $O(nm)$ . Sunday [4] designed an algorithm quick search which scans the character of the window in any order and computes its shift with the occurrence shift of the character  $T$  immediately after the right end of the window. The FC-RJ [11] algorithm searches the whole text string for the first character of the pattern and maintains an occurrence list by storing the index of the corresponding character. Time and space complexity of preprocessing is  $O(n)$ . FC\_RJ uses an array equal to size of the text string for maintaining occurrence list.

Ukkonen [15] proposed automation method for finding approximate patterns in strings. He proposed the idea using a DFA for solving the inexact matching problem. Though automata approach doesn't offer time advantage over Boyer-Moore algorithm [3] for exact pattern matching. The complexity of this algorithm in worst and average case is  $O(m+n)$ . In this every row denotes number of errors and column represents matching a pattern prefix. Deterministic automata approach exhibits  $O(n)$  worst case time complexity. The main difficulty with this approach is construction of the DFA from NFA which takes exponential time and space. Wu.S.Manber.U [16] proposed the algorithm for fast text searching allowing errors. The first bit-parallel method is known as "shift-or" which

searches a pattern in a text by parallelizing operation of non deterministic finite automation. This automation has  $m+1$  states and can be simulated in its non deterministic form in  $O(mn)$  time. The filtering approach was started in 1990. This approach is based upon the fact it may be much easier to tell that a text position doesn't match. It is used to discard large areas of text that cannot contain a match. The advantage in this approach is the potential for algorithms that do not inspect all text characters.

By using dynamic programming approach especially in DNA sequencing Needleman-Wunsch [9] algorithm and Smith-waterman algorithms [14] are more complex in finding exact pattern matching algorithm. By this method the worst case complexity is  $O(mn)$ . The major advantage of this method is flexibility in adapting to different edit distance functions. The Raita algorithm [10] utilizes the same approach as Horspool algorithm[6] to obtaining the shift value after an attempt. Instead of comparing each character in the pattern with the sliding window from right to left, the order of comparison in Raita algorithm [10] is carried out by first comparing the rightmost and leftmost characters of the pattern with the sliding window. If they both match, the remaining characters are compared from the right to the left. Intuitively, the initial resemblance can be established by comparing the last and the first characters of the pattern and the sliding window. Therefore, it is anticipated to further decrease the unnecessary comparisons.

The Aho-Corasick algorithm[1] developed at Bell Labs in 1975 by Alfred Aho and Corasick is an extension of the KMP algorithm [7]. The AC algorithm consists of constructing a finite state pattern matching machine from the keyword and then using the machine to process the text in a single pass. It can find an occurrence of several patterns in the order of  $O(n)$  time, where  $n$  is the length of the text, with pre-processing of the patterns in linear time.

Two dimensional pattern matching methods are commonly used in computer graphics. Takaoka and Zhu proposed using a combination of the KMP[6] and RK methods in an algorithm developed for two dimensional cases. The second approach that runs faster when the row length of the pattern increases and is significantly faster than previous methods proposed. Three dimensional pattern matching is useful in solving protein structures, retinal scans, finger printing, music, OCR and continuous speech. Multi-dimensional matching algorithms are a natural progression of string matching algorithms toward multi-dimensional matching patterns including tree structure, graphs, pictures, and proteins structures.

### 3. 2-JUMP DNA Search Multiple Pattern Matching Algorithm

In this method we use combination of both the techniques

- Index Based Search
- ASCII sum

The index based search has been well established. Here we created index table of the input data and our search skips primarily on the index-row of the first character of the pattern. However in our proposed work, we go one step ahead and rather than using primitive method of comparing single character at a time, we rather compare sum of two characters of both input sequence data and pattern. This reduces our comparisons by one-third (we count one comparison for sum). After we match it completely we go for order checking in the subgroups sequentially until there is a mismatch or it completely matches.

#### 3.1. Algorithm

```
Input[n] : Input character array of length n.
Patt[m] : Pattern character array of length m.
IndexTable[4][n] – index Table of input of length 4*n (ACGT)
Let i,j,startIndex,flag,compare,counter integer variables
i=j=start Index=compare=counter=0.
Flag=1
1. Create the index table.
2. Fetch startIndex as per first letter of pattern.
   startIndex = IndexTable [firstLet][i];
3. while(n-startIndex > m)
   while(j<m)
if(m-j==1) // odd no. of characters in pattern.
   if(input[startIndex+j] != pat[j])
   compare++;
   flag=0;
   break;
   Inp2 =input[startIndex+j]+input[startIndex+j+1];
   Pat2 = pat[j]+pat[j+1];
   Compare++;
   If(inp2!=pat2)
   Flag=0;
   Break;
   Else
   compare++;
If(input[startIndex+j] != pat[j]|| input[startIndex+j+1] != pat[j+1])
   flag=0;
   break;
   If(flag == 1)
   Counter++;
   Else
   Flag=1;
   J=0;
   StartIndex = IndexTable[firstLet][++i];
```

#### 3.2. Index Based Search

This method has been invented and used to reduce the search time drastically. In this method we make an Index table of given input on the basis of characters involved which in our case are  $A,C,G,T$ . So, we have a (4xSize of input) table. Now we concentrate only on the index row of first character of our pattern and continue our comparison



technique from the first index onwards. Based upon our comparisons results of success or failure we can directly jump to next potential occurrence of pattern by moving to the next index in the row chosen. We continue above operations till we finish all indexes of that row. In this way we need not move serially through the input, but rather we only concentrate only on the potential strings.

### 3.3 ASCII SUM (or 2-Jump)

Our unique comparison method adds further benefits to our Index Based Search. Here we use unique property of characters involved in our search patterns and input. As we are dealing with only genetic data, so our domain confines to following four characters *A, C, G, T*. Further reducing these characters to single digits by mod formula.

Table.1. Subscript values of DNA sequence characters

S.No	DNA	ASCII Value	ASCII Value-64	(ASCIIValue-64)%5	Array Subscript
1	A	65	1	1	1
2	C	67	3	3	3
3	G	71	7	2	2
4	T	84	20	0	0

Now we can use unique property of above integers. Any sum of above in combination of two gives a unique number in return.

$$\begin{aligned}
 A + A &\sim 1 + 1 = 2 \\
 A + T &\sim 1 + 0 = 1 \\
 A + G &\sim 1 + 2 = 3 \\
 A + C &\sim 1 + 3 = 4
 \end{aligned}$$

And so on for other integers too. Now we can use this to reduce our both input size and patterns to half the length they actually are, *i.e.*, we combine two neighboring alphabets (or their reduced integers) to give single integers.

E.g. *Sequence=ATTGCCATA*  
 Equivalent integers: 100233101  
*Pattern=GCCA*  
 Equivalent integers: 2 3 3 1

Here the first character of pattern is 'G'. From our sequence we find that first index of character 'G' is at 4. So we start forming groups from 4<sup>th</sup> index onwards. 2-Sum groups starting at 'G' of sequence: (2+3), (3+1) = 5,4. 2-Sum groups of pattern: (2+3), (3+1) = 5,4.

So, now rather than comparing each character/integer separately we can compare two of them in one go. If in one go we find that our pattern string matches a substring of the input, and then we can go further and compare the two characters. This will be necessary as the two characters may exist in reverse order form as compared to that of pattern.

E.g. *input- AT*  
*Pattern- TA*

But, such comparison will be required only if pattern matches. Thus over all we find following result: Say, comparisons found over pattern lengths in general are 'n'. By our methods we reduce them to halves *i.e.*, 'n/2'. Further adding the single comparisons if our pattern matched:  $n/2 + p$ . Where  $p$  is length of pattern, which is generally quite small. Thus taking  $p \rightarrow 0$ . We get total number of comparisons is  $n/2$ . The conversion of input can be done on the fly or while creation of index table.

### 3.4. Trivial Cases in Comparisons

- Case i:* If  $S = \phi$  *i.e.*,  $|S| = 0$  and  $P = \phi$  *i.e.*,  $|P| = 0$  then the number of occurrences of  $P$  in  $S$  is 0.
- Case ii:* If  $S = \phi$  *i.e.*  $|S| = 0$  and for any  $|P| \geq 0$  then the number of occurrences of  $P$  in  $S$  is 0.
- Case iii:* If  $S \neq \phi$  *i.e.*,  $|S| \neq 0$  and for any  $|P| = 0$  then the number of occurrences of  $P$  in  $S$  is 0.
- Case iv:* If  $S \neq \phi$  *i.e.*,  $|S| \neq 0$ ,  $P \neq \phi$  *i.e.*,  $|P| \neq 0$  and  $|S| \leq |P|$  then the number of occurrences of  $P$  in  $S$  is 0.

3.4. To understand the algorithm assume a string  $S=AGAATGCAGCTACAAGGTTCCATTCGTCTCGCACTA$  of 37 characters and pattern  $P=ATGCAG$ . Therefore the string can be viewed as follows in an indexing table.

Table.2. Index values of A,C,G and T sequence characters

<i>T 0</i>	5	11	18	19	23	24	26	28	30	36
<i>A 1</i>	1	3	4	8	12	14	15	22	34	37
<i>G 2</i>	2	6	9	16	17	27	32			
<i>C 3</i>	7	10	13	20	21	25	29	31	33	35

As 'A' being our first character of pattern the target indexes are 1, 3, 4, 8, 12, 14, 15, 22, 34 and 37. Here  $S_2$  and  $P_2$  refer to combination of two characters of input string and pattern respectively.  $S$  and  $P$  refer to whole input and pattern s1. First we begin at index 1 because 'A' is starting from index 1. We then form 2-groups of input and pattern both.  
*i.e.*,  $S_2 = A+G$   
 $P_2 = A+T$   
 Clearly  $S_2 \neq P_2$  therefore  $S \neq P$ . So we skip and go to next index.

2. At index 3 we get another probable match. We form 2-groups of input and pattern both.  
*i.e.*,  $S_2 = A+A$   
 $P_2 = A+T$   
 Again we find  $S_2 \neq P_2$ , so we can match directly from next index.

4. Next we move to index 4. Here,

$$S2 = A+T$$

$$P2 = A+T$$

So we get  $S2=P2$ , we move further to next subgroup,

$$S2 = G+C$$

$$P2 = G+C$$

As  $S2=P2$  we proceed further,

$$S2=A+G$$

$$P2=A+G,$$

As all subgroups have matched we go for checking order in our subgroups. In case of first subgroup, we find character in same order as pattern, so we go for next subgroup. Here also characters are in same order as per pattern. Same follows up to the last subgroup. So we do three more comparisons and over all in 6 comparisons we are getting our pattern matched.

Thus  $S=P$ . We now proceed to next index.

5. Next we move to index 8. Here,

$$S2 = A+G$$

$$P2 = A+T$$

Clearly  $S2 \neq P2$ . Thus we conclude  $S \neq P$  and move to further index.

6. However at 12, we find

$$S2 = A+C$$

$$P2 = A+T.$$

Here too we find  $S2 \neq P2$  giving us  $S \neq P$ . We check for next index now.

7. At index 14,

$$S2 = A+A$$

$$P2 = A+T$$

So  $S2 \neq P2$ . Without further checking we skip to next index.

8. Next at index 15,

$$S2 = A+G$$

$$P2 = A+T$$

Again we have  $S2 \neq P2$ . We need not check further and continue our search from next index.

9. Next at index 22,

$$S2 = A+T$$

$$P2 = A+T$$

We find successful match in this subgroup so we check for next subgroup too,

$$S2=T+C$$

$$P2=G+C$$

But here we find mismatch *i.e.*,  $S2 \neq P2$ . Without checking further we can skip to next index.

10. However at next index *i.e.*, 34 we find that remaining length of input string  $S$  is 4 characters, while our pattern string  $P$ 's length is 6 characters. Therefore it is not possible to match pattern with sequence. So we skip remaining comparisons.

**Proof:** Let  $N$ =Input String say, ATTTGACCTTGAAA...

By converting the string to equivalent numerical sequence using formula,

$$N[i] = (N[i] - 64) \% 5, i = \text{Length of Input.}$$

Now we apply same to Pattern  $P$ ,

$$P[i] = (P[i] - 64) \% 5, i = \text{Length of Pattern.}$$

First we prepare  $P$ ,

$$P[j] = P[i] + P[i+1]$$

$$j++, i+2$$

Where  $P'$  is another array of length half that of  $P$ .

Now we process  $N$ ,

$$2Sum = N[i] + N[i+1], \text{ where } i < \text{length of } P$$

Compare ( $P'[j]$ , 2Sum)

Where Compare function compares the two quantities and breaks the whole operation if it find mismatch.

Thus we see effectively maximum number of comparisons require.

$$\text{Max}(\text{length of}(P'), (\text{length of}(N))/2);$$

in case of even comparisons and

$$\text{Max}(\text{length of}(P'), (\text{length of}(N))/2) + 1;$$

in case of odd comparisons. Also the comparisons are finally going to end as length of  $N$  is finite.

#### 4. Experimental Results and Discussions

In this section we present several experiments comparing our algorithm to the existing algorithms and evaluating with the number and size of patterns on the performance. Each experiment was performed on different pattern sizes and the comparison results are noted. The text file which we used for our experiments was a collection of 1024 nucleotide sequence characters. From the below figure we can draw the following conclusions. As the size of the pattern increases the number of comparisons increases but in the proposed technique as the size increases the number of comparisons decreases in some of the cases. The patterns are randomly chosen from the given file size of 1024 characters.

4.1. The below DNA sequence dataset has been taken for the testing of 2-jump algorithm .The DNA biological sequence  $S \in \Sigma^*$  of size  $n=1024$  and pattern  $P \in \Sigma^*$ . Let S be the following DNA sequence.

“AGAACGCAGAGACAAGGTTCTCATTGTGTCTCGC  
 AATAGTGTACCAACTCGGGTGCCTATTGGCCTCC  
 AAAAAAGGCTGTTCAACGCTCCAAGCTCGTGACCT  
 CGTACTACGACGGCGAGTAAGAACGCCGAGAAG  
 GTAAGGGAATAATGACGCGTGGTGAATCCTATG  
 GGTTAGGATCGTGTCTACCCCAAATTCTTAATAAA  
 AAACCTAGGACCCCTTCGACCTAGACTATCGTAT  
 TATGGACAAGCTTTAACTGTCGTAAGTGGAGGCT  
 TCAAAACGGAGGGACCAAAAAATTTGCTTCTAGC  
 GTCAATGAAAAGAAGTCGGGTGTATGCCCCAATTC  
 CTTGCTGCCCGGACGGCCAGTTCATAATGGGACAC  
 AACGAATCGCGGCCGGATATCACATCTGCTCCTGT  
 GATGGAATTGCTGAATGCGCAGGTGTGCTTATGTA  
 CAATCCACGCGTACTACATCTTGTCTTATGTA  
 GGGTTCAGTTCCTCGCGCAATCATAGCGGTACGAA  
 TACTGCGGCTCCATTCGTTTTGCCGTGTTGATCGG  
 GAATGCACCTCGGGACTTTCGATACGACCTGGG  
 ATTTGGCTATACTCCATTCCTCGCAGTATTTTCGATT  
 GCTCATTAGGCTTTGCGGTAAGTAAGTTCTGGCCA  
 CCCACTTCGAGAAGTGAATGGCTGGCTCCTGAGCG  
 CGTCCCTCCGTACAATGAAGACCGGTCTCGCGCTAA  
 ATTTCCCCAGCTTGTACAATAGTCCAGTTTATTAT  
 CAAAGATGCGACAAATAAATTGATCAGCATAATC  
 GAAGATTGCGGAGCATAAGTTTGAAAACCTGGGA  
 GGTGCCAGAAAACCTCCGCGCCTACTTTCGTCAGG  
 ATGATTAAGAGTATCGAGGCCCGCCGTCAATACC  
 GATGTTCTTCGAGCGAATAAGTACTGCTATTTTGC  
 AGACCCTTTGCCAGGCCTTGTCTAAAGGTATGTTA  
 CTTAATATTGACAATACATGCGTATGGCCTTTTCC  
 GGTAACTCCCTG”.

The index table (*index Tab[4][1024]*) for sequence S is very large in number of DNA sequence characters . For different patterns sizes which has been chosen randomly from the above DNA sequence the number of occurrences and the number of comparisons is shown in the Table. 3. To check whether the given pattern is present in the sequence or not we need an efficient algorithm with less comparison time and complexity. By the current technique different patterns are analyzed and the graph is plotted by using these results and analyzed accordingly. From the below experimental results, improvement can be seen that 2-JUMP algorithm gives good performance compared to the some of the popular methods shown in the Table.4. Here we have taken five fields in the Table .3. The pattern text, number of characters in the pattern, number of occurrences of a pattern, the proposed method and the number of comparisons and comparisons per character. The number of comparisons per character (CPC ratio) which is equal to (Number of comparisons /file size) can be used as a

measurement factor, this factor affects the complexity time, and when it is decreased the complicity also decreases.

Table .3.Experimental results analysis of 2-jump algorithm

S.No	Pattern	Pattern Length	No. of Occur	2-jump	CPC
1	A	1	259	259	0.2
2	AG	2	53	312	0.3
3	CAT	3	11	335	0.3
4	AACG	4	5	434	0.4
5	AAGAA	5	2	441	0.4
6	AAAAAA	6	3	456	0.4
7	AGAACGC	7	2	379	0.3
8	AAAAAAGG	8	1	460	0.4
9	GCTCATTAG	9	1	390	0.3
10	CCTTTCCGG	10	1	377	0.3
11	TTTGCCGTGT	11	1	431	0.4
12	TTCTTAATAAAA	12	1	435	0.4
13	GGGACCAAAAAAT	13	1	392	0.3
14	TTTGCCGTGTTGA	14	1	432	0.4
15	CCTCCAAAAAAGGCT	15	1	382	0.3
16	GGCTGTTCAACGCTCC	16	1	392	0.3
17	TTTTCGATTGCTCATA	17	1	432	0.4
18	GGGATTTGGCTATACTCC	18	1	395	0.3
19	GGCCTTGTCTAAAGGTATG	19	1	393	0.3
20	CCTGAGCGCGTCTCCGTCA	20	1	382	0.3

From the below Table.4. results analysis it has been observed the following in terms of relative performance of our algorithm with some of existing algorithms. To measure the performance of the proposed algorithm with the existing popular algorithm we have used two parameters like CPC (Character per comparison ratio) and number of comparisons which are shown in Table.4. The proposed algorithm gives good performance with the algorithms like MSMPMA, Brute-force, Tri-Match, IKPMPM and Naïve string matching algorithms. From the Table.4. We have taken different pattern sizes from 1 to 16 and analyzed accordingly. In all the different cases the proposed technique gives better performance with existing algorithms.

Table .4.Comparisons of different algorithms with 2-jump

Pattern	2-JUMP		BKMPMPM		MSMPMA		Brute-Force		Tri-Match		Naïve String	
	No.of Com	CPC	No.of Com	CPC	No.of Com	CPC	No.of Com	CPC	No.of Com	CPC	No.of Com	CPC
A	259	0.2	259	0.2	1024	1.0	1024	1.0	1025	1.0	1024	1.0
AG	312	0.3	518	0.5	1230	1.2	1282	1.2	1284	1.2	1281	1.2
CAT	335	0.3	542	0.5	1298	1.2	1318	1.2	1321	1.2	1310	1.2
AACG	434	0.4	614	0.6	1359	1.3	1376	1.3	1380	1.3	1376	1.3
AAGAA	441	0.4	607	0.5	1375	1.3	1388	1.3	1393	1.3	1387	1.3
AAAAAAGG	460	0.4	623	0.6	1394	1.3	1409	1.3	1417	1.3	1407	1.3
TTCTTAATAAAA	435	0.4	634	0.6	1390	1.3	1390	1.3	1402	1.3	1399	1.3
GGCTGTTCAACGCTCC	392	0.3	580	0.5	1349	1.3	1349	1.3	1365	1.3	1349	1.3

Fig.1. Shows comparison of different algorithms with 2-JUMP. The proposed algorithm outperforms when compared with some of the popular algorithms. The current technique gives good performance in reducing the number of comparisons compared with other algorithms. The dotted line shows the 2-jump proposed model where as MSMPMA, Brute-Force, Trie-matching IKPMPM and Naïve string searching are shown by solid lines. From the below graph towards the X-axis we have the pattern size whereas towards Y-axis shows the number of comparisons. If we see the experimental analysis all the other algorithms will give more than 1000 comparisons whereas the proposed technique gives less than 500 comparisons due to the indexed method.

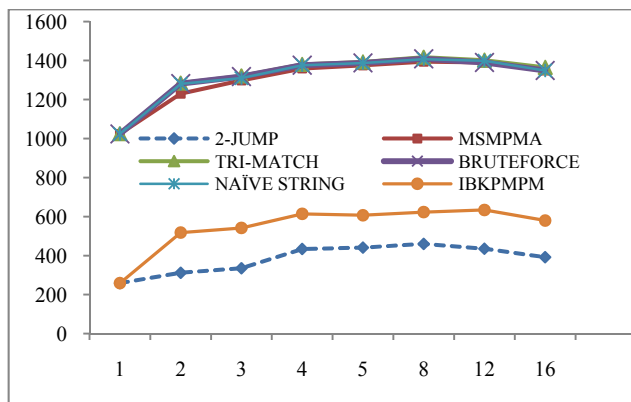


Fig.1. Comparison of different algorithms with 2-JUMP.

The following are observed from the experimental results.

- Reduction in number of comparisons.
- The ratio of comparisons per character has gradually reduced and is less than 1.
- Suitable for unlimited size of the input file.
- Once the indexes are created for input sequence we need not create them again.
- For each pattern we start our algorithm from the matching character of the pattern which decreases the unnecessary comparisons of other characters.
- It gives good performance for DNA related sequence applications.

### Applications in Bioinformatics

Different biological problems of bioinformatics involve the study of genes, proteins, nucleic acid structure prediction, and molecular design.

- Alignment and comparison of DNA, RNA, and protein sequences.
- Gene mapping on chromosomes.
- Gene finding and promoter identification from DNA sequences.
- Interpretation of gene expression and micro-array data.
- Gene regulatory network identification.

- Construction of phylogenetic trees for studying evolutionary relationship.
- DNA and RNA structure prediction.
- Protein structure prediction and classification.
- Molecular design.
- Organize data and allow researchers to access existing information and submit new entries.
- Develop tools and resources which are used for analysis and management of biological data.
- Use sequence data to analyze and interpret the results in a biologically meaningful manner.
- To help researchers in the pharmaceutical industry in drug design process.
- Finding similarities among strings such as proteins of different organisms.
- Finding similarities among parts of spatial structures.
- Constructing of phylogenetic trees called the evolution of organisms.
- Classifying new data according to previously clustered sets of annotated data.

### 5. Conclusion

In this paper we have proposed a new algorithm for DNA pattern matching called 2-jump index based search for DNA pattern matching. The proposed technique enhances the comparison time and the CPC ratio when compared with some of the popular techniques. The proposed algorithm is implemented, analyzed, tested and compared. The experimental result shows that there is a large amount of performance improvement due to this the overall performance increases.

### References

- [1] Aho, A. V., and M. J. Corasick, "Efficient string matching: an aid to bibliographic Search," Communications of the ACM **18** (June 1975), pp. 333-340.
- [2] Berry, T. and S. Ravindran, 1999. A fast string matching algorithm and experimental results. In: Proceedings of the Prague Stringology Club Workshop '99, Liverpool John Moores University, pp: 16-28.
- [3] Boyer R. S., and J. S. Moore, "A fast string searching algorithm" Communications of the ACM **20**, 762-772, 1977.
- [4] D.M. Sunday, A very fast substring search algorithm, Comm. ACM **33** (8) (1990) 132-142.
- [5] Devaki-Paul, "Novel Devaki-Paul Algorithm for Multiple Pattern Matching" International Journal of Computer Applications (0975 - 8887) Vol 13- No.3, January 2011.
- [6] Horspool, R.N., 1980. Practical fast searching in strings. Software practice experience, 10:501-506
- [7] Knuth D., Morris. J Pratt. V Fast pattern matching in strings, SIAM Journal on Computing, Vol 6(1), 323-350, 1977.

- [8] Kurtz, S, Approximate string searching under weighted edit distance. In proceedings of the 3<sup>rd</sup> South American workshop on string processing. Carleton Univ Press, pp. 156-170, 1996
- [9] Needleman, S.B Wunsch, C.D(1970). "A general method applicable to the search for similarities in the amino acid sequence of two proteins." J.Mol.Biol.48,443-453.
- [10] Raita, T. Tuning the Boyer-Moore-Horspool string-searching algorithm. Software - Practice Experience 1992, 22(10), 879-884.
- [11] Rami H. Mansi, and Jehad Q. Odeh, "On Improving the Naive String Matching Algorithm," Asian Journal of Information Technology, Vol.8, No. 1, ISS N 1682-3915,2009, pp. 14-23.
- [12] Raju Bhukya, DVLN Somayajulu, "An Index Based Forward backward Multiple Pattern Matching Algorithm, 'World Academy of Science and Technology..June 2010, pp347-355
- [13] Raju Bhukya, DVLN Somayajulu, "An Index Based K-Partition Multiple Pattern Matching Algorithm", Proc. of International Conference on Advances in Computer Science 2010 pp 83-87.
- [14] Smith, T.F and Waterman, M (1981). Identification of common molecular subsequences T.mol.Biol.147,195-197.
- [15] Ukkonen, E., Finding approximate patterns in strings J.Algor. 6, 1985, 132-137.
- [16] Wu S., and U. Manber, "Agrep — A Fast Approximate Pattern-Matching Tool," Usenix Winter 1992 Technical Conference, San Francisco (January 1992), pp. 153-162.
- [17] Wu, S., Manber U., and Myers, E. 1996, A sub-quadratic algorithm for approximate limited expression matching. Algorithmica 15,1,50-67, Computer Science Dept, University of Arizona, 1992.
- [18] Ziad A.A Alqadi, Musbah Aqel & Ibrahim M.M.EI Emary, Multiple Skip Multiple Pattern Matching algorithms. IAENG International Vol 34(2), 2007.



Raju Bhukya has received his B.Tech in Computer Science and Engineering from Nagarjuna University in the year 2003 and M. Tech degree in Computer Science and Engineering from Andhra University in the year 2005. He is currently working as an Assistant Professor in the Department of Computer Science and Engineering in National Institute of Technology,

Warangal, Andhra Pradesh, India. He is currently working in the areas of Bio-Informatics.

Somayajulu DVLN has received his M. Sc and M. Tech



degrees from Indian Institute of Technology, Kharagpur in 1984 and in 1987 respectively, and his Ph. D degree in Computer Science & Engineering from Indian Institute of Technology, Delhi in 2002. He is currently working as Professor and Head of Computer Science & Engineering at National Institute of Technology, Warangal. His current research interests are bio-

informatics, data warehousing, database security and Data Mining.



# Data Structure & Algorithm for Combination Tree To Generate Test Case

Ravi Prakash Verma<sup>1</sup>, Bal Gopal<sup>2</sup> and Md. Rizwan Beg<sup>3</sup>

<sup>1</sup> Department of Computer Science and Engineering, Integral University  
Lucknow, Utter Pradesh, 226026 India

<sup>2</sup> Department of Computer Applications, Integral University  
Lucknow, Utter Pradesh, 226026 India

<sup>3</sup> Department of Computer Science and Engineering, Integral University  
Lucknow, Utter Pradesh, 226026 India

## Abstract

The combinations play an important role in software testing. Using them we can generate the pairs of input parameters for testing. However until now we have the tabular representations for combination pairs or simply the charts for them. In this paper we propose the use of combination trees which are far easier to visualize and handle in testing process. This also gives the benefits of the remembering the combination of input parameters which we have tested and which are left, giving further confidence on the quality of the product which is to be released.

**Keywords:** *Software testing, combination trees, Data structures, algorithm*

## 1. Introduction

The software testing is one the most important activity in the SDLC [4]. It authenticate whether the software being developed solves the intended purpose or not [2]. "Software systems continuously grow in scale and functionality" [1]. Therefore large size and complexity of software can introduces more error, bugs and faults, in this situation testing becomes more important to uncover errors, bug & faults before software is actually put to use. Software testing also confirms that software being developed as per requirements [5]. At present it is mostly done manually and the test cases are written by the tester, it is a manual activity [3] [6]. This is most error prone area as important path or case may be missed out by the tester [3]. The testers develop test cases on the basis of the combinations of value of input parameters taken one at a time, these test cases are represented in the tabular

form. It becomes difficult to remember that all the combination have been listed out or not. Further it difficult to visualize that whether we have covered all input parameters decisions that can be taken by the user. The combination trees can show the decision or action taken by the uses in a sequence which is very important for the software developer and tester to prove the robustness of the software system being developed. Testing done on the bases of combination trees [7] ensures that we are covering every possible action that can be taken by the user or at least can ensure that software system performs correctly if valid condition & action are chosen. In this paper we present a formal data structure and algorithm to generate the combination trees from the set of elements represented in array.

## 2. Proposed work

The number of  $k$ -combinations from a given set  $S$  of  $n$  elements (distinct and no repeating) is often denoted by

${}^n C_k$  which is  $\binom{n}{k} = \frac{n(n-1)\dots 1}{k(k-1)\dots 1}$ . When  $k > n/2$  then

$\binom{n}{k} = \binom{n}{n-k}$  for  $0 \leq k \leq n$ . The total number of

combination from  $n$  distinct elements is  $= {}^n C_0 + {}^n C_1 + {}^n C_2 + \dots + {}^n C_{n-1} + {}^n C_n$  which is  $2^n$  or  $\sum_{i=0}^{i=n} {}^n C_i$ . As we see that  ${}^n C_0$

represents null or empty elements in the set, however this

is not the case in testing as this represents the case where we do not have input, so ignoring this we have =  ${}^n C_1 +$

${}^n C_2 + \dots + {}^n C_{n-1} + {}^n C_n$  which is  $2^n - 1$  or  $\sum_{i=1}^{i=n} {}^n C_i$ . For

example if we have  $S = \{a, b, c\}$ ,  $n = |S| = 3$ . The total number of combination are given by  ${}^3 C_1 + {}^3 C_2 + {}^3 C_3 = 3 + 3 + 1 = 7$  or  $2^3 - 1 = 8$ . The sets are given as follows  $\{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}$ . If we want to generate combination tree for this set  $S$  we start with root, which represents the null or empty set initially, this is level zero. For making level 1 then we add all the distinct elements from the set and make root as their parent not that the number of levels in the combination tree are  $n+1$  where  $n$  represents number of distinct nodes, the level start from 0, 1, 2, ...,  $n$ . After that we add (make child) next element from the set  $S$  higher in some order preferably lexicological order to the first child at level 1, once these are fixed we select next child and here also we take element higher in lexicological order and add them until all elements in the set are exhausted. Then the same is repeated until all levels are occupied. The combination tree representation of the combination just generated is shown by the tree in figure 1.

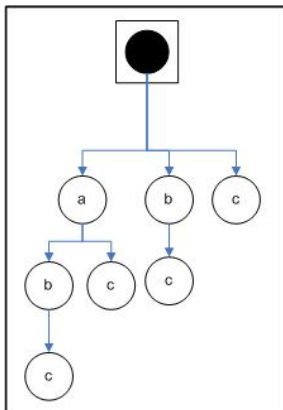


Figure 1. Showing combination tree

The sets and its element can be represented as conditions or the input given to the software module. The combination trees connects these conditions and input values and we can it imitate the users action and choices if follow a particular part in the combination tree. It gives complete listing of action that users can do. The testers can follow a particular path and decide what software should be doing under a situation and decide whether the software module should pass or fail on particular path.

Now we formalize the above method into algorithm and give its supporting data structures. First of all we need a

structure to represent a tree node having data, pointer to parent and pointers to child, which is given as follows.

```

struct node { char [ ] value ;
              int iChild;
              structure node *Parent;
              structure node *Child [Max];
            }
  
```

The Max can take value of  $N$ , where  $N$  is the number of elements in the set  $S$  represented by array. Next we define auxiliary function to create a node, which is given as follows.

```

struct node * Root = NULL;

node * makeNode(char data, int nC, int i)
{ node * temp = (node *) malloc(sizeof(node));
  if (i == 1)
    { temp->value = data; }
  else
    { temp->value = 'R'; }
  temp->Parent = NULL;
  temp->iChild = nC;
  for (int j = 0; j < temp->iChild; ++j)
    { temp->Child[j] = (node *) malloc(sizeof(node));
      temp->Child[j] = NULL;
    }
  return (temp);
}
  
```

The root of the tree is the special node having no data but it has pointers to its children and its parent field is set to NULL. The auxiliary function to create root node is called with “ $nC$ ” as “Max” and “1” as “0”.

We need an auxiliary array or list to store the nodes at given level which server as parent to the child below the current level. The linked list representation of pointers to nodes is used to store intermediate result. One of the advantages provided by this storage is that it avoids back tacking and traversal. The size of this pointer array first increases then it starts to reduce and finally reduces to zero size in length. This happens because in  $\sum_{i=1}^{i=n} {}^n C_i$ ,  ${}^n C_i$

equals  ${}^n C_{n-i}$ , which is  $2^{(n-1)} - 1$ . For this we define node structure PPNode and “addParentPointer” auxiliary functions to add nodes in the list and “removeNodeFromHead()” to delete the added nodes from the beginning in FIFO order. The PPHead & PPTail are pointers to handle the list. These are as follows.

```

struct PPNode { struct node * N;
               struct node * next;
             };
  
```

```

struct PPNode * PPHead = NULL;
struct PPNode * PPTail = NULL;

void addParentPointer(node * n)
{ PPNode * temp = (PPNode*) malloc
    (sizeof(PPNode));
  temp->N = n;
  temp->Next = NULL;
  if (PPHead == NULL && PPTail == NULL)
  { PPHead = temp;
    PPTail = PPHead;
    Root = n;
  }
  else
  { PPTail->Next = temp;
    PPTail = temp;
  }
}

void removeNodeFromHead()
{ PPNode * temp = (PPNode *) malloc
    (sizeof(PPNode));
  temp = PPHead;
  if (PPHead != NULL && PPHead->Next !=NULL)
  { PPHead = PPHead->Next; }
  else
  { PPHead = NULL; }
  free(temp);
}
    
```

Another auxiliary function is used to set the index value such that the element in the Array is greater than its parent in terms of lexicographical order, this is given as follows.

```

int setIndex(PPNode * T)
{ int j = 0;
  char x = T->N->value;
  for (int i = 0; i < Max; ++i)
  { if (x == Array[i])
    { j = i;
      i = Max;
    }
  }
  return (j+1);
}
    
```

Last we need an array to store the distinct elements and Max is the number of elements in array. To start creating the tree we set head & tail of the linked list to NULL and root of the tree to NULL. Finally the "createCombinationTree" function creates the combination tree and is given as follows.

```

void cCTree(int _Max)
{
1. addParentPointer(makeNode(NULL, _Max, 0));
2. i = 0;
3. while (PPHead != NULL)
4. { j = 0;
5.   while ( i < _Max)
6.   { node * n = makeNode(Array[i], _Max-i-1, 1);
7.     n->Parent = PPHead->N;
8.     PPHead->N->Child[j] = n;
9.     addParentPointer(n);
10.    i = i + 1;
11.    j = j + 1;
12.  }
13.  j = 0;
14.  removeNodeFromHead();
15.  i = setIndex(PPHead);
}
}
13. temp = temp->next;
14. removeNodeFromHead();
15. i = setIndex(temp);
}
}
    
```

### 3. Proof and analysis

For a set of elements S containing n elements a combination tree can be generated, where the elements are distinct and repetition in generated combination are not allowed. In order to prove that combination tree algorithm generates all the combination successfully and the loops terminate and the algorithm halts, we use the loop invariance method [8], which is given as follows:

#### 3.1. Proof

**Initialization:** Prior to the beginning of the loop the link list "ParentPointerNode" is empty.

**Maintenance:** To see that, at each iteration maintains the loop invariance we start with the root, that is the first node that is added, i is initialized to zero and the immediate child of the root gets insert into the tree as well as in the list. Once the insertion is complete we remove the first node root from the list and this time the i gets the new value 1 and this time also the list is not empty but contains the new roots at next level. Once the value of i is exceeds the maximum number of elements then new node are not being added to the list instead they are removed from the head.

**Termination:** At termination we see that node are removed one by one as i get the value always higher then

maximum, therefore nodes are removed one by one and finally the list becomes empty.

### 3.2. Complexity Analysis

To establish the upper bound in the proposed algorithm, to represent the worst case run time, we have to do approximation at various places in order to simplify the analysis. We start by measuring the upper bound of various auxiliary procedure used and then using them in the proposed algorithm for final rough estimation. The function "makeNode(data)", "makeRootNode()" and "setIndex(struct PPNode \* T)" have the complexity of  $O(n)$ . The complexity of "setIndex(struct PPNode \* T)" is the approximate value as the complexity decreases as the node starts taking it places in the tree since first time it get called it takes  $n$  units of time, second time it takes  $n-1$  units of time and finally it stats taking  $O(1)$  time. The functions void "addParentPointer(struct \* node)" and void "removeNodeFromHead()" take  $O(1)$  time. for the algorithm "createCombinationTree" we start with step 1 which takes  $O(n)$  time, step 2 takes  $O(1)$  time, step 3 has a loop which executes taking  $({}^nC_1 + {}^nC_2 + \dots + {}^nC_{n-1} + {}^nC_n = 2^n - 1)$   $O(2^n)$  time, step 4 take  $O(2^n)$  time, step 5 is loop taking maximum time of  $O(n2^n)$ , 6 takes  $O(n)$  time step 7-12 take  $O(1)$  individually & they are in two loops therefore take total time of  $O(n2^n)$ , step 13 take  $O(1)$  and finally step 14 takes total time of  $O(n2^n)$ . Summing up the total time of each step we get

$$= O(n) + O(1) + O(1) + O(2^n) + O(2^n) + O(n2^n) + O(1) + O(1) + O(1) + O(1) + O(1) + O(1) + O(1) + O(1) + O(n2^n).$$

$$= O(n) + 10O(1) + 2O(2^n) + 2O(n2^n)$$

Ignoring constant we have

$$= O(n2^n) + O(2^n) + O(n)$$

$$= O(2^n(n+1)) + O(n)$$

Ignoring lower order terms we have

$$= O(n2^n)$$

So the approximate worst case complexity of the creating combination tree is  $O(n2^n)$ .

### 4. Conclusion & future work

The combinations can be generated by reading the vertices and follow leading edges as path to other vertices, when we start from a root & descend to child, the combination pair is, all node encountered while descending from root to the leaves of the tree. There fore to generate combination pair having 2 elements we have to descend to depth of two. The root of the tree is at depth zero, so we follow every path from the root to depth of two. This is how we have generated the combination tree which assumes that there are distinct

elements in the set  $S$  having  $n$  number of elements. We have generated non repeating combination with over all complexity of  $O(n2^n)$ . For the future work we should try to establish more accurate upper bound on the algorithm and also reduce the fixed space take by each node as the number of child of a node in the combination tree varies, these are maximum for the roots & decrease when we descend in the tree, therefore memory requirement drops and also the number of sub paths decrease.

### References

- [1] Kaschner, K., Lohmann, N., "Automatic Test Case Generation for Interacting Services". In Proc. of ICSSOC 2008 Workshops. Volume 5472 of Lecture Notes in Computer Science. (2009)
- [2] Tony Hoare, "Towards the Verifying Compiler", In The United Nations University / International Institute for Software Technology 10th Anniversary Colloquium: Formal Methods at the Crossroads, from Panacea to Foundational Support, Lisbon, March 18–21, 2002. Springer Verlag, 2002.
- [3] Robert V. Binder, "Testing Object-Oriented Systems: Models, Patterns, and Tools", Addison Wesley Longman, Inc., 2000.
- [4] S. S. Riaz Ahamed, " Studying the feasibility and importance of software testing: An Analysis", International Journal of Engineering Science and Technology, Vol.1(3), 2009, 119-128.
- [5] Glenford J. Myers, "The Art of Software Testing", Second Edition, John Wiley & Sons, Inc.
- [6] B. Beizer "Software Testing Techniques", Van Nostrand Reinhold, 2nd edition, 1990.
- [7] Jaroslav Nesetril, "ASPECTS OF STRUCTURAL COMBINATORICS (Graph Homomorphisms and Their Use)", TAIWANESE JOURNAL OF MATHEMATICS Vol. 3, No. 4, pp. 381-423, December 1999
- [8] Thomas H Cormen, Clifford Stein, Ronald L Rivest, Charles E Leiserson, "Introduction to Algorithms (2001)", McGraw-Hill

# Generation of test cases from software requirements using combination trees

Ravi Prakash Verma<sup>1</sup>, Bal Gopal<sup>2</sup> and Md. Rizwan Beg<sup>3</sup>

<sup>1</sup> Department of Computer Science and Engineering, Integral University  
Lucknow, Utter Pradesh, 226026 India

<sup>2</sup> Department of Computer Applications, Integral University  
Lucknow, Utter Pradesh, 226026 India

<sup>3</sup> Department of Computer Science and Engineering, Integral University  
Lucknow, Utter Pradesh, 226026 India

## Abstract

Requirements play an important role in conformance of software quality, which is verified and validated through software testing. Usually the software requirements are expressed natural language such as English. In this paper we present an approach to generate test case from requirements. Our approach takes requirements expressed in natural language and generates test cases using combination trees. However until now we have the tabular representations for combination pairs or simply the charts for them. In this paper we propose the use of combination trees which are far easier to visualize and handle in testing process. This also gives the benefits of remembering the combination of input parameters which we have tested and which are left, giving further confidence on the quality of the product which is to be released.

**Keywords:** *Software testing, combination trees, Data structures, algorithm, Software Requirements, test cases*

represented in the tabular form. It becomes difficult to remember that all the combination have been listed out or not. Further it difficult to visualize that whether we have covered all input parameters decisions that can be taken by the user. The trees can show the decision or action in a sequence which is very important for the software developer and tester to prove the robustness of the software system being developed. Testing done on the bases of combination trees [7] ensures that we are covering every possible action that can be taken by the user or at least can ensure that software system performs correctly if valid condition & action are chosen. In this paper we have proposed the algorithm to generate the test cases from by the use of combination trees and then we combine these trees to generate a single tree. The path traced from root to the node and finally to the leave nodes give the test case.

## 1. Introduction

The software testing is one the most important activity in the SDLC [4]. It authenticate whether the software being developed solves the intended purpose or not [2]. "Software systems continuously grow in scale and functionality" [1]. Software testing confirms that software being developed as per requirements [5]. At present it is mostly done manually and the test cases are written by the tester, it is the Ad-hoc activity [3] [6]. This is most error prone area as important path or case may be missed out by the tester [3]. The testers develop test cases on the basis of the combinations of value of input parameters taken one at a time, these test cases are

## 2. Proposed work

For the sake of understanding we take one example of the requirement and demonstrate the how the test cases are to be generated from software requirements using combination trees. As we know there are lots of software systems being developed which are GUI based. We pick one of common software requirement which is part of in fact every software system which is GUI based, which is "the user should be able to log in to the system". From here onwards we formalize our approach which is as follows.



## 2.1 Identification of classes of input

As we see that there are six controls on the Login Form namely two Textboxes, two Labels and two buttons. This login form is shown in the figure below (figure 1)

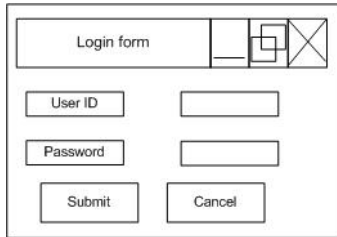


Figure 1. Sample login form

Let us establish which control receives which type of input from the user the “UserID” & “Password” textboxes receive user ID & password respectively, while the labels have fixed caption for the same. The buttons “Submit and Cancel receive the click Events. On the basis of the classes of input controls used in the form we can separate the distinct classes, over here in his case we have “textbox” and “Buttons”.

The “Text” input to the control textbox can be any value from the superset as the set

AN = {alpha-numeric characters like a-z, A-Z}  
 SC = {Special characters like '\$','#','!', '~','\*', ...}  
 NC = {(numeric characters like 0-9)}

Text = {AN, SC, NC}

Any input can be classified into valid & invalid class and the in case of text it is constraint by length possibly  $c_1 \leq k \leq c_2$ , where  $c_1$  and  $c_2$  are finite and  $c_1 \neq c_2$ . Now we define the input into valid, invalid and show the desired length. Now lets us give each cell a number so that it could be differentiated with each other and handling becomes easy, from now onwards we will use these numbers and to understand what they are indicating to we have to refer the following tables.

Table 1. classification of inputs of Textboxes

SN	Input	Length	Valid	Invalid
1	Text <sub>UID</sub>	>6 (1)	alpha-numeric characters {a-z, A-Z} (2)	Special characters like {'\$', '#', '!', '~', '*', ...} numeric characters like {0-9} i.e. Text - AN(3)
2	Text <sub>P</sub>	> 6 (4)	Text (5)	-

Table 2. classification of inputs of buttons

SN	Object/Control	Event	Embedded procedure/function	Action
1	Submit Button	Event Click <sub>S</sub> B (6)	Calls Match: which matches user name & password (7)	If Match successful: Go to Home Page (8) If Match unsuccessful: Display Message (9)
2	Cancel Button	Event Click <sub>C</sub> B (10)	Calls Clear All Textboxes (11)	All text boxes are cleared (12)

The condition or statement represented by any number can be complimented as, For example we see that (1) in table 1 represents that the textbox which accepts the user id of the user should allow a user id greater than the length six, so notation (1') means that user id is less than length six. We that the input that is accepted by this form under the above requirement should have (1)·(2) and another statement can be generated by taking the compliment of (1)·(2) which is (1')·(2) which mean the input is any combination from the set AN but length is less than six. “.” implies that both the statements are to be imposed simultaneously. Now we individually take one row from the table and put it into arrays. For table 1, row 1 the arrays elements are 1.2 & 1.3 and it compliment is 1'·2 & 1'·3. For table 1, row 2 the arrays elements are 4.5 and it compliment is 4'·5. Similarly for table 2, row 1 the array elements are 6.7, 8, 9 and for table 2, row 2 the array elements are 10.11 & 10.12. For the array we are generating a combination tree with the following algorithm and creating an orchid with trees representing each array. We will need a following data structure:

```
struct node {
    char [ ] value ;
    structure node *Parent;
    structure node *Child [Max];
}
```

Roots is an array of node which are used to store the different roots of the tree and is defined as follows

```
struct Roots {
    struct node * N;
    struct node * next;
} Roots[MaxNumberOfArrays];
```

```

struct Roots * RootsHead = NULL;
struct Roots * RootsTail = NULL;

void addRoot(struct * node)
{ if (RootsHead == NULL && RootsTail == NULL)
  { RootsHead = (struct *Roots) malloc(sizeof(struct
Roots));
  RootsTail = (struct *Roots) malloc(sizeof(struct
Roots));
  RootsHead->N = node;
  RootsHead->next = NULL;
  RootsTail = RootsHead;
  }
else
{ struct Roots * temp = (struct *Roots)
malloc(sizeof(struct Roots));
temp = RootsTail;
temp ->N = node;
temp->next = NULL;
RootsTail->next = temp;
RootsTail = temp;
Free(temp);
}
}

void removeNodeFromHead()
{ if (RootsHead != NULL)
  { struct Roots * temp = (struct *Roots)
malloc(sizeof(struct Roots));
temp = RootsHead;
temp = temp->next;
RootsHead = temp;
}
}

int countRoots(struct Roots * RootsHead)
{ if (RootsHead != NULL)
  { int i = 1;
  struct Roots * temp = (struct *Roots)
malloc(sizeof(struct Roots));
temp = RootsHead;
while (temp != RootsTail)
  { temp = temp->next;
  i = i + 1;
  }
return (i);
}
else
{ return (0); }
}

struct node * makeRootNode(char [] NameOfArray)
{ struct node * temp = (struct * Roots)
malloc(sizeof(struct Roots));

```

```

temp->value = NameOfArray;
temp-> Parent = NULL;
for (int i = 0; i < MAX + 1; ++i)
  { temp-> Child[i] = NULL }
return (temp);
}

```

```

bool match(char [ ] NameOfArray)
{ struct node * temp = (struct * Roots)
malloc(sizeof(struct Roots));
temp = RootsHead;
while (temp != RootsTail)
  { if (temp->value = NameOfArray)
    { return (True) ;
    temp = RootsTail;
    }
temp = temp->next
}
return (False);
}

```

The linked list representation of pointers to nodes is used to store intermediate result. One of the advantages provided by this storage is that it avoids back tacking and traversal. The size of this pointer array first increases then it starts to reduce and finally reduces to zero size in length. This happens because of  $\sum_{i=1}^{i=n} n c_i$ , which is  $2^{(n-1)} - 1$ .

```

struct ParentPointerNode { struct node * N;
                          struct node * next;
};

```

```

struct ParentPointerNode * ParentPointerHead = NULL;
struct ParentPointerNode * ParentPointerTail = NULL;

```

```

void addParentPointer(struct * node)
{ if (ParentPointerHead == NULL &&
ParentPointerTail == NULL)
  { ParentPointerHead = (struct *ParentPointerNode)
malloc(sizeof(struct ParentPointerNode));
  ParentPointerTail = (struct *ParentPointerNode)
malloc(sizeof(struct ParentPointerNode));

  ParentPointerHead->N = node;
  ParentpointerHead->next = NULL;
  ParentPointerTail = ParentPointerHead;
}
else
{ ParentPointerTail ->next = node;
  ParentPointerTail = node;
}
}

```

```

void removeNodeFromHead()

```



8. system shows home page

With the help of this procedure we can connect the orchid into a single tree

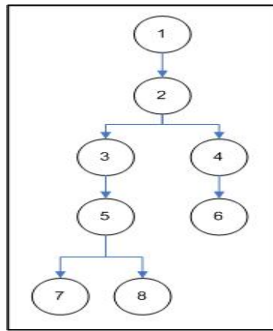


Figure 3. Control flow graph of the procedure to login to the system using login form

### 2.4 Elimination of child

Combination tree shows all possible combination, it does not considers where they are meaning full or not, certain combinations generated by the above algorithm are impossible to realize for example in the above case we can see that if by pressing “Submit” button the use may go with situation 8 or 9 (see table ) but not the both one after the other or if “Click” event of the button is not fired then either 8 nor 9 can be possible. Therefore there could be many such cases present in the combination tree which are infeasible, absurd or not possible altogether. To eliminate such cases we have to parse the entire collection of tree under certain rules which eliminate these combinations. This rule should be developed only for the trusted & standard components, whose behaviors is known and has been thoroughly tested. For example in our case it’s the “Button”. Following rules can be defined using a rule set.

**Definition:** Rule set is the set of edges or set of possible productions. Let S be set of rules and L be the set of symbols denoted by  $L = \{L_1, L_2, L_3, \dots, L_n\}$ , with which we express the rules or productions. For example in our case the set of symbols is  $L = \{6.7, 8, 9\}$  and the rule set S is defined as follows:

$$S \rightarrow 6.7S \mid S8 \mid S9$$

Now we can produce all applicable rules with the production system these are as follows

**Rule 1**

$$S \rightarrow 6.7S$$

$$S \rightarrow 6.78$$

**Rule 2**

$$S \rightarrow 6.7S$$

$$S \rightarrow 6.79$$

We define the production set  $P = \{6.78, 6.79\}$  and apply it over the orchid then we eliminate edges from root to 8, root to 9, and 8 to 9. Similarly for others and the resulting orchid is given in figure below (figure 4).

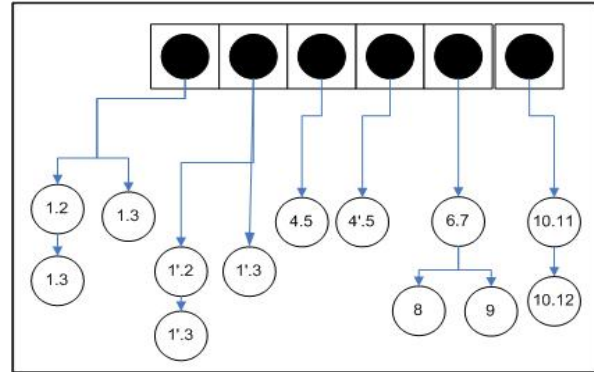


Figure 4. After elimination of children

### 2.5 Elimination of roots

The elimination of roots is possible by merging the trees which represent the complimentary conditions originating from same steps of control flow graph. As Roots [0] & Roots [1] originate from same step 1 of the flow control and Roots [2] & Roots [3] also originate from same step 1 of the flow control. The new orchid is shown in figure 5.

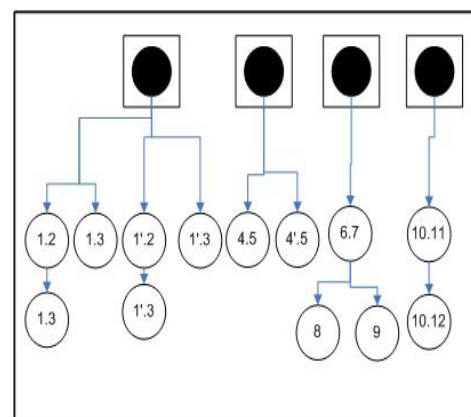


Figure 5. After elimination of roots

## 2.6 Combining trees

We can see that if we do not reduce the combination tree then we would have huge number of possibility and number of test case generated will be very large. As we have developed a control flow graph for the object under test, if we use that then we could limit the number of possibilities by which user can interact with the form, with the help of this we fix the merger of tree as follows (figure 6)

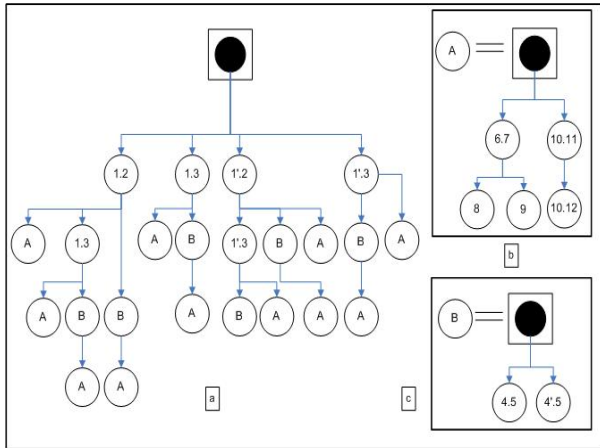


Figure 6. After combining trees

Now we add two additional nodes an extension node, expected result pass node and expected result fail node. The expected result pass node is the node where the software/module/form should comply with the intended purpose of the software requirement further its child fields are set to NULL (see figure 7). The expected result fail node is not actually indicate the failure of the software/module/form instead it indicate that software/module/form should raise an error message or it should not allow users to continue. Here also the child fields of expected result fail node are set to NULL. The aforementioned nodes are graphically shown in the figure below. These nodes are attached as leaflet of the tree forming external nodes. We can fix these nodes with help of tables and flow control generated finally we get the following (see figure 8)

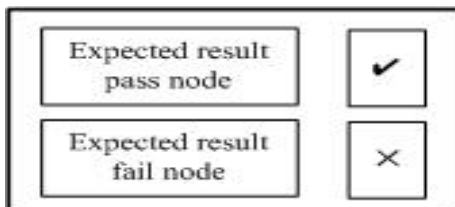


Figure 7. Additional nodes

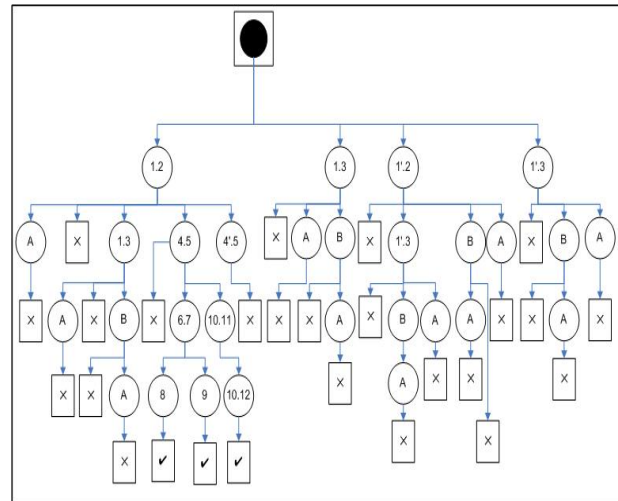


Figure 8. Final combination tree

## 3. Result and analysis

To get the test case we have to descend from the root to its child and where ever we find a terminating leaves we list the nodes encountered and that becomes the test case with the expected result motioned in the leaves whether it passes or fails. In doing so we get 4 test cases at level 2, 22 test case at level 3, 28 test case at level 4 and finally 9 at level 5. So in total we have 63 test cases. Among all test cases generated so far we have 3 test cases where we have the expected results pass. As we can see that in this simple case can produce enormous amount of test case, however in practice only some are created and only few are executed.

## 4. Conclusion and future work

It has been impossible to think about such number when we create test cases on ad-hoc bases, however it may not be possible to execute all of them but at least we discover the test cases in which the system should pass successively under given choices of inputs and action by user. We can deliver the system on the bases of selecting the test case in which there is expected result pass while maturing & increasing our confidence on system by performing more test as system is operational. If we find any bugs or fault we can fix them later on. The optimal testing is necessary to establish quality control. Our future work will be to release a tool to support our claim as it is not possible to manually generate such amounts



of test case and we would probabilistically determine the optimality in execution of test cases over such standard software components such as login form.

## References

- [1] Kaschner, K., Lohmann, N., "Automatic Test Case Generation for Interacting Services". In Proc. of ICSOC 2008 Workshops. Volume 5472 of Lecture Notes in Computer Science. (2009)
- [2] Tony Hoare, "Towards the Verifying Compiler", In The United Nations University / International Institute for Software Technology 10th Anniversary Colloquium: Formal Methods at the Crossroads, from Panacea to Foundational Support, Lisbon, March 18–21, 2002. Springer Verlag, 2002.
- [3] Robert V. Binder, "Testing Object-Oriented Systems: Models, Patterns, and Tools", Addison Wesley Longman, Inc., 2000.
- [4] S. S. Riaz Ahamed, " Studying the feasibility and importance of software testing: An Analysis", International Journal of Engineering Science and Technology, Vol.1(3), 2009, 119-128.
- [5] Glenford J. Myers, "The Art of Software Testing", Second Edition, John Wiley & Sons, Inc.
- [6] B. Beizer "Software Testing Techniques", Van Nostrand Reinhold , 2nd edition, 1990.
- [7] Jaroslav Nesetril, "ASPECTS OF STRUCTURAL COMBINATORICS (Graph Homomorphisms and Their Use)", TAIWANESE JOURNAL OF MATHEMATICS Vol. 3, No. 4, pp. 381-423, December 1999

# Evolutionary Biclustering of Clickstream Data

R.Rathipriya<sup>1a</sup>, Dr. K.Thangavel<sup>1b</sup>, J.Bagyamani<sup>2c</sup>

<sup>1</sup>Department of Computer Science, Periyar University, Salem, Tamilnadu, India

<sup>2</sup>Department of Computer Science, Government Arts College, Dharmapuri, Tamilnadu, India

## Abstract

Biclustering is a two way clustering approach involving simultaneous clustering along two dimensions of the data matrix. Finding biclusters of web objects (i.e. web users and web pages) is an emerging topic in the context of web usage mining. It overcomes the problem associated with traditional clustering methods by allowing automatic discovery of browsing pattern based on a subset of attributes. A coherent bicluster of clickstream data is a local browsing pattern such that users in bicluster exhibit correlated browsing pattern through a subset of pages of a web site. This paper proposed a new application of biclustering to web data using a combination of heuristics and meta-heuristics such as K-means, Greedy Search Procedure and Genetic Algorithms to identify the coherent browsing pattern. Experiment is conducted on the benchmark clickstream msnbc dataset from UCI repository. Results demonstrate the efficiency and beneficial outcome of the proposed method by correlating the users and pages of a web site in high degree. This approach shows excellent performance at finding high degree of overlapped coherent biclusters from web data.

**.Keywords:** — Biclustering, Clickstream data, Coherent Bicluster, Genetic Algorithm, Greedy Search Procedure, Web Mining.

## 1. Introduction

The World Wide Web is the one of the important media to store, share, and distribute information in the large scale. Nowadays web users are facing the problems of information overload and drowning due to the significant and rapid growth in the amount of information and the number of users. As a result, how to provide web users with more exactly needed information is becoming a critical issue in web-based information retrieval and web applications.

Web mining [5,15] discovers and extracts interesting pattern or knowledge from web data. It is classified into three types as web content mining, web structure, and web usage mining. Web usage mining is the intelligent data mining technique to mine clickstream data in order to extract usage patterns. These patterns are analyzed to

determine user's behavior which is an important and challenging research area in the web usage mining.

Clickstream data is a sequence of Uniform Resource Locators (URLs) browsed by the user within a particular period of time. To discover pattern of group of users with similar interest and motivation for visiting the particular website can be found by clustering. Traditional clustering [6] is used to cluster the web users or web pages based on the existing similarities. When a clustering method is used for grouping users, it typically partitions users according to their similarity of browsing behavior under all pages. However, it is often the case that some users behave similarly only on a subset of pages and their behavior is not similar over the rest of the pages. Therefore, traditional clustering methods fail to identify such user groups.

To overcome this problem, concept of Biclustering or Coclustering was introduced. Biclustering[2-4] was first introduced by Hartigan and called it direct clustering[11]. The application of biclustering in web mining is ideal when users have multiple interest/behavior over different subsets of web pages. Biclustering attempts to cluster web user and web pages simultaneously based on the users' behavior recorded in the form of clickstream data. It identifies the subset of users which show similar interest/behavior under a specific subset of web pages. These browsing pattern play vital role in E-commerce based applications. Recommender systems analyze patterns of user browsing interest and to provide personalized services which match user's interest in most business domains, benefiting both the user and the merchant.

The objective of the proposed method is to identify set of subgroup of users and set of subgroup of pages with maximum volume such that these users and pages are highly correlated.

The rest of the paper is organized as follows. Section 2 describes some of the biclustering approaches available in the literature. Methods and materials required for biclustering approach are described in the section 3. Section 4 focuses on the proposed Biclustering framework

using Genetic Algorithm. Analysis of experimental results is discussed in the Section 5. Section 6 concludes the paper with features for future enhancements.

## 2. Related Work

Koutsonikola, V.A. et al.[13] proposed a bi-clustering approach for web data, which identifies groups of related web users and pages using spectral clustering method on both row and column dimensions. S. Araya et al.[14] proposed methodology for target group identification from web usage data which improved the customer relationship management e.g. financial services. Sujatha et al.[16] proposed a novel method to improve the cluster quality using Genetic Algorithm (GA) for web usage data. Guandong et al.[10] presented an algorithm using bipartite spectral clustering to extract bicluster from web users and pages and the impact of using various clustering algorithms is also investigated in that paper.

The followings are the some of the biclustering algorithm available in the literature. Cho et al. [4] introduced K-Means based biclustering algorithms that identifies 'm' row clusters and 'n' column clusters while monotonically decreasing the Mean Square Residue score defined by Cheng and Church. Dhillon et al.[8] proposed an innovative co-clustering algorithm that monotonically increases the preserved mutual information by intertwining both the row and column clusterings at all stages. Tang et al.[17] introduced a framework for unsupervised analysis of gene expression data which applies an interrelated two-way clustering approach on the gene expression matrices. Kluger et al.[12] proposed a method to discover the biclusters with coherent values and looked for checkerboard structures in the data matrix by integrating biclustering of rows and columns with normalization of the data matrix. Another approach called Double Conjugated Clustering (DCC) which aims to discover biclusters with coherent values defined using multiplicative model of bicluster by Busygin et al.[3]

Coupled Two Way Clustering algorithm[9] was introduced by Getz et al. which performs one way clustering on the rows and columns of the data matrix using stable clusters of row as attributes for column clustering and vice versa. Bleuler et al. [2] propose a evolutionary algorithm framework that embeds a greedy strategy. Chakraborty et al. [5] use genetic algorithm to eliminate the threshold of the maximum allowable dissimilarity in a bicluster.

In literature, biclustering algorithms are widely applied to the gene expression data. Most of these algorithms are failed to extract the coherent pattern from

the data matrix. In web mining, there is no related work that has been applied specific biclustering algorithms for discovering the coherent browsing patterns.

In this paper, Greedy Search Procedure and evolutionary approach namely Genetic Algorithm (GA) is introduced to obtain the optimal coherent browsing patterns. The results show that GA outperforms the greedy procedure by identifying coherent browsing patterns. These patterns are very useful in the decision making for target marketing.

## 3. Methods and Materials

### 3.1 Preprocessing

Clickstream data pattern is converted into web user access matrix A by using (1) in which rows represent users and columns represent pages of web sites. Let A(U, P) be an 'n x m' user access matrix where U be a set of users and P be a set of pages of a web site. It is used to describe the relationship between web pages and users who access these web pages. Let 'n' be the number of web user and 'm' be the number of web pages. The element  $a_{ij}$  of A(U, P) represents frequency of the user  $U_i$  of U visit the page  $P_j$  of P during a given period of time.

$$a_{ij} = \begin{cases} \text{Hits}(U_i, P_j), & \text{if } P_j \text{ is visited by } U_i \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where  $\text{Hits}(U_i, P_j)$  is the count/frequency of the user  $U_i$  accesses the page  $P_j$  during a given period of time.

### 3.2 Coherent Bicluster

A bicluster with coherent values is the subset of users and subsets of pages with coherent values on both rows and columns. A measure called Average Correlation Value (ACV)[1] is used to measure the degree of coherence of the biclusters.

### 3.3 Average Correlation Value

It is used to evaluate the homogeneity of a bicluster. Matrix B = ( $b_{ij}$ ) has the ACV which is defined by the following function,

$$ACV(B) = \max \left\{ \frac{\sum_{i=1}^n \sum_{j=1}^n |r_{-rowij}| - n}{n^2 - n}, \frac{\sum_{k=1}^m \sum_{l=1}^m |r_{-colkl}| - m}{m^2 - m} \right\} \quad (2)$$

$r_{-rowij}$  is the correlation between row i and row j,

$r_{col_{kl}}$  and is the correlation between column k and column l. A high ACV suggests high similarities among the users or pages. ACV can tolerate translation as well as scaling. And also works well for biclusters in which there's a linear correlation among the users or pages.

### 3.4. Greedy Search Procedure

A greedy algorithm repeatedly executes a search procedure which tries to maximize the bicluster based on examining local conditions, with the hope that the outcome will lead to a desired outcome for the global problem. This approach employs simple strategies that are easy to implement and most of the time quite efficient.

#### Structure of Greedy Search Procedure

Step 1: Start with initial bicluster.

Step 2: For every iteration

Add/ remove the element(user/page) to/from the bicluster which maximize the objective function.

End for

In this paper, objective function is to maximize ACV of a bicluster.

### 3.5 Encoding of Biclusters

Each enlarged and refined bicluster is encoded as a binary string. The length of the string is the number of rows plus the number of columns of the user access matrix A (U, P). A bit is set to one when the corresponding user or page is included in the bicluster. These binary encoded biclusters are used as initial population for genetic algorithm.

### 3.6 Volume of Bicluster

The number of elements in bicluster B (I, J) is called the volume of bicluster B (I, J) and denoted as VOL (B (I, J)).

$$\text{VOL} (B (I, J)) = |I| \times |J| \quad (3)$$

where, |I| is the number of users in the B and |J| is the number of pages in B.

## 4. Coherent Biclustering Approach Using Evolutionary Algorithm

The proposed algorithm is used to identify the optimal coherent biclusters in terms of volume and quality in three subsequent steps. First step is to identify the initial biclusters called seeds by using K-Means clustering

algorithm. Second step is to enlarge and refine these seeds using greedy search procedure which results in local optimum. Third step is to obtain global optimum of biclusters using evolutionary technique called genetic algorithm. These overlapped coherent biclusters have high degree of correlation among subset of users and subset of related pages of a web site.

This algorithm identifies the coherent browsing pattern from the web usage data which plays vital role in the direct marketing and target marketing. One-to-one relation between web users and pages of a web site is not appropriate because web users are not strictly interested in one category of web pages. Therefore, the proposed algorithm is tuned to discover the overlapping coherent biclusters from clickstream data patterns.

### 4.1 Bicluster Formation using K-Means Algorithm

In this paper, K-Means clustering method is applied on the web user access matrix A(U, P) along both dimensions separately to generate  $k_u$  user clusters and  $k_p$  page clusters. And then combine the results to obtain small co-regulated submatrices ( $k_u \times k_p$ ) called biclusters. These correlated biclusters are called seeds.

### 4.2 Enlargement and Refinement of Bicluster Using Greedy Search Procedure

In this step, seeds are enlarged and refined by adding /removing the rows and columns to enlarge their volume and improve their quality respectively. The main goal of the greedy search procedure is to maximize the volume of the bicluster seed without degrading the quality measure.

Here, ACV is used as merit function to grow the seeds. Insert/Remove the users/pages to /from the bicluster if it increases ACV of the bicluster.

#### Algorithm 1: Seed Enlargement and Refinement using Greedy Search Procedure

**Input:** User Access Matrix A

**Output:** Set of enlarged and refined biclusters

- Step 1. Compute  $k_u$  user clusters and  $k_p$  page clusters from preprocessed clickstream data.
- Step 2. Combine  $k_u$  and  $k_p$  clusters to form  $k_u \times k_p$  biclusters called seeds.
- Step 3. For each seed do  
Call Seed Enlargement(Seed(U, P))

Call Seed Refinement(Seed( U, P))  
 Step 4. Return enlarged and refined biclusters

**Algorithm 2: Seed Enlargement (Seed(U, P))**

**Input:** Set of seeds.

**Output:** Set of enlarged seeds.

Step 1. Set of users ‘u’ not in U  
 Step 2. Set of pages ‘p’ not in P  
 Step 3. For each node u/p do  
     If  $ACV(\text{union}(\text{Seed}, u/p)) > ACV(\text{Seed}(U, P))$  then  
         Add u/p to Seed(U, P)  
     End(if)  
 End(for)  
 Step 4. Return Enlarged Seed

**Algorithm 3: Seed Refinement (Enlarged Seed(U, P))**

**Input:** Set of seeds.

**Output:** Set of refined seeds.

Step 1. For each node u/p in U/P  
     Remove node u/p from Enlarged Seed ,  
     U’/P’ be set of rows/columns in U/P but  
     not contained u/p  
     If  $ACV(\text{Enlarged Seed}(U', P')) > ACV(\text{Enlarged Seed}(U, P))$   
     Update U/P  
 End(if)  
 End(for)  
 Step 2. Return Refined seed as bicluster

Enlarging and refining the seed starts from page list followed by user list until ACV is increased using greedy search procedure.

**4.3 Coherent Biclustering Framework using Genetic Algorithm (GA)**

The GA is a stochastic global search method that mimics the metaphor of natural biological evolution. GA operates on a population of potential solutions applying the principle of survival of the fittest to produce better and better approximations to a solution. At each generation, a new set of approximations is created by the process of selecting individuals according to their level of fitness in the problem domain and breeding them together using operators borrowed from natural genetics. This process leads to the evolution of populations of individuals that are better suited to their environment than the individuals that they were created from, just as in natural adaptation.

Biclustering approach is viewed as optimization problem with the objective of discovering overlapping coherent biclusters with high ACV and high volume. In this paper, Genetic Algorithm (GA) is used for optimization of bicluster. The important feature of GA is that it provides a number of potential solutions to a given problem and the choice of final solution is left to the user.

Usually, GA is initialized with the population of random solutions. In order to avoid random interference, biclusters obtained from greedy search procedure are used to initialize GA. This will result in faster convergence compared to random initialization. Maintaining diversity in the population is another advantage of initializing with these biclusters.

**Fitness Function**

The main objective of this work is to discover high volume biclusters with high ACV. The following fitness function  $F(I, J)$  is used to extract optimal bicluster.

$$F(I, J) = \begin{cases} |I|*|J|, & \text{if } ACV(\text{bicluster}) \geq \delta \\ 0, & \text{Otherwise} \end{cases} \quad (4)$$

Where  $|I|$  and  $|J|$  are number of rows and columns of bicluster and  $\delta$  is defined as follows

$$ACV \text{ threshold } \delta = \text{Max}(ACV(P))$$

Here, the objective function should be maximized. P is the set of biclusters in each population,  $mp$  is the probability of mutation,  $r$  is the fraction of the population to be replaced by crossover in each population,  $cp$  is the fraction of the population to be replaced by crossover in each population,  $n$  is the number of biclusters in each population. The biclustering framework using genetic algorithm is given below.

**Algorithm 4: Evolutionary Biclustering Algorithm**

**Input:** Set enlarged and refined seed

**Output:** Optimal Bicluster

Step 1. Initialize the population  
 Step 2. Evaluate the fitness of individuals  
 Step 3. For  $i=1$  to  $\text{max\_iteration}$   
     Selection()  
     Crossover()  
     Mutation()  
     Evaluate the fitness  
 End(For)



Step 4. Return the optimal bicluster

Selection: The most commonly used form of GA selection is Roulette Wheel Selection (RWS), is used for the selection operator. When using RWS, a certain number of biclusters of the next generation are selected probabilistically, where the probability of selecting a bicluster solution  $S_{n_i}$  is given by

$$Pr(S_{n_i}) = \frac{Fitness(S_{n_i})}{\sum_{N=1 \dots n} Fitness(S_n)} \quad (4)$$

With RWS, each solution will have a probability to survive by being assigned with a positive fitness value. A solution with a high volume has a greater fitness value and hence has a higher probability to survive. On the other side, weaker solutions also have a chance to survive the selection process. This is an advantage, as though a solution may be weak, it may still contain some useful components.

Crossover and Mutation: Then  $cp$  of parents is chosen probabilistically from the current population and the crossover operator will produce two new offsprings for each pair of parents using one point crossover technique on genes and conditions separately. Now the new generation contains the desired number of members and the mutation will increase or decrease the membership degree of each user and page with a small probability of mutation  $mp$ .

### 5. Experimental Results and Analysis

The experiments are conducted on the well-known benchmark clickstream dataset called msnbc dataset which was collected from MSNBC.com portal. This data set is taken from UCI repository, where the original data is preprocessed using equation 1. There are 989,818 users and only 17 distinct items, because these items are recorded at the level of URL category, not at page level, which greatly reduces the dimensionality.

The length of the clickstream record starts from 1 to 64. Average number of visits per user is 5.7. Intuitively, very small and very large number of URL category visited may not provide any useful information about the user's behavior. Thus, the length of the record having less than 5 is considered as a very small and record length greater than 15 is considered as a very large. During data filtering process, small and large records are removed from the dataset.

The metric index R is used to evaluate the overlapping degree between biclusters. It quantifies the amount of overlapping among biclusters. Degree of overlapping[7], is used as quantitative index to evaluate quantitatively the quality of generated biclusters. The degree of overlapping among all biclusters is defined as follows

$$R = \frac{1}{|U| * |P|} \sum_{i=1}^{|U|} \sum_{j=1}^{|P|} T_{ij}$$

where

$$T_{ij} = \frac{1}{(N-1)} * \left( \sum_{k=1}^N W_k(a_{ij}) - 1 \right) \quad (5)$$

where N is the total number of biclusters, |U| represents the total number of users, and |P| represents the total number of pages in the data matrix A. The value of  $w_k(a_{ij})$  is either 0 or 1. If the element (point)  $a_{ij}$   $\in$  A is present in the  $k^{th}$  bicluster, then  $w_k(a_{ij}) = 1$ , otherwise 0. Hence, the R index represents the degree of overlapping among the biclusters. If R index value is higher, then degree of overlapping of the generated biclusters would be high. The range of R index is  $0 \leq R \leq 1$ .

During the bicluster formation step, K-Means clustering algorithm is applied along the both dimensions to generate  $k_u$  and  $k_p$  clusters and combined these clusters to get  $k_u * k_p$  initial biclusters called seeds. Seeds are the biclusters whose volume is small. During the second step, seeds are enlarged and refined iteratively using Greedy Search Procedure. These seeds are enlarged and refined at incremental of ACV to reach high volume which is evident from the Table 1 and Table 2.

Table 1: Performance of Biclustering using Greedy Search Procedure

	Seed Formation Phase	Seed Enlargement and Refinement Phase
No. of Seeds	114	114
Average ACV	0.4711	0.9413
Average Volume	494.9	1599.8

Table 2: Step-Wise Performance of Biclustering using Greedy Search Procedure

	Average ACV	Average Volume
Initial Bicluster	0.4711	494.9
Seed After Column Insertion	0.8420	758.2
Seed After Row Insertion	0.8848	3488.6
Seed After Column deletion	0.9395	1600.5
Seed After Row deletion	0.9413	1599.8

Each bicluster seed underwent four stages of seed enlargement and refinement step. During each stage their ACV is incremented which is shown in Fig 2. Since the quality of the bicluster is more important than the volume, the volume adjusted in order to achieve the high ACV in various stages of the second phase which is portraits in Fig1.

To avoid random interference, very tightly correlated biclusters obtained using greedy search procedures are used as initial population for GA. Moreover, it results in quick convergence and provides number of potential biclusters. These biclusters have high ACV and high volume which is obvious from table 4. This approach shows excellent performance at finding high degree of overlapped coherent biclusters from web data.

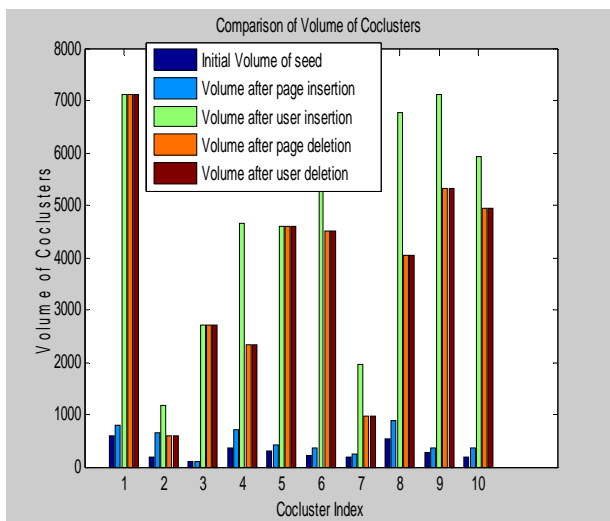


Fig 1. Volume of Biclusters in Various Stages

Table 3: Parameter Setting for GA

Crossover Probability	0.7
Mutation Probability	0.01
Population Size	114
Generation	100
ACV Thersold	0.95

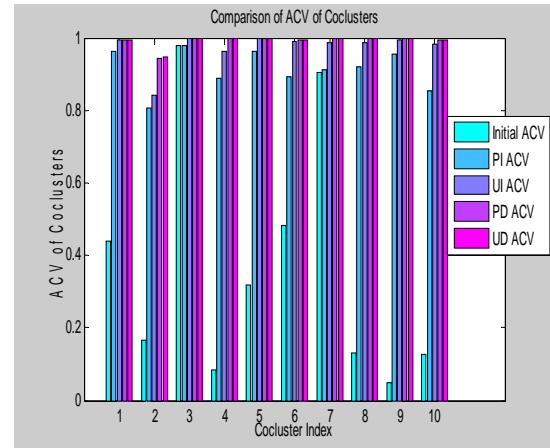


Fig 2. ACV of Biclusters in Various Stages

Table 4: Performance of Biclustering using GA

Mean Volume	Mean ACV	Row Percentage	Column percentage	Overlapp -ing Degree
12715	0.9609	99.9	82.35	0.2152

Table 5: Comparison of Average Volume and Homogeneity of biclusters

	Average Volume	Average ACV	Overlapping Degree
Two-Way K-Means	494.9	0.4711	0
Greedy Search Procedure	1599.8	0.9413	0.0192
Genetic Algorithm	12715	0.9609	0.2152

These biclusters exhibit coherent pattern on a subset of dimensions. In clickstream analysis, the frequency of visiting the pages of a web site of two users may rise or fall synchronously in response to a set of their interest. Though the magnitude of their interest levels may

not be close, but the pattern they exhibit can be very much similar. Our proposed biclustering frame work is interested in finding such coherent patterns of bicluster of users and with a general understanding of users' browsing interest. This method makes significant contribution in the field of web mining, E-Commerce applications and etc.

From the results, it is obvious that it correlates the relevant users and pages of a web site in high degree of homogeneity. Analyzing these overlapping coherent biclusters could be very beneficial for direct marketing, target marketing and also useful for recommending system, web personalization systems, web usage categorization and user profiling. The interpretation of biclustering results is also used by the company for focalized marketing campaigns to improve their performance of the business.

## CONCLUSION

The main contribution of this paper is twofold namely, development of coherent biclustering framework using GA to identify overlapped coherent biclusters from the clickstream data patterns and a coherence quality measure called ACV is used to get coherent biclusters in last two phases of the biclustering framework. The interpretation of the biclustering results can also be used towards improving the website's design, information availability and quality of provided services. The overlapping nature of the proposed framework can significantly contribute towards this direction. This method has potential to identify the coherent patterns automatically from the clickstream data. Future work aims at extending this framework by enriching clustering process would result to enhanced clusters' quality and a more accurate definition of relation coefficients.

## REFERENCES

- [1] Bagyamani, J., Thangavel, K., SIMBIC: similarity based biclustering of expression data. Communications in Information Processing and management, Book chapter, Springer, vol: 70, pp: 437-441, 2010.
- [2] Bleuler, S., Prelic, A., Zitzler, E.: An EA framework for biclustering of gene expression data. In: Congress on Evolutionary Computation CEC2004, vol:1, pp:166-173, 2004.
- [3] Busygin S, Jacobsen G, Kramer E, Double conjugated clustering applied to leukemia microarray data. SIAM data mining workshop on clustering high dimensional data and its applications, 2002.
- [4] Cho H, Dhillon IS, GuanY, Sra S. Minimum sum-squared residue co-clustering of gene expression data. In: Proceedings of the fourth SIAM international conference on data mining, 2004.
- [5] Chakraborty, A., Maka, H, Biclustering of gene expression data using genetic algorithm, IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology, pp:1-8, 2005.
- [6] Chu-Hui Lee, Yu-Hsiang Fu, Web Usage Mining Based on Clustering of Browsing Features, Eighth International Conference

- on Intelligent Systems Design and Applications, vol. 1, pp:281-286, 2008
- [7] Das C, Maji P, Chattopadhyay S, A Novel Biclustering Algorithm for Discovering Value-Coherent Overlapping  $\sigma$ -Biclusters, Advanced Computing and Communications, pp:148-156, 2008.
- [8] Dhillon IS, Mallela S, Modha DS. Information-theoretic co-clustering. In: Proceedings of the ninth ACM SIGKDD international conference on knowledge discovery and data mining(KDD). pp: 89-98, 2003.
- [9] Getz G, Levine E, Domany E. Coupled two-way clustering analysis of gene microarray data, PNAS ,pp:12079-12079, 2000.
- [10] Guandong Xu, Yu Zong, Peter Dolog and Yanchun Zhang, Co-clustering Analysis of Weblogs Using Bipartite Spectral Projection Approach, Knowledge-Based and Intelligent Information and Engineering Systems, Lecture Notes in Computer Science, Vol: 6278, pp: 398-407, 2010.
- [11] Hartigan JA., Direct clustering of a data matrix, Journal of the American Statistical Association ,pp:123-9, 1972.
- [12] KlugerY, Basri R, Chang JT, Gerstein M., Spectral biclustering of microarray data: biclustering genes and conditions, Genome Research ,pp:703-716, 2003.
- [13] Koutsonikola, V.A. and Vakali, A. ,A Fuzzy bi-clustering approach to correlate web users and pages, Int. J. Knowledge and Web Intelligence, Vol. 1, No. 1/2, pp.3-23, 2009.
- [14] Sandro Araya, Mariano Silva, Richard Weber, A methodology for web usage mining and its application to target group identification, Fuzzy Sets and Systems, pp:139-152, 2004.
- [15] Srivastava, J., Cooley R., Deshpande, M., Tan, P.N., Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data. SIGKDD Explorations, Vol. 1, No. 2, pp:12-23, 2000.
- [16] Sujatha N, Iyakutty K, Refinement of Web usage Data Clustering from K-means with Genetic Algorithm, European Journal of Scientific Research, Vol.42, No.3 pp:478-490, 2010.
- [17] Tang Cand Zhang A, Interrelated Two-Way Clustering: An Unsupervised Approach for Gene Expression Data Analysis, Proc. Second IEEE Int'l Symp. Bioinformatics and Bioeng., Vol. 14, pp:41-48, 2001.

# Transmission Power Level Selection Method Based On Binary Search Algorithm for HiLOW

Lingeswari V Chandra, Selvakumar Manickam, Kok-Soon Chai and Sureswaran Ramadass

National Advanced IPv6 Centre, Universiti Sains Malaysia  
National Advanced IPv6 Centre,  
6<sup>th</sup> Floor, School of Computer Science Building,  
Universiti Sains Malaysia (USM),  
11800 Minden, Penang,  
Malaysia

## Abstract

Recently the sensor communication research has introduced an IP-based communication known as 6LoWPAN to sensor network. 6LoWPAN was introduced to give a new perspective to sensor network by enabling IPv6 to be applied to wireless sensors as well as wired sensor. Dedicated routing protocols based on 6LoWPAN was soon introduced and Hierarchical Routing Protocol for 6LoWPAN (HiLOW) is one of them. HiLOW clearly defines the routing tree setup process, address allocation technique and the data routing process but there is some shortcomings in terms of transmission power selection. HiLOW does not highlight how the suitable transmission power is being selected for sensor communication purpose and this leads to the assumption that at all time and all scenarios the sensors are using maximum transmission power. In the case the sensors are using maximum transmission power for communication even when it is not necessary then power depletion for sensors will be amplified and the network lifetime will be significantly reduced. In this paper we present a brief introduction to 6LoWPAN, a concise review on HiLOW, a highlight on issues revolving each process in HiLOW and propose a new idea on transmission power selection method for HiLOW.

**Keywords:** *Wireless Sensor Network, Hierarchical Routing, 6LoWPAN, HiLOW, Transmission Power Selection.*

## 1. Introduction

The Wireless Sensor Network (WSN) has been getting much focus from research community in recent years due to its foreseen potential as a tool in solving many problems from day to day problems such as home monitoring [1] up to ecological problems such as mountain side monitoring to detect possible landslides. Initially WSN started a technology used and researched only for military usage for the purpose of detecting enemies, land mines and identifying own man, but WSN has been extended to home monitoring, office monitoring, environmental monitoring and many other areas. In future WSN could exist in every

part of our life due to the engineering contribution. Even though enormous improvement in being observed from the hardware engineering perspective but till today WSN still faces the limitation in power and computational capacities and memory usage[2].

In WSN there are two crucial activities which take place; first is sensing activity done by the sensors then followed by communication between sensors. Radio communication for transmission is typically the most energy consuming activity [3] and the reception energy is often as high as the transmission energy. Thus the network protocol as well as the routing protocol being designed specifically for WSN needs to take into consideration the energy usage in setting up the network as well as complexity of computation during routing. Flaws in routing tree setup could increase the request of retransmission and also cause the node to use more power to reach the parent node or child node to transmit; this will lead towards energy wastage and further shorten the network life and reliability.

IPv6 over Low-Power Wireless Area Network (6LoWPAN) was introduced by Internet Engineering Task Force (IETF) as a standardization effort of IPv6 networking over IEEE 802.15.4. Prior to the introduction of 6LoWPAN many other communication protocols have been introduced to Wireless Sensor Network namely 802.15.1 Bluetooth [4], WirelessHart [5], ZWave [6], ZigBee [6] and others. 6LoWPAN[7] is significant compared to the prior communication protocols as it is the first to introduce IPv6 to be applied not only to wireless but also wired sensor network.

IPv6 was introduced to Low-Power Wireless Area Network compared to IPv4 as IPv6 is the future network. IPv6 was introduced to overcome the address depletion problem in IPv4. Thru introduction of IPv6 other weakness in IPv4 is also handled such as inefficiency of

header processing, lack of standardisation on mobility flow, control, security, M/c and re-configuration. As IPv6 will soon be the addressing scheme as well as the additional benefits it possesses then the low powered sensor network is also introduced with IPv6 capability in contrary to IPv4.

6LoWPAN defines the network layer protocol as well as the transport layer protocol which can be deployed to any IEEE 802.15.4[8, 9] compliant sensors. The 6LoWPAN stack is minimum 30KB in size which is smaller compared to the named protocols above. The routing protocol for 6LoWPAN is an open area for research as it is not specifically defined. As of now, there are three prominent routing protocols which have been designed specifically for 6LoWPAN namely Hierarchical Routing Protocol (HiLOW) [11, 12], Dynamic MANET On-demand for 6LoWPAN (DYMO Low) [13] and 6LoWPAN Ad Hoc On-Demand Distance Vector Routing (LOAD) [14].

The remainder of this paper is organized as follows: Section 2 reviews the processes defined in HiLOW protocol briefly and highlights the issues pertaining the protocol and other works undertaken to improve HiLOW. Section 3 explains in detail the proposed power selection method. Section 4 presents the conclusion.

## 2. HiLOW Protocol and Existing Issues

A hierarchical routing protocol (HiLow) for 6LoWPAN was introduced by K. Kim in 2007 [11]. HiLOW exploits the dynamic 16-bits short address assignment capabilities of 6LoWPAN. HiLOW makes an assumption that the multi-hop routing occurs in the adaptation layer by using the 6LoWPAN Message Format. The operations in HiLOW ranging from the routing tree setup operation up to the route maintenance operation and the issues revolving each operation level will be discussed in the rest of the section.

### 2.1 Hilow Routing Tree Setup, Issues and Other Works done.

The process of setting up the routing tree in HiLOW consists of a sequence of activities. The process is started by a node which tries to locate an existing 6LoWPAN network to join into. The new node will either use active or passive scanning technique to identify the existing 6LoWPAN network in its Personal Operation Space (POS).

If the new node identifies an existing 6LoWPAN it will then find a parent which takes it in as a child node and

obtain a 16 bit short address from the parent. Parent node is a node which is already attached to the network. The parent will assign a 16 bit short address to a child by following the formula as in (1). An important element of HiLOW is that the Maximum Allowed Child (MC) need to be fixed for every network and all the nodes in the network is only able to accept child limited to the set MC. In the case where no 6LoWPAN network is discovered by the node then it will initiate a new 6LoWPAN by becoming the coordinator and assign the short address as 0.

FC : Future Child Node's Address

MC : Maximum Allowed Child Node

N : Number of child node inclusive of the new node.

AP : Address of the Parent Node

$$FC = MC * AP + N \quad (0 < N \leq MC) \quad (1)$$

Three potential issues have been identified in this process. The first issue involving this protocol is that the nodes are assumed to communicate using maximum power transmission. Using maximum power transmission to communicate to its parent's node is not advantageous. This method could lead towards enhanced power drainage of a child node. For example in a scenario that a child communicates with a parent using maximum power transmission (power level 10) even though it could communicate via lower transmission (power level 5) then its power drainage is heightened by nearly 50%. So in this paper we are proposing a power selection method during the routing tree setup by implementing binary search algorithm with LQI value as qualifier. The proposed method is expected to reduce power wastage and heighten the network lifetime.

The second issue would be when the child node gets respond from more than one potential parent. There is no clear mechanism rolled out in selecting the suitable parent to attach with. If the new node chooses to join the first responding parent node, it could be bias to the parent as some parent might be burdened with more parents meanwhile other parents which is in the same level has less child or none at all. Selecting the parent based on first responded potential parent could also lead to fast depletion of energy to certain parent causing the life span of the network to be shorter and the stability to be jeopardized. Selection of parent without considering the link quality could cause towards high retransmission rate which will consume energy from the child node as well as parent node.

In [15] a mechanism to overcome the issue was suggested. Their mechanism suggests the potential parent node to provide the new child with its existing child node count



(child\_number). By issuing the child\_number the node could select suitable parent which has less child nodes. The suggested mechanism performs well only when the potential parent node has same depth, same energy level and has different number of existing child. Their mechanism also does not take into consideration the quality of the link established between the parent node and child node. Therefore the suggested mechanism does not solve the arising issue completely. In order to overcome the weakness in the selection method a comprehensive parent selection method that takes into consideration the link quality, the existing energy of the potential parent as well as the depth of the parent has been proposed in [16]. The paper theoretically discusses how the proposed method could overcome bias child attachment in different scenarios.

Third issue revolves around the MC value which is being fixed for all nodes. The current scenario works well in a homogenous powered sensor environment where all the sensors' power source is the same; for example all is battery powered with same type of battery or all sensors are non-battery powered and having same power source. Meanwhile in a heterogeneous power source sensor environment this method is not advantageous as sensors which are main power and affluent in energy should be assigned with more child compared to battery powered sensor. This is an open issue to be addressed in HiLOW and assumption that all nodes having same energy conservation have to be made. The activity of disseminating the MC value to joining nodes is also left in gray. This issue is not addressed in this paper.

## 2.2 Routing Operation in HiLOW

Sensor nodes in 6LoWPAN can distinguish each other and exchange packet after being assigned the 16 bits short address. HiLOW assumes that all the nodes know its own depth of the routing tree. The receiving intermediate nodes can identify the parent's node address through the defined formula (2). The '[' symbol represents floor operation

AC : Address of Current Node  
 MC : Maximum Allowed Child

$$AP = [(AC-1) / MC] \quad (2)$$

By using the above formula the receiving intermediate nodes can also identify whether it is either an ascendant node or a descendant node of the destination. When a node receives a packet, the node determines the next hop node to forward the packet by following the three cases (3) as

shown in [10]. So far no issues have been identified in this process.

SA : Set of Ascendant nodes of the destination node  
 SD : Set of Descendant nodes of the destination node  
 AA(D,k): The address of the ascendant node of depth D of the node k  
 DC : The depth of current node  
 C : The current node

Case 1: C is the member of SA (3)

The next hop node is AA (DC+1, D)

Case 2: C is the member of SD

The next hop node is AA (DC-1, C)

Case 3: Otherwise

The next hop node is AA (DC-1, C)

## 2.3 Route Maintenance in HiLOW

Each node in HiLOW maintains a neighbor table which contains the information of the parent and the children node. When a node loses an association with its parent, it should to re-associate with its previous parent by utilizing the information in its neighbor table. In the case of the association with the parent node cannot be recovered due to situation such as parent nodes battery drained, nodes mobility, malfunction and so on, the node should try to associate with new parent in its POS [11]. Meanwhile if the current node realizes that the next-hop node regardless whether its child or parent node is not accessible for some reason, the node shall try to recover the path or to report this forwarding error to the source of the packet.

Even though a route maintenance mechanism has been defined in HiLOW, the mechanism is seen as not sufficient to maintain the routing tree. An Extended Hierarchical Routing Over 6LoWPAN which extends HiLOW was presented by in [16] in order to have better maintained routing tree. They suggested two additional fields to be added to the existing routing table of HiLOW namely, Neighbour\_Replace\_Parent (NRP) and Neighbour\_Added\_Child (NAC). This NRP doesn't point to the current parent node but to another node which can be its parent if association to current parent fails. Meanwhile NAC refers to the newly added child node. More work need to be done on this mechanism on how many nodes allowed to be adapted by a parent node in

addition to the defined MC and whether this mechanism will have any impact on the routing operation, however this topic is beyond the scope of this paper.

### 3. Transmission Power Level Selection Method for HiLOW

A transmission power level selection method by implementing binary search algorithm coupled with maximum search round and LQI value as qualifier is being presented in this paper. The suggested method is able to reduce number of nodes communicating using maximum transmission power with its parent node, by doing so the energy used in transmission is reduced and network lifetime is heightened.

Binary search method coupled with maximum is selected compared to incremental search or pure binary search in order to reduce the number of rounds the nodes undergoes to search for parent. Table 1 displays the number of maximum search rounds in worst case scenario which is possible based on the different number of power levels for three different searches. From the table it can be easily deduced that Binary Search Algorithm is more efficient in worse case scenarios. Meanwhile the mechanism suggested in this paper ensures that the number of search is even more limited as the energy is very crucial for sensor nodes.

An assumption that all the nodes have mapped their Tx Power Setting to output power and the Tx Power setting is incremental by 1 from each other, for example as set by default in Atmel Raven Nodes as shown in Table 2. An assumption that different power level setting uses different battery consumption is also made for example in Atmel Raven when the output Power is 0dBm the amount of battery power is quoted to be less than 13mA and in the case of full output power ( $\approx -17$  dBm) the battery consumed battery power is 16-17mA.

Table 1: Maximum Search Rounds in worst case scenario for three different type of search method

Power Level (N)	Incremental / Linear Search	Binary Search ( $\log_2 N$ )	Suggested Search (MR =4)
5	5	3	3
10	10	4	4
20	20	5	4
30	30	5	4
40	40	6	4

The transmission power selection process starts when a node starts when the node looking to join a network in it

POS as shown in Fig. 1. Before starting a scan the node needs to determine and the Lowest Transmit Power (LP) and Highest Transmit Power (HP). The node then sets the Optimum Transmit Power (OP) value to be equivalent to HP. Then the Search Transmit Power (SP) value has to be determined. The SP value is determined following mathematical equation in (4). The Current Search (CR) value is also set to 0.

Table 2: Default Power Mapping in Atmel Raven Sensor Nodes [19]

TX Power Setting	Output Power[dBm]
0	3
1	2.6
2	2.1
3	1.6
4	1.1
5	0.5
6	-0.2
7	-1.2
8	-2.2
9	-3.2
10	-4.2
11	-5.2
12	-7.2
13	-9.2
14	-12.2
15	-17.2

Two values are to be set during compile time; one is Maximum Search Round (MR), the MR value has to be set for all nodes and the value should be same for all nodes. MR is basically the number of maximum search round the nodes can go through before they terminate the search. Second value is the accepted LQI value.

SP : Search Transmit Power Level

HP : Highest Transmit Power Level

LP : Lowest Transmit Power Level

$$SP = [ ( (HP - LP) - 1) / 2 ] \quad (4)$$

The node then will use the SP to search for the potential parent.

In the case 1: Where the node does not find any potential parent it will set the LP value to be SP value.

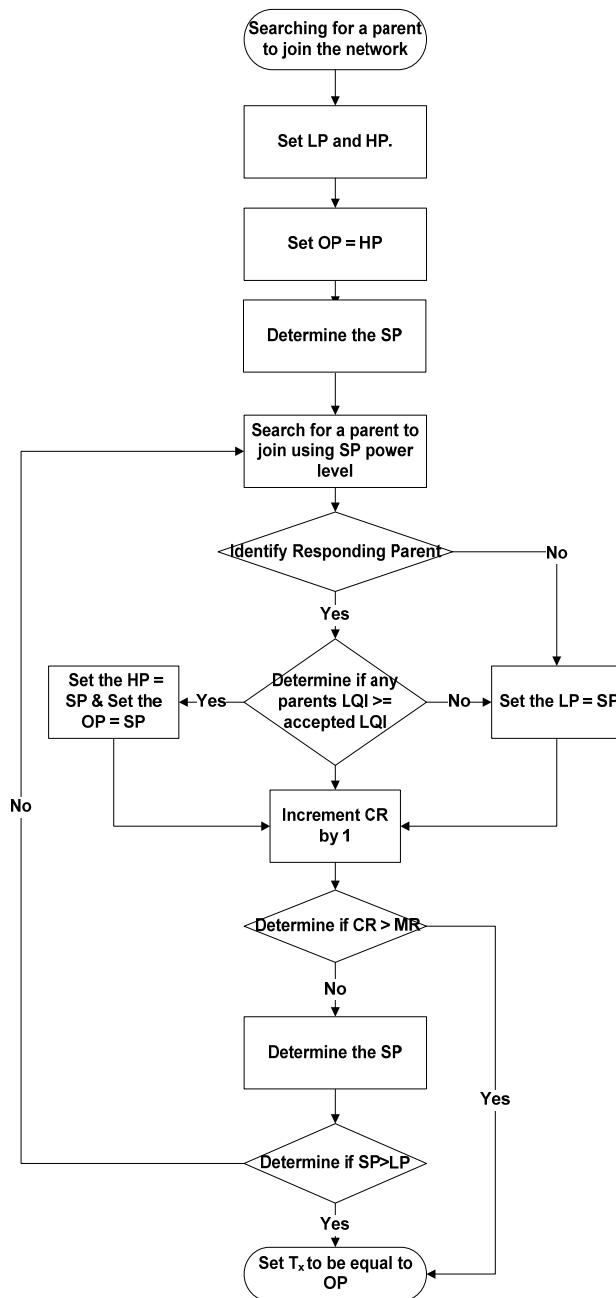


Fig. 1 Proposed power level selection method for HiLOW

In the case 2: Where the node does find a potential parent it will compare the LQI value with the accepted LQI value. In the case where the LQI value is more than accepted LQI then the HP will be set to equal with SP and OP will also be set to equal SP. In the other case the LP value will be set to SP value and the OP value remain unchanged.

Regardless which ever case the node encountered, the node will then continue to the same process which is increment CR by 1, then determine if the CR more than MR. If the condition is true then the node terminates the search process and set the transmission power level to be equivalent to OP. In the case the condition is not true then the SP is again determined, then the new SP is compared with the LP to ensure that is higher than LP if it is not then the process is also terminated and the transmission power level is set to be equivalent to OP. In the case the condition is true then the process loops back to the process for a parent using the SP power level.

#### 4. Conclusions

In this paper review on HiLOW, issues revolving each process in HiLOW and other works done in this area are presented. A new idea on transmission power level selection method by implementing binary search algorithm coupled with maximum search round and LQI value as qualifier is presented in this paper. The presented power level selection method is believed to be able to overcome the problem of maximum power usage for every transmission; by which the network lifetime could be increased. The presented power selection method is also better than linear search method and pure binary search method as discussed in our paper as it has it exits search in fixed number of rounds compared to the latter. Even though the method is suggested for HiLOW, the method could be easily adapted to other type of hierarchical routing. Our future research will be focused on validating the suggested mechanism as well as adapting it to other routing protocols such as LEACH.

#### Acknowledgments

The author would like to acknowledge Universiti Sains Malaysia (USM) for funding of USM Fellowship Scheme 2009/10.

#### References

- [1] [1] Lee, S.hyun. & Kim Mi Na., "Wireless Sensors for Home Monitoring - A Review", Recent Patents on Electrical Engineering, 2008, Vol. 1, No. 1, pp32-39.
- [2] Ian F.Akyildiz, Weilian Su, Yogesh Sankarasubramaniam and Erdal Cayirci, "A Survey on Sensor Networks", Communication Magazine, IEEE, Volume 40, 2002.
- [3] A. Dunkels, F. "Osterlind, N. Tsiftes, and Z. He, "Software-based online energy estimation for sensor node ", Proceedings of the Fourth IEEE Workshop on Embedded Networked Sensors (Emnets IV), Cork Ireland, 2007.
- [4] [2] L.Martin, B.D Mads, B.Philippe, (2003) "Bluetooth and sensor networks: a reality check", *Proceedings of the 1st international conference on Embedded networked sensor systems, 2003*

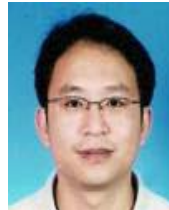
- [5] S.Jianping, et al., "WirelessHART: Applying Wireless Technology in Real-Time Industrial Process Control", In Proceedings of IEEE Real-Time and Embedded Technology and Applications Symposium, 2008.
- [6] B.Chiara, C.Andrea, D.Davide, V.Roberto, "An Overview on Wireless Sensor Networks Technology and Evolution", Sensors 2009, Sensors 2009, 9, 6869-6896; doi:10.3390/s90906869
- [7] N. Kushalnagar, et al., "Transmission of IPv6 Packets over IEEE 802.15.4 Networks", rfc4944, September 2007.
- [8] IEEE Computer Society, "802.15.4-2006 IEEE Standard for Information Technology- Telecommunications and Information Exchange Between Systems- Local and Metropolitan Area Networks- Specific Requirements Part 15.4: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks (WPANs)"
- [9] K. Kim, S.Yoo, S.Daniel, J.Lee, G.Mulligan, "Problem Statement and Requirements for 6LoWPAN Routing", draft-ietf-6lowpan-routing-requirements-02, March 2009
- [10] K. Kim, S.Yoo, S.Daniel, J.Lee, G.Mulligan, "Commissioning in 6LoWPAN", draft-6lowpan-commissioning-02, July 2008
- [11] K. Kim, et al., "Hierarchical Routing over 6LoWPAN (HiLOW)", draft-daniel-6lowpan-hilow-hierarchical-routing-01, June 2007.
- [12] K. Kim, et al., "Hierarchical Routing over 6LoWPAN (HiLOW)", draft-daniel-6lowpan-hilow-hierarchical-routing-00, June 2005.
- [13] K. Kim, G.Montenegro, S.Park, I.Chakeres, C.Perkins, "Dynamic MANET On-demand for 6LoWPAN (DYMO-low) Routing", draft-montenegro-6lowpan-dymo-low-routing-03, June 2007.
- [14] K.Kim, S.Daniel, G.Montenegro, S.Yoo, N.Kushalnagar, "6LoWPAN Ad Hoc On-Demand Distance Vector Routing (LOAD)", draft-daniel-6lowpan-load-adhoc-routing-02, March 2006
- [15] K.Kim, S.Daniel, G.Montenegro, S.Yoo, N.Kushalnagar, "6LoWPAN Ad Hoc On-Demand Distance Vector Routing (LOAD)", draft-daniel-6lowpan-load-adhoc-routing-02, March 2006
- [16] Hun-Jung-Lim, Tai-Myoung Chung, "The Bias Routing Tree Avoiding Technique for Hierarchical Routing Protocol over 6LoWPAN", 2009 Fifth International Joint Conference on INC, IMS and IDC.
- [17] C.Nam, H.Jeong, D.Shin, "Extended Hierarchical Routing Protocol over 6LoWPAN", MCM2008, September 2008.
- [18] C.Lingeswari et al., "Bias Child Node Association Avoidance Mechanism for Hierarchical Routing Protocol in 6LoWPAN", Proceedings of the Third IEEE International Conference on Computer Science and Information Technology
- [19] AVR2002 : Raven Radio Evaluation Software
- [20] Zhu Jian, Zhao Lai, "A Link Quality Evaluation Model in Wireless Sensor Networks", Proceedings of the 2009 Third International Conference on Sensor Technologies and Applications



**Lingeswari V.Chandra** obtained her BIT with Management degree from AIMST University in 2008. She is the university gold medalist. She obtained her software engineering foundation training from Infosys, Bangalore. She is currently pursuing her PhD in National Advanced IPv6 Center, Universiti Sains Malaysia. Her research interest is in Wireless Sensor Network particularly in Hierarchical Routing.



Mr. Selvakumar Manickam obtained his Bachelor of Computer Science and Master of Computer Science from Universiti Sains Malaysia in 1999 and 2003 respectively. He is a lecturer and domain head of industrial & community linkages of the National Advanced IPv6 Centre of Excellence (NAV6) in Universiti Sains Malaysia. His research areas are information architecture, network technology and management as well as IPv6 in Bioinformatics.



**Kok-Soon Chai** is a certified Project Management Professional by Project Management Institute, USA. He received his MSc and Ph.D. (2003) degrees from the University of Warwick, UK. He worked for more than seven years as a senior R&D software engineer, embedded software manager, and CTO at Motorola, Agilent, Plexus Corp., Wind River in Singapore (now a division of Intel Corp.), and NeoMeridian. He holds one US patent, with two US patents pending. His main interests are wired and wireless sensor networks, green technology, embedded systems, consumer electronics, and real-time operating systems. Dr. Chai is a senior lecturer at the National Advanced IPv6 Centre of Excellence (NAV6) in Universiti Sains Malaysia



**Sureswaran Ramadass** obtained his BsEE/CE (Magna Cum Laude) and Master's in Electrical and Computer Engineering from the University of Miami in 1987 and 1990, respectively. He obtained his Ph.D. from Universiti Sains Malaysia (USM) in 2000 while serving as a full-time faculty in the School of Computer Sciences. Dr. Sureswaran Ramadass is a Professor and the Director of the National Advanced IPv6 Centre of Excellence (NAV6) in Universiti Sains Malaysia.

# Setting up of an Open Source based Private Cloud

Dr.G.R.Karpagam<sup>1</sup>, J.Parkavi<sup>2</sup>

<sup>1</sup>Professor, Department of Computer Science and Engineering,  
PSG College of Technology, Coimbatore-641 004, India

<sup>2</sup>ME, Software Engineering,  
Department of Computer Science and Engineering,  
PSG College of Technology, Coimbatore-641 004, India

## Abstract

Cloud Computing is an attractive concept in IT field, since it allows the resources to be provisioned according to the user needs[11]. It provides services on virtual machines whereby the user can share resources, software and other devices on demand. Cloud services are supported both by Proprietary and Open Source Systems. As Proprietary products are very expensive, customers are not allowed to experiment on their product and security is a major issue in it, Open source systems helps in solving out these problems. Cloud Computing motivated many academic and non academic members to develop Open Source Cloud Setup, here the users are allowed to study the source code and experiment it. This paper describes the configuration of a private cloud using Eucalyptus. Eucalyptus an open source system has been used to implement a private cloud using the hardware and software without making any modification to it and provide various types of services to the cloud computing environment.

**Keywords:** *Cloud Computing, Open Source, Private Cloud.*

## 1. Introduction

Cloud computing is a computing environment, where resources such as computing power, storage, network and software are abstracted and provided as services on the internet in a remotely accessible fashion. Billing models for these services are generally similar to the ones adopted for public utilities. On-demand availability, ease of provisioning, dynamic and virtually infinite scalability is some of the key attributes of cloud computing [6].

The main concept behind cloud computing is providing services. It provides various types of services, some of the important services are SaaS, PaaS and IaaS. Software as a service is a model of software deployment whereby a provider licenses an application to customers for use as a service on demand. Platform as a service generates all facilities required to support the complete cycle of construction and delivery of web-based applications wholly available in Internet without the need of downloading software or special installations by

developers and finally Infrastructure as a service provides informatics resources, such as servers, connections, storage and other necessary tools to construct an application design prepared to meet different needs of multiple organizations, making it quick, easy and economically viable [4].

Cloud computing is mainly classified into three types based on the deployment model; Public cloud, Private cloud and Hybrid cloud. If the services are provided over the internet then it is public cloud or external cloud and if it is provided within an organization through intranet then it is named as private cloud or internal cloud and Hybrid cloud is an internal/external cloud which allows a public cloud to interact with the clients but keep their data secured within a private cloud [7].

This paper explains about EUCALYPTUS: an open-source system that enables the organization to establish its own cloud computing environment. Eucalyptus is structured by various components which interact with each other through well-defined interfaces. It is used for implementing on-premise private and hybrid clouds using the hardware and software infrastructure that is in place, without modification.

## 2. Eucalyptus

Eucalyptus (Elastic Utility Computing Architecture for Linking Your Programs To Useful Systems) was released in May 2008, creator of the leading Open-Source Private Cloud platform. They were incorporated as an organization in January 2009 Headquartered in Santa Barbara, California.

Eucalyptus software is available under GPL (General Public License) that helps in creating and managing a private or even a publicly accessible cloud. It provides an EC2 (Elastic Compute Cloud)-compatible cloud computing platform and S3 (Simple Storage Service)-compatible cloud storage platform.



Eucalyptus is one of the key for open source cloud platforms which makes it much popular. The client tools used for Eucalyptus is same as that of AWS, because Eucalyptus services are available through EC2/S3 compatible APIs [6].

### 2.1 Amazon AWS Compatibility

API compatibility layer is build on top of Eucalyptus that explores the functionality in terms of Amazon’s API. Amazon tools, infrastructure and other work that someone put into building for Amazon would also be compatible with Eucalyptus inside the datacenter. In Fig.1 we can see various components of Amazon and Eucalyptus. The EC2 (Elastic Compute) component of Amazon which handles the provisioning of virtual machine and its resources are replaced here with cloud controller similarly Amazon provides storage mechanism EBS (Elastic Block Storage) which provides block storage devices to virtual machines are replaced by Storage Controller and S3 (Simple Storage System) simple object based get put mechanism, here it is implemented as walrus.

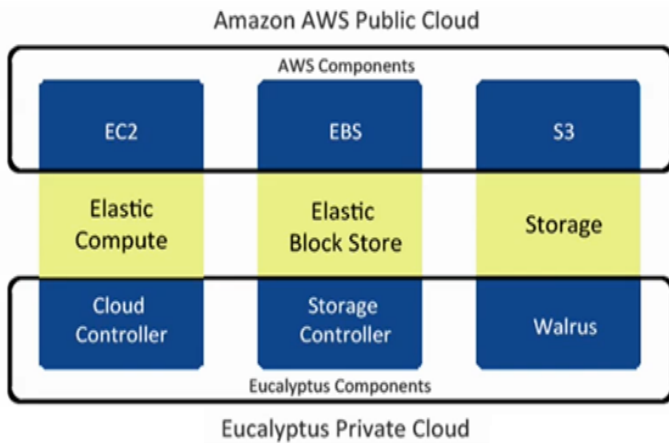


Fig.1. Services from AWS and Eucalyptus [2].

### 3. UEC Architecture

Ubuntu Enterprise Cloud UEC, is a private cloud set up for developing its our own IT infrastructure. UEC comes up with many open source software and Eucalyptus is one among them and it makes the installation and configuration of the cloud easier. Canonical also provides commercial technical support for UEC. The basic architecture of UEC consists of A front end which runs one or more Cloud Controller (CLC), Cluster Controller (CC), Walrus (WS3), Storage Controller (SC) and One or more nodes[6]. The architecture of UEC is shown in Fig 2. A CLC manages the whole cloud and includes multiple CC’s. There will be a WS3 attached to a CLC. A CC can contain multiple NC’s and SC’s. Ultimately the VM’s will be running in the NC making use of its physical resources [5].

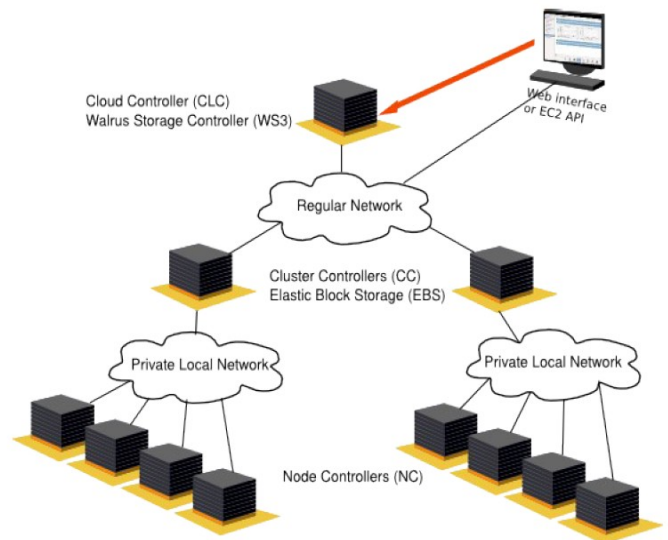


Fig.2. Architecture of Ubuntu Enterprise Cloud [13].

### 4. Building a private cloud

Private Cloud is also called an internal cloud which is mainly designed to control the data of an organization, than by getting the resources from other hosted services [12].

This section describes about the basic installation and configuration of Ubuntu Enterprise Cloud as well as the steps for creating a virtual machine image and uploading the image to the private cloud.

#### 4.1 Installation and Configuration

The UEC setup in Fig.3. Includes two servers (Server 1 and Server 2) which will run a Lucid 64-bit server version and the third system which will run a Lucid Desktop 64-bit version (Client 1) [6] [8].

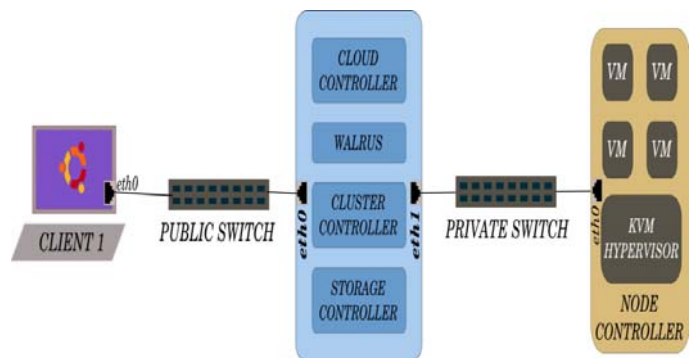


Fig.3.UEC basic setup with Three Machines [6].

## 5. Steps in Configuring an Open Source Private Cloud

Steps	Description/commands
<b>Installation Procedure for Server 1</b>	
Install Ubuntu Server 10.04 CD in Server 1	<b>Boot the Server off for installation</b>
Setup the IP address details.	<b>192.168.4.145. ( Please do that for eth0)</b>
Cloud Controller Address	<b>Leave this blank as Server1 is the Cloud Controller in this setup</b>
Cloud Installation Mode	<b>Select "Cloud controller", "Walrus storage service", "Cluster controller" and "Storage controller".</b>
Network interface for communication	<b>Select eth1 node</b>
Eucalyptus cluster name	<b>Cluster 1</b>
Eucalyptus IP range	<b>192.168.4.155-192.168.4.165</b>
<b>Installation Procedure for Server 2</b>	
Install Ubuntu Server 10.04 CD in Server 2	<b>Boot the Server off for installation</b>
Setup the IP address for one interface	<b>Please do that for eth0 by setting up the private IP - 192.168.4.146</b>
Cloud Controller Address	<b>192.168.4.145</b>
Cloud Installation Mode	<b>Select "Node Controller"</b>
Gateway	<b>192.168.4.145 (IP of the CC)</b>
<b>Installation Procedure for Client 1</b>	
Install Ubuntu Desktop 10.04 CD in Client	<b>Boot the Desktop off for installation</b>
IP Address	<b>The Desktop will be on the enterprise network and will obtain an IP address through DHCP</b>
Install KVM	<b>To help us to install images on KVM platform and bundle them</b>

<b>Invoke the Web Interface</b>	
Login to the web interface of CLC	<a href="https://192.168.4.145:8443/">https://192.168.4.145:8443/</a> The default username is "admin" and the default password is "admin".
Download the credentials	From <a href="https://192.168.4.145:8443/#credentials">https://192.168.4.145:8443/#credentials</a> and save it in the ~/.euca directory
Extract the credentials archive	<code>\$ cd .euca</code> <code>\$ unzip mycreds.zip</code>
Source eucarc	<code>\$ . ~/.euca/eucarc</code>
Verify euca2ools communication with UEC	<code>\$ euca-describe-availability-zones verbose</code>
<b>Running Instances</b>	
Installing Images	From Canonical over the internet (no proxy), check Store tab.
Checking the available Images	<code>\$ euca-describe-images</code>
Installing a Keypair	<code>\$ euca-add-keypair mykey &gt; ~/.euca/mykey.priv</code> <code>\$ chmod 0600 ~/.euca/mykey.priv</code>
Running an Instance ( using terminals)	<code>\$ euca-run-instances -g Ubuntu 9.10 -k mykey -t c1.medium emi-E08810 7E</code>
Hybridfox	Used to run the instances using GUI
Life cycle of an Instance	<b>Pending - Running - Shutting down - Terminated - Reboot.</b> <code>\$ euca-run-instances</code> <code>\$ euca-terminate-instances</code> <code>\$ euca-reboot-instances</code>

Table.1. Configuration Steps

## 6. ALGORITHM

### 6.1 Installing server1

1. Boot the server off the Ubuntu Server 10.04 CD. At the graphical boot menu, select "Install Ubuntu Enterprise Cloud" and proceed with the basic installation steps.
2. Installation only lets you set up the IP address details for one interface. Please do that for eth0.
3. We need to choose certain configuration options for UEC, during the course of the install.
4. Cloud Controller Address - Leave this blank as Server1 is the Cloud Controller in this setup.
5. Cloud Installation Mode - Select "Cloud controller", "Walrus storage service", "Cluster controller" and "Storage controller".

6. Network interface for communication with nodes - eth1
7. Eucalyptus cluster name – cluster1
8. Eucalyptus IP range - 192.168.4.155-192.168.4.165 [6].

## 6.2 Installing server 2

1. Boot the server off the Ubuntu Server 10.04 CD. At the graphical boot menu, select “Install Ubuntu Enterprise Cloud” and proceed with the basic installation steps.
2. Installation only lets us to set up the IP address for one interface. Please do that for eth0 by setting up the private IP - 192.168.4.146.
3. Then choose certain configuration options for UEC, during the course of the install. Ignore all the settings, except the following:
4. Cloud Controller Address - 192.168.4.145
5. Cloud Installation Mode - Select “Node Controller”
6. Gateway - 192.168.4.145 (IP of the CC) [6].

## 6.3 Installing Client 1

The purpose of Client1 machine is to interact with the cloud setup, for bundling and registering new Eucalyptus Machine Images (EMI).

1. Boot the Desktop off the Ubuntu Desktop 10.04 CD and install. The Desktop will be on the enterprise network and will obtain an IP address through DHCP.
2. Install KVM to help us to install images on KVM platform and bundle them:  
\$apt\_get install qemu\_kvm [6].

## 6.4 Algorithm for Invoking the Web Interface

1. Login to the web interface of CLC by using the following link <https://192.168.4.145:8443>. The default username is “admin” and the default password is “admin”.
2. Note that the installation of UEC installs a self signed certificate for the web server. The browser will warn us about the certificate not having been signed by a trusted certifying authority. Authorize the browser to access the server with the self signed certificate.
3. When you login for the first time, the web interface prompts to change the password and provide the email ID of the admin. After completing this mandatory step, download the credentials archive from <https://192.168.4.145:8443/#credentials> and save it in the ~/.euca directory.
4. Extract the credentials archive:  
\$ cd .euca  
\$ unzip mycreds.zip
5. Source eucarc script to make sure that the environmental variables used by euca2ools are set properly.  
\$ . ~/.euca/eucarc

6. To verify that euca2ools are able to communicate with the UEC, try fetching the local cluster availability details shown in Fig.4.

\$ euca-describe-availability-zones verbose

```
cloud@eucalyptus:~$ cd .euca
cloud@eucalyptus:~/.euca$ source eucarc
cloud@eucalyptus:~/.euca$ euca-describe-availability-zones verbose
AVAILABILITYZONE      cluster1      192.168.4.145
AVAILABILITYZONE      |- vm types   free / max    cpu  ram  disk
AVAILABILITYZONE      |- m1.small   0002 / 0002   1    192  2
AVAILABILITYZONE      |- c1.medium  0002 / 0002   1    256  5
AVAILABILITYZONE      |- m1.large   0001 / 0001   2    512  10
AVAILABILITYZONE      |- m1.xlarge  0001 / 0001   2   1024 20
AVAILABILITYZONE      |- c1.xlarge  0000 / 0000   4   2048 20
cloud@eucalyptus:~/.euca$
```

Fig .4 Snapshot for list of Available Resources

7. If the free/max VCPUs are set as 0 in the above list, it means that the node did not get registered automatically. Use the following on Server1 and approve when prompted to add 192.168.4.146 as the Node Controller:

\$sudo euca\_conf --discover-nodes [6].

## 7. Running Instances

### 7.1 Installing Cloud Images

No images exist by default in the Store (web Interface). Running an instance or VM in the cloud is only based on image. Images can be installed directly from Canonical online cloud image store or we can also build custom image, bundle it, upload and register them with the cloud. The “Store” tab in the web interface will show the list of images that are available from Canonical over the internet [6].

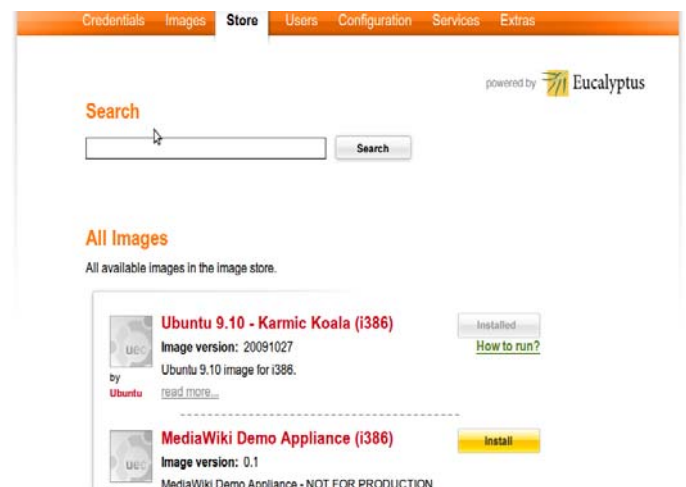


Fig.5. List of Images from Store

## 7.2 Checking Images

“euca-describe-images” is the command-line equivalent of clicking the “Images” tab in the Eucalyptus administrative web interface. This shows the emi-xxxxxx identifier for each image/bundle that will be used to run an instance.

```
$ euca-describe-images
```

```
IMAGE emi-E088107E image-store-1276733586/ image.
manifest.xml admin
available public x86_64machine eki-F6DD1103 eri-0B3E1166
IMAGE eri-0B3E1166 image-store-1276733586/ ramdisk.
manifest.xml admin
available public x86_64ramdisk
IMAGE eki-F6DD1103 image-store- 1276733586/ kernel.
manifest.xml admin
available public x86_64kernel
```

## 7.3 Installing a Keypair

Build a keypair that will be injected into the instance allowing us to access it via ssh:

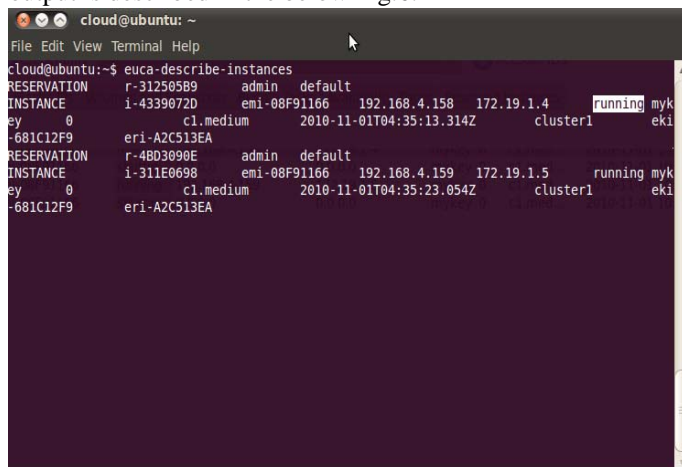
```
$ euca-add-keypair mykey > ~/.euca/mykey.priv
$ chmod 0600 ~/.euca/mykey.priv [6]
```

## 7.4 Running the Instances

1. Now we are finally ready to begin running instances. We'll start by creating an instance of our image and connections will be allowed on ports ssh and http:

```
$ euca-run-instances -g Ubuntu 9.10 -k mykey -t c1.medium
emi-E088107E
```

2. After issuing the “euca-run-instances” command to run an instance, we can track its progress from pending to running state by using the euca-describe-instances command and the output is described in the below Fig.6.



```
cloud@ubuntu: ~
File Edit View Terminal Help
cloud@ubuntu:~$ euca-describe-instances
RESERVATION r-312505B9 admin default
INSTANCE i-4339072D emi-08F91166 192.168.4.158 172.19.1.4 running myk
ey 0 c1.medium 2010-11-01T04:35:13.314Z cluster1 eki
-681C12F9 eri-A2C513EA
RESERVATION r-4B03090E admin default
INSTANCE i-311E0698 emi-08F91166 192.168.4.159 172.19.1.5 running myk
ey 0 c1.medium 2010-11-01T04:35:23.054Z cluster1 eki
-681C12F9 eri-A2C513EA
```

Fig.6. Snapshot of Running Instances

## 7.5 Hybridfox

Hybridfox provide compatibility between Amazon Public cloud and Eucalyptus Private Cloud [9]. Hybridfox tool is a modified or extended elasticfox that enables us to switch seamless between different cloud clusters in order to manage the overall cloud computing environment. Hybridfox can perform all the functions that can be done by elasticfox, on the Eucalyptus Computing environment like Manage Images, Raise and Stop Instances, Manage Instances, Manage Elastic IPs, Manage Security Groups, Manage Key pairs and Manage Elastic Block Storage[3].Running a different instance by using Hybridfox is shown below in the Fig.6.

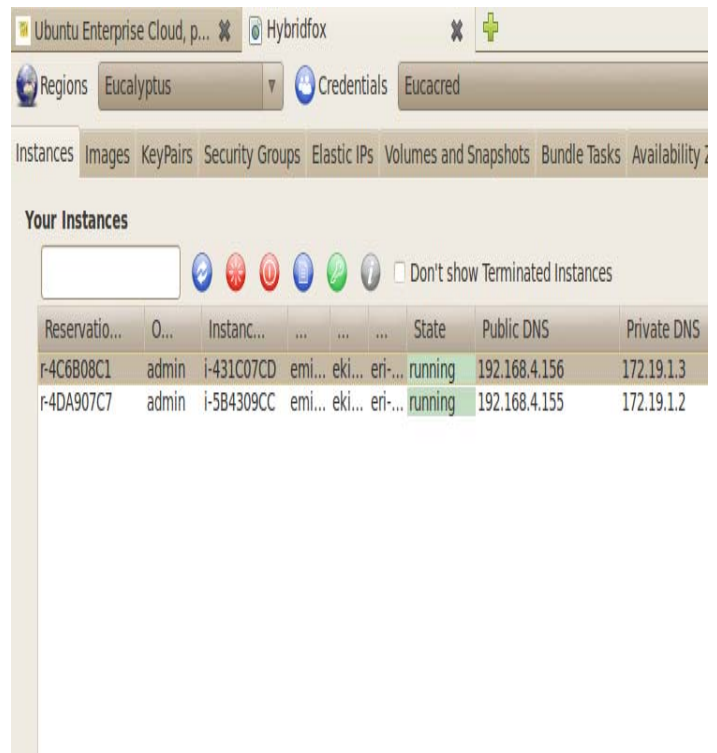


Fig.7. Running the Instance by Hybridfox

## 7.6 Life cycle of an Instance

When “euca-run-instances” command is invoked (or when run instance is chosen from Hybridfox/Elasticfox), the running process will be in a sequential manner as shown in Fig.7. Here are some few things that happen on various components of UEC:

1. Authentication/Authorization of the user request to ensure that we have permission to launch the instance
2. Identification of CC to take responsibility for deploying the instance and identification of the NC for running the instance.
3. Downloading the image from WS3 to NC (images are cached so that starting multiple instances of the same machine image downloads that image only once) [6]



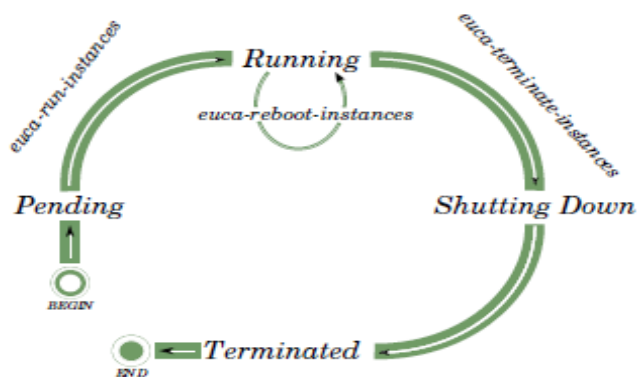


Fig.7. Life Cycle of an Instance [6].

## 8. Future Scope

### Types of Services

A cloud can provide service either to private or public cloud. In public cloud, based on demand the services are provided to the client and in a private cloud the service is provided to a single client [10]. The combination of both public and private cloud is called hybrid private cloud, here the private cloud is hosted in a public cloud. Services that are included to the cloud setup are listed in Table.2.

Apache Service	Software as a Service
GNU C++ Service	Compiler as a Service
IDE as a Service	Software as a Service
Search Engine	Search Engine as a Service

Table.2. List of Services

### 8.1 Web Service

A user can access a web page from any computer connected to the cloud by using Apache web server. Install the Apache web server in the instance and get accessed to the service.

```
$sudo apt-get install apache2
```

### 8.2 Compiler as a Service

This service is provided to compile the c++file. Even if the client doesn't have the compiler, it can be compiled with the compiler available from cloud. The user is ssh'ed to the instance with certain privileges and allowed to compile and see the result.

```
$ssh cloud@<IP Address>
```

Installing gnu c++ compiler in an instance:

```
$sudo apt-get install build-essential
```

### 8.3 Other Services

IDE as a service can be obtained by installing Apache Tomcat6 and Search Engine as a service can be achieved by accessing through the Web Service.

## 9. Conclusion

Cloud computing is an everlasting computing environment where data are delivered on-demand to authenticated devices in a secured manner and users utilize a shared and elastic Infrastructure. This paper briefly explains the set up of a private cloud in a cluster based environment using open source technologies like Eucalyptus, KVM, and euca2ools. The virtual machine images are available in the cloud and upon user request; its instances are created and run. Services were included successfully and made available to the user. The current implementation of this paper provides Infrastructure as a service (IaaS) and Software as a Service (SaaS).

## References

- [1] Cloud Computing (2010), Wikipedia;en.wikipedia.org/wiki/
- [2] Dr. Rich Wolski, (2010) Enterprise Cloud Control.
- [3] Ezhil Arasan Babaraj, (2009), Driving Technology Direction on Cloud Computing Platform, Blog post; Hybridfox: Cross of Elasticfox and Imagination, ezhil.sys-con.com/.
- [4] Glossary, (2010), MasterBase, www.en.masterbase.com/support/glossary.asp.
- [5] Installing the Eucalyptus Cloud/Cluster/Storage Node on Ubuntu Karmic 9.10 dustinkirkland, www.YouTube.com
- [6] Johnson D, Kiran Murari, Murthy Raju, Suseendran RB, Yogesh Girikumar (2010), Eucalyptus Beginner's Guide - UEC Edition, CSS Open Source Services, UEC Guide.v1.0. (Ubuntu Server 10.04 - Lucid Lynx).
- [7] Judith H, Robin B, Marcia K, and Dr. Fern H, Dummies.com, Comparing-Public-Private-and-Hybrid-cloud- computing. Wiley Publishing, Inc.2009.
- [8] Kefa Rabah, (2010) Build Your Own Private Cloud Using Ubuntu 10.04 Eucalyptus Enterprise Cloud Computing Platform v1.2.
- [9] Mitchell pronsc, (2009) Hybridfox: Elasticfox for Eucalyptus.
- [10] Partha Saradhi K S (2010), Types of Cloud Computing Services, Information Security.
- [11] Patrícia T Endo, Glauco E Gonçalves, Judith K, Djamel S (2010), A Survey on Open-source Cloud Computing Solutions, VIII Workshop em Clouds, Grids e Aplicações, pp. 3-16.
- [12] Private cloud, (2008) SearchCloudComputing.com, Definitions; Whatls.com
- [13] Simon Wardley, Etienne Goyer & Nick Barcet, (2009), CANONICAL ,Technical White Paper, Ubuntu Enterprise Cloud Architecture.



# Real-Time Strategy Experience Exchanger Model [Real-See]

Mostafa Aref<sup>1</sup>, Magdy Zakaria<sup>2</sup> and Shahenda Sarhan<sup>3</sup>

<sup>1</sup> Faculty of Computers and Information, Ain-Shams University  
Ain-Shams, Cairo, Egypt

<sup>2</sup> Faculty of Computers and Information, Mansoura University  
Mansoura, Egypt

<sup>3</sup> Faculty of Computers and Information, Mansoura University  
Mansoura, Egypt

## Abstract

For many years, researchers tried and succeeded to develop agents that can adapt their behavior to face new opponent scenarios and beating them. So in this paper we introduce an experience exchanging model that allow a game engine to update all other engines with the game reaction against new surprising un-programmed opponent scenarios that face the computer player through exchanging new cases among engines case-based reasoning systems. We believe this will reveal game players from downloading a new engine of the game and loosing their saved episodes.

**Keywords:** *Real-Time Strategy Games, Case-based Reasoning, Feature Similarity.*

## 1. Introduction

Artificial Intelligence (AI) [2][4][18] is the area of computer science focusing on creating intelligent machines. The ability to create intelligent machines has intrigued humans since ancient times. Today with the advent of the computer and 60 years of research into AI programming techniques, the dream of smart machines is becoming a reality.

Researchers are creating systems as intelligent agents that can autonomously decide about the desired results without user interaction, script or even fixed execution plan. They can mimic human thought, understand speech and beat the best human chess-player. This has two benefits, first, they allow for a high-level definition of

the problem. Secondly, agents are better reusable and more robust than fixed programs. These benefits make agents a suitable area for computer AI games.

AI games has existed since 1951 when Christopher Strachey wrote a checkers program [16][18]. As 3D rendering [16] hardware and resolution quality of game graphics improved, AI games had increasingly become one of the critical factors determining a game's success. From this we can refer to AI games as techniques used in computer and video games to produce the illusion of intelligence [16][18] in the behavior of non-player characters (NPCs). While the non-player character is a character that is controlled by the game master so it is a part of the program, but not controlled by a human.

The real-time performance requirements of computer AI games, the demand for humanlike interactions [5], appropriate animation sequences, and internal state simulations for populations of scripted agents have impressively demonstrated the potential of academic AI research and AI games technologies.

## 2. Background

### 2.1 Real-Time Strategy Games

A real-time strategy game (RTS) is a strategic war [5][9] game in which multiple players operate on a virtual battlefield, controlling bases and armies of military units. It typically ends with the destruction of the enemy.

The better balance you get among economy, technology, and army, the more chances you have to win.

Although many studies exist on learning to win games with comparatively small search spaces, few studies exist on learning to win complex strategy games. Some researchers argued that agents require sophisticated representations and reasoning abilities to perform well in these environments, so they are challenging to construct.

Fortunately, Ponsen and Spronck (2004) [14] developed a good representation for WARGUS, a moderately complex RTS game. They also employed a high-level language for game agent actions to reduce the decision space. Together, these constrain the search space of useful plans and state-specific sub-plans, allowing them to focus on the performance task of winning RTS games.

Marthi, Russell, and Latham (2005) [11] applied hierarchical reinforcement learning (RL) in a limited RTS domain. This approach used reinforcement learning augmented with prior knowledge about the high-level structure of behavior, constraining the possibilities of the learning agent and thus greatly reducing the search space.

Ponsen, Muñoz-Avila, Spronck and Aha (2006) [12] introduced the Evolutionary State-based Tactics Generator (ESTG), which focuses on the highly complex learning task of winning complete RTS games and not only specific restrained scenarios.

## 2.2 Case-based Reasoning

Case-based Reasoning (CBR) is a plausible generic model of an intelligence and cognitive science-based method by the fact that it is a method for solving problems by making use of previous, similar situations and reusing information and knowledge about such situations. CBR [13] combines a cognitive model describing how people use and reason from past experience with a technology for finding and presenting such experience. The processes involved in CBR can be represented by a schematic cycle as shown in figure (1).

1. **Retrieval** is the process of finding the cases in the case-base that most closely match the current information known (new case) [1][8].
2. **Reuse** is the step where [1] matching cases are compared to the new case to form a suggested solution.
3. **Revision** is the testing of the suggested [8] solution to make sure it is suitable and accurate.
4. **Retention** is the storage of new cases for future reuse.

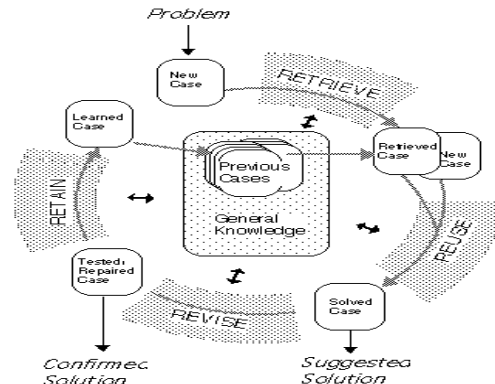


Fig.1 Aamodt Case-based reasoning cycle [1]

### 2.2.1 Case-based Reasoning related to RTS

In this section we will try to summarize some case-based reasoning researches on real-time and/or strategy games. Some CBR researches has targeted real-time individual games, as Goodman's (1994) [7] projective visualization for selecting combat actions, and predicting the next action of a human playing Space Invaders.

MAYOR (1996) [6] used a causal model to learn how to reduce the frequency of failed plan executions in SimCity, a real-time city management game. Where Ulam et al.'s (2004) [17] meta-cognitive approach performs failure-driven plan adaptation for Freeciv game. They employed substantial domain knowledge, and addressed a gaming sub-task (i.e., defend a city).

Molineaux and Ponsen (2005) [2] relax the assumption of a fixed adversary, and develop a case-based approach that learns to select which tactic to use at each state. They implemented this approach in the Case-based Tactician (CAT). They reported learning curves that demonstrate its performance quickly improves with training, even though the adversary is randomly chosen for each WARGUS game. CAT is the first case-based system designed to win against random opponents in a RTS game.

Santiago et al.,(2007) proposed Darmok [15] as the base reasoning system, which is a case-based planning system designed to play real-time strategy (RTS) games. In order to play WARGUS, Darmok learns plans from expert demonstrations, and then uses case-based planning to play the game reusing the learnt plans.

In this section, different concepts and topics related to RTS games were explained. All challenges that face RTS games were concerned with increasing game intelligence through improving tactics, reinforcement learning, player satisfaction and modeling opponents. But our concern

was different; we tried to increase game intelligence not through learning but through exchanging experiences between game engines. That we will try to explain in next section.

### 3. Real-Time Strategy Experience

#### Exchanger Model [Real-See]

As usual if you want to update any application you just need to download its update from its web site but what would you do if your engine of the application is more updated than the source itself ?!. Usually this cannot happen in ordinary applications, but here we are talking about RTS games which depend on agents trained by the recent RL techniques. This means that they can update themselves according to any changes in their environment.

In this paper we introduce our model that allowed an RTS game engine to update all other engines with the game reaction against new surprising un-programmed opponent scenarios that face the computer player. We believe this will reveal game players from downloading a new engine of the game and loosing their saved episodes. But we first needed to discuss the existing case representations and whether we can use them or we will need one of our own.

#### 3.1 Proposed Case Representation

Many case representations are depending on the game or the researcher point of view. We here tried to make use of the former representations to get a case representation that suits our model and could be applied in different RTS games. For example Aha et.al (2005) [2] defined a case C as a four-tuple:

$$C = [\text{BuildingState}, \text{Description}, \text{Tactic}, \text{Performance}]$$

Where we can consider the BuildingState as a part of the Description. We can also notice that they didn't mention the goal of the case while it is an important factor in case retrieval. From all of this we proposed a case representation of our own to use it through our model

$$C = \langle \text{State}, \text{Action}, \text{Goal}, \text{Problems to avoid}, \text{Performance} \rangle$$

- State is a vector composed of features representing game state that the system has already experienced.
- Action set is a list of case actions the agent can take at that level in the architecture.
- Goal: is a list of Goals to be achieved

- Problems to avoid: is a list of Problems to avoid
- Performance is a value in  $[0, 1]$ , reflects the utility of choosing that tactic for that state.

Our case representation concentrates on making case retrieval more accurate and easier depending first on the case state features then on goal and performance. We here used the famous Missionaries and Cannibals problem as an example of our proposed case representation as following:

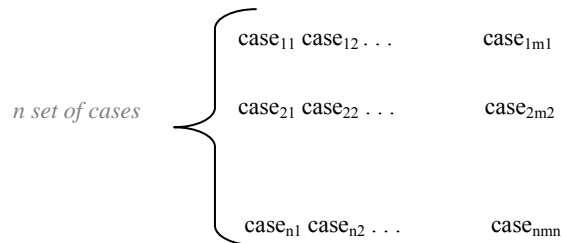
- **State** =  $\langle M, C, B, P \rangle$   
State =  $\langle 3, 3, 1, 2 \rangle$

Where M: no. of missionaries  
C: no. of Cannibals  
B: no. of boats  
P: no. of people a boat can accommodate at a time

- **Actions**  
Move (D<sub>1</sub>, D<sub>2</sub>)  
Return (D<sub>1</sub>, 0)  
Move (S<sub>1</sub>, S<sub>2</sub>)  
Return (S<sub>1</sub>, D<sub>1</sub>)  
Move (S<sub>1</sub>, S<sub>3</sub>)  
Return (D<sub>2</sub>, 0)  
Move (D<sub>2</sub>, D<sub>1</sub>)  
Return (D<sub>2</sub>, 0)  
Move (D<sub>2</sub>, D<sub>3</sub>)
- **Goals**: Cross the river
- **Problems to avoid** : Cannibals eat Missionaries
- **Performance**: Less time to solve the problem equals higher performance.

#### 3.2 Real-See Model

We supposed that n sets of cases from N engines were sent to the receiver engine figure (2). Each set consists of M<sub>n</sub> cases.



These cases represent the input of the case comparator. The case comparator compare each case of them with the cases in the case-base that most closely match the current information known, and if it found a match it discards the received case and repeat the operation on the next

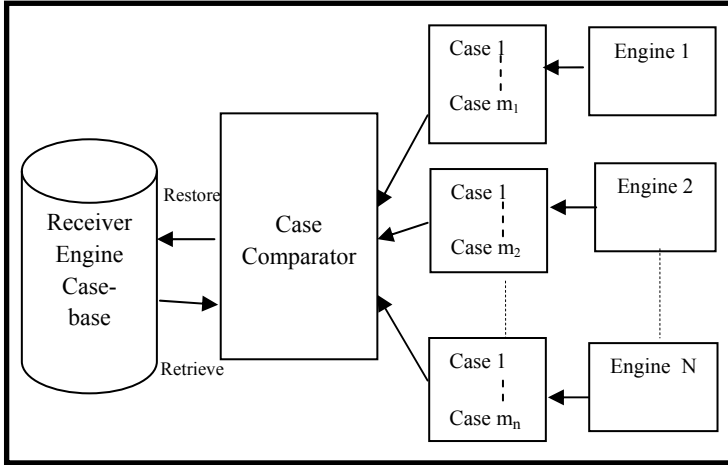


Fig.2 Real-See Model

one till it finishes all the  $M \times N$  cases. The cases that didn't have a match in the case-base will be stored in the receiver engine case-base and the rest will be deleted.

In Real-See model the case comparator plays the major role as it dose all the job. In the next section we will discuss the case comparator in details.

### 3.2.1 Case Comparator

The case comparator compare each received case with the cases in the case-base, in order to do that we will need to make use of the similarity metrics. If the case comparator did not found a similar case to the received one it will add it to the case-base but if it found a similar one it will act according to the similarity degree.

Given a received case  $P$ , the matching of case  $P$  and a retrieved case  $C$  is guided by the similarity metric in equation (1).

$$\text{similarity}(P, C) = \frac{\sum_{i=1}^k w_i \times \text{sim}(p_i, c_i)}{\sum_{i=1}^k w_i} \quad (1)$$

Where  $w_i$  is the weight of a feature  $i$ ,  $\text{sim}$  is the similarity function of features, and  $p_i$  and  $c_i$  are the values for feature  $i$  in the target and retrieved cases respectively.

But before calculating cases  $P$  and  $C$  similarity, we first needed to calculate the value of individual features similarity,  $\text{sim}(p_i, c_i)$ . The feature  $i$  similarity of both cases  $P$  and  $C$  is related to the distance between them. Many equations were used to calculate the feature similarity depending on the distance, for example

○ Euclidian distance [10][18]

$$d(P, C) = \sum_{i=1}^k \sqrt{p_i^2 - c_i^2} \quad (2)$$

○ Hamming distance [10][18]

$$H(P, C) = k - \sum_{(i=1, k)} p_i \cdot c_i - \sum_{(i=1, k)} (1-p_i) \cdot (1-c_i) \quad (3)$$

○ Absolute distance [18]

$$d(P, C) = \sum_{i=1}^k |p_i - c_i| \quad (4)$$

Here we chose to use the absolute distance divided by the feature values range specially that we are dealing with un-scaled discrete values not vectors, which is computed by:

■ Distance for *Numeric* features

$$d_i(P, C) = |p_i - c_i| / (p_i + c_i) \quad (5)$$

■ Distance for *Symbolic* features

$$d_i(P, C) = 0 \text{ if } p_i = c_i \\ = 1 \text{ otherwise} \quad (6)$$

From equations (5) and (6) we can say that

$$\text{Sim}(p_i, c_i) = 1 - d_i \quad \text{where } 0 \leq \text{Sim}(p_i, c_i) \leq 1 \quad (7)$$

The next step is to calculate feature  $i$  weight. The feature weight may be calculated using many ways for example the distance inverse but this way will be a problem if the feature values were equal which means that the distance will be zero. Here we used the inverse of the squared standard deviation; as the standard deviation represents a sample of the whole feature values population and is a measure of how widely values are dispersed from the average value. In this case of feature values equality the weight is discarded and the feature similarity value will equal 1. We here calculated the weight using equation (8).

$$w_i = 1 / \sigma(i)^2 \quad (8)$$

The last step is to calculate case  $P$  and case  $C$  similarity using equation (1), and to check its value relating to a threshold value  $\alpha$  according to our Real-See algorithm in figure (3).

In figure (3), a received case  $P$  is retained as long as its similarity value relative to case  $C$  is not above  $\alpha$ . As the result we get a set  $Q$  of retained cases as:

$$Q = \{P \in M_n \mid \text{Sim}(P, C) \leq \alpha\}$$

Where  $M_n$  is the received cases and  $\text{Sim}(P, C)$  denotes the degree of similarity of  $C$  respect to  $P$ . The elements in  $Q$  along with their similarity scores are delivered to the receiver engine case-base for to be retained.

But what happened to the cases its similarity value relative to C is above  $\alpha$ ? Shall we decline them or what? Here in our model we tried to make use of the case goal.

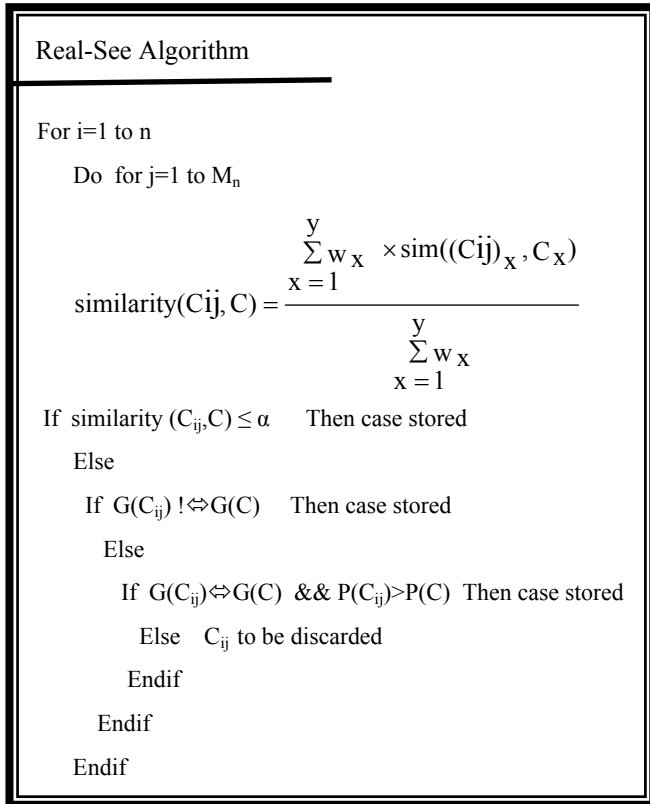


Fig.3 Real-See algorithm

Till now similarity metrics depends on the case description. In our model this means to decline cases similar to the retrieved ones. So we tried to apply the similarity metrics on the case goals, if case P similarity value relative to case C is above  $\alpha$  ( $\alpha=0.5$ ) the case comparator will compare case P and case C goals

But to calculate the goal similarity we first need to check the similarity of its parts. If there is a similarity we can express it by one else by zero. The calculated similarities is then applied in equation (9)

$$\text{MoS} = \frac{1}{R} \sum_{i=1}^R B_i \quad (9)$$

Where  $B_i$  represents the predicate  $i$  of the goal, R is the number of predicates used in similarity calculation and MoS represents the arithmetic mean of the predicates similarities and we used it as the goal similarity. We can then evaluate the mean of similarities (Mos) using equation (10)

$$\text{Goal similarity} = \begin{cases} \text{Not Similar} & \text{MoS} \leq \frac{1}{2} & 364 \\ \text{Similar} & \text{MoS} > \frac{1}{2} \end{cases} \quad (10)$$

If it found a goal match and case P performance is greater than case C performance, case P will be stored otherwise case P is declined. But if there was no goal match case P will be stored. We will explain it clearly in the next section with real picked cases.

#### 4. Testing Real-See Model on Real Cases



Fig.4 Glest – 3D RTS game

For more explanation we needed to test the Real-See algorithm on some real cases. We selected a 3D RTS game called Glest figure (4) to pick up some cases of it to go one with our algorithm testing.

- **Example 1:** We first chose a stored case called the three towers (case C) to compare it with a received case called defend the castle (case P). In the next five steps we calculated the similarity between the two cases using 14 features (table 1) to representing each case.

- The first step is to calculate feature  $i$  similarity. So we calculated the absolute distance using equations (5) and (7).
- The second step is to calculate feature  $i$  weight using equation (8).
- The third step is to calculate the similarity between case P and C using equation (1).

We can notice from table (2) that the features of value zero in both cases are discarded and were not contributed in the calculation, as it has no effect on the similarity degree which can finally be calculated as following:

$$\text{Similarity}(P,C) = 3.066/6.732 = 0.456$$



Table 1: The data set of Three\_Towers

and Destroy\_Villag cases erepresenting14 features.

Features		Three_Towers ( Case C)	Defend the Castle (Case P)
Resources	Gold	200	500
	Wood	200	500
	Stone	250	500
	Food	0	50
# of Enemy Units	Castle	0	2
	defense_tower	2	1
	Worker	0	0
	Swordman	0	0
	Archer	3	2
	Guard	0	0
	Cow	0	0
	battle_machine	0	1
	Armor	30	15
	Sight value	5	2

Table 2: Three\_Towers and Destroy\_Villag cases similarity calculations

C	P	$d_i(P,C)$	Sim ( $p_b c_i$ )	$w_i$	$w_i^*$ Sim( $p_b c_i$ )
200	500	0.429	0.571	2.222E-05	1.26984E-05
200	500	0.429	0.571	2.22222E-05	1.26984E-05
250	500	0.333	0.667	0.000032	2.13333E-05
0	50	1	0	0.001	0
0	2	1	0	0.5	0
2	1	0.333	0.667	2	1.333333333
0	0	Discarded	Discarded	Discarded	Discarded
0	0	Discarded	Discarded	Discarded	Discarded
3	2	0.2	0.8	2	1.6
0	0	Discarded	Discarded	Discarded	Discarded
0	0	Discarded	Discarded	Discarded	Discarded
0	1	1	0	2	0
30	15	0.333	0.667	0.009	0.005925926
5	2	0.429	0.571	0.222	0.126984127
Sum				6.732	3.066

- The fourth step is to check the result of the previous similarity equation according to the Real-See algorithm. From which we can see that the  $Sim(P,C) \leq 0.5$  which means that the received case (defend the castle) similarity to the stored one (the three towers) is weak and that the received case will be stored in the receiver engine case-base.
  - The last step is to pick the next new received case and start over from the first step.
- **Example 2:** To be sure of the results we had to repeat the previous steps on another new received case called tower\_of\_souls table (3) and table (4).

$$\text{Similarity}(P,C) = 7.226/10.541 = 0.686$$

Table 3: The data set of Three\_Towers and Tower\_of\_Souls cases erepresenting14 features.

Features		Three_Towers ( Case C)	Tower_of_Souls (Case P)
Resources	Gold	200	3000
	Wood	200	300
	Stone	250	1000
	Food	0	60
# of Enemy Units	Castle	0	2
	defense_tower	2	1
	Worker	3	1
	Swordman	1	2
	Archer	2	3
	Guard	1	2
	Cow	0	0
	battle_machine	0	0
	Armor	30	20
	Sight value	5	15

Table 4: Three\_Towers and Tower\_of\_Souls cases similarity calculation

C	P	$d_i(P,C)$	$Sim(p_i,c_i)$	$w_i$	$w_i^* Sim(p_i,c_i)$
200	3000	0.875	0.125	2.551E-07	3.189E-08
200	300	0.2	0.8	0.0002	0.0002
250	1000	0.6	0.4	3.555E-06	1.422E-06
0	60	1	0	0.0006	0
0	2	1	0	0.5	0
2	1	0.333	0.667	2	1.333
3	1	0.5	0.5	2	1.6
1	2	0.333	0.667	2	1.333
2	3	0.2	0.8	2	1.6
1	2	0	1	2	1.333
0	0	discarded	discarded	discarded	Discarded
0	0	discarded	discarded	discarded	Discarded
30	20	0.2	0.8	0.02	0.016
5	15	0.5	0.5	0.02	0.01
Sum				10.541	7.226

From table (4) we can see that the  $Sim(P,C) > 0.5$  Which means that the received case (tower\_of\_souls) and the stored one (the three towers) are so similar and that the received case will not be stored in the receiver engine case-base till the goal and performance similarities according to our algorithm is checked as following.

o The three towers goal is

winner (player):-  
Objective (“destroy\_towers”),  
towercount (0).

o The tower\_of\_souls goal is

winner (player):-  
Objective (“defend\_from\_attack”),  
unitcount (0),  
towercount (1).

To check the similarity of the cases goals we first need to check the similarity of its parts see table (5).

Table 5: goal similarity calculation

	Three towers goal	Tower_of_souls goal	Similarity(s)	
P <sub>1</sub>	Objective (“destroy towers”)	Objective (“defend from attack”)	No	0
P <sub>2</sub>	Missing	unitcount (0)	discarded	
P <sub>3</sub>	towercount (0)	towercount (1)	No	0
P <sub>4</sub>	winner (player)	winner (player)	yes	1

After that using equation (9), the MoS value is calculated and then evaluated according to equation (10).

$$MoS = \frac{1}{3} \sum_{i=1}^3 \{0, 0, 1\} = 1/3$$

$$Goal\ similarity = \begin{cases} \text{Not Similar} & MoS \leq 1/2 \\ \text{Similar} & MoS > 1/2 \end{cases} \quad (10)$$

Finally from equation (10) we founded out that the three towers case goal is not similar to the tower\_of\_souls case goal, but as we mentioned before that the three towers case is similar to the tower\_of\_souls case. So from all the previous and according to the Real-See algorithm we can conclude that the tower\_of\_souls case will be stored in the receiver engine case based.

- **Example 3:** to test the last case of Real-See algorithm the performance value comparison, we used a stored case called duel and a new received case called tough\_battle in table (6).

Table 6: The data set of duel and tough\_battle cases representing 14 features.

Features		Duel (Case C)	Tough_battle (Case P)
Resources	Gold	2000	500
	Wood	300	400
	Stone	1500	1000
	Food	30	60
# of Enemy Units	Castle	0	0
	defense_tower	0	0
	Worker	2	3
	Swordman	2	3
	Archer	0	0
	Guard	1	2
	Cow	0	0
	battle_machine	1	3
	Armor	10	40
	Sight value	15	10

Table 7: Duel and Tough\_battle cases similarity calculations

<i>C</i>	<i>P</i>	$d_i(P,C)$	$Sim(p_i,c_i)$	$w_i$	$w_i^* Sim(p_i,c_i)$
2000	500	0.6	0.4	8.88889E-07	3.55556E-07
300	400	0.143	0.857	0.0002	0.0002
1500	1000	0.2	0.8	0.000008	0.0000064
30	60	0.333	0.667	0.002	0.002
0	0	Discarded	Discarded	Discarded	Discarded
0	0	Discarded	Discarded	Discarded	Discarded
2	3	0.2	0.8	2	1.6
2	3	0.2	0.8	2	1.6
0	0	Discarded	Discarded	Discarded	Discarded
1	2	0.333	0.667	2	1.333
0	0	discarded	Discarded	Discarded	discarded
1	3	0.5	0.5	0.5	0.25
10	40	0.6	0.4	0.002	0.001
15	10	0.2	0.8	0.08	0.064
Sum				6.585	4.849

$$\text{Similarity}(P,C)=4.849/6.585=0.73654$$

From table (7) we can see that the  $Sim(P,C)>0.5$  Which means that the received case (tough\_battle) and the stored one (duel) are so similar and that the received case will not be stored in the receiver engine case-base till the goal and performance similarities according to our algorithm is checked.

As in example 2 we will check the duel and tough\_battle cases goal similarities as following:

- The duel goal is  
winner (player):-  
objective (“defend\_from\_attack”),  
unitcount (0).
- The tough\_battle goal is  
winner (player):-  
Objective (“defeat\_enemy”),  
unitcount (0).

To check the similarity of the cases goals we first need to check the similarity of its parts see table (8)

Table 8: Goal Similarity Calculation

	<i>Duel Goal</i>	<i>Tough_Battle Goal</i>	<i>Similarity(S)</i>	
P <sub>1</sub>	Objective (“defend_from_attack”)	Objective (“defeat_enemy”)	No	0
P <sub>2</sub>	unitcount (0).	unitcount (0).	Yes	1
P <sub>3</sub>	winner (player)	winner (player)	Yes	1

After that using equation (9) the MoS value is calculated and then evaluated according to equation (10)

$$\text{MoS} = \frac{1}{3} \sum_{i=1}^3 \{0, 1, 1\} = 2/3$$

From equations (9) and (10) we can see that the duel case goal is similar to the tough\_battle case goal, we also mentioned before that the duel case is similar to the tough\_battle case. From this and according to Real-See algorithm we can definitely say that we need to check the last case of our algorithm the performance value case.

Suppose that the performance of duel case is 0.63 while the performance of the tough\_battle case is 0.7 this means that the received tough\_battle case will be stored in the engine case-based. And after storing the case, the case comparator will start over again from the first step with a new received case.

## 5. Conclusions

In this paper, we have presented an experience exchanging model to improve the performance of RTS game engines through exchanging experiences of facing new un-programmed opponent scenarios. Our model is based on the game case-based reasoning system specially on adding new cases to it. New cases are sent by other engines that faced new opponent scenarios and beat them to help the engine dealing with these scenarios if it faces them in the future. We believe this will reveal game players from downloading a new engine of the game and loosing their saved stages.

Our main priority here was to be sure that these received cases are all new to the system and have no matching cases in the game CBR. In order to do that we also introduced an algorithm which we call Real-See to check the similarity of these received cases to the stored ones. This algorithm is not concentrating on the case description only but on the case goal and performance too. We tested the Real-See algorithm on real picked cases from 3D-Glest RTS game and it performed well.

## Future Work

In the future we plan to pursue several future researches on the case-based situation assessment depending on Real-See algorithm and whether it helps to enlarge the case-based or to shrink it. We also will try to introduce an implementation of the Real-See algorithm in both Glest and Waragus open-source real time strategy games.

## References

- [1] Aamodt A. and Plaza E., "Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches" In Proceedings of AICOM - Artificial Intelligence Communications, IOS Press, Vol. 7: 1,1994, pp. 39-59
- [2] Aha D., Molineaux M. and Ponsen M., "Learning to win: Case-based plan selection in a real-time strategy game". In Proceedings of Sixth International Conference on Case-Based Reasoning, 2005, pp. 5-20. Springer.
- [3] Balla R. and Fern A., "UCT for Tactical Assault Planning in Real-Time Strategy Games". In Proceedings of the 21st international JONT conference on Artificial intelligence, 2009, pp.40-45, Pasadena, California, USA.
- [4] Copeland, J., "Artificial intelligence: A Philosophical Introduction". United Kingdom: Blackwell Publishers, 1993.
- [5] Dimitriadis V. K., "Reinforcement Learning in Real Time Strategy Games Case Study on the Free Software Game Glest". Department of Electronic and Computer Engineering Technical University of Crete. China, 2009.
- [6] Fasciano M., "Everyday-World Plan Use". The University of Chicago, Computer Science Department. Chicago, Illinois, 1996.
- [7] Goodman M., "Results on Controlling Action With Projective Visualization". In Proceedings of the Twelfth National Conference on AI. 1994, pp. 1245-1250. Seattle, WA: AAAI Press.
- [8] Hammond K., "Case-Based Planning: A Framework for Planning from Experience". In Proceedings of Journal of Cognitive Science. Ablex Publishing, Norwood, NJ. Vol. 14, 1994.
- [9] Kok, E. "Adaptive reinforcement learning agents in RTS games". University Utrecht, The Netherlands, 2008.
- [10] Li M., et al., "The Similarity Metric". In Proceedings of IEEE Transactions on Information Theory, Aug. 2004.
- [11] Marthi B, Russell S., and Latham D., "Writing Stratagus Playing Agents in Concurrent Alisp". In Proceedings of Workshop on Reasoning Representation and Learning in Computer Games (IJCAI-05), 2005, pp. 67-71.
- [12] Ponsen M., Muñoz-Avila H., Spronck P., and Aha D., "Automatically Generating Game Tactics via Evolutionary Learning", Proceedings of AI Magazine, vol.(27), 2006, pp. 75-84.

[13] Simpson R., "A Computer Model of Case-based Reasoning in Problem Solving". PhD thesis, Georgia Institute of Technology, 1985.

[14] Ponsen M. and Spronck P., "Automatically acquiring domain knowledge for adaptive AI games using evolutionary learning". In Proceedings of the 17th conference on Innovative applications of artificial intelligence. Vol.(3) .Pittsburgh, Pennsylvania, 2004, pp:1535-1540.

[15] Santiago O., Mishra K., Sugandh N. and Ram A., "Case-based planning and execution for real-time strategy games". In Proceedings of ICCBR-2007, 2007, pp 164-178.

[16] Tracy B., "Game Intelligence AI Plays Along". In Proceedings of the Computer Power User. Volume 2, Issue 1, 2002, pp 56-60.

[17] Ulam P., Goel A. & Jones J., "Reflection in Action: Model-Based Self-Adaptation in Game Playing Agents". In Proceedings of D. Fu & J. Orkin (Eds.) Challenges in Game Artificial Intelligence: Papers from the AAAI Workshop (TR WS-04-04). San Jose, CA: AAAI Press, 2004.

[18] [WWW.Wikipedia.org](http://WWW.Wikipedia.org)

Mostafa Aref is a Professor and Chairman of Computer Science Department, Ain Shams University. He got his B.Sc. from Ain Shams University, Egypt; his M.Sc. from University of Saskatchewan, Canada; and his Ph.D. from University of Toledo, Ohio, USA. He worked in Several Universities in USA, Saudi Arabia and Egypt. Currently he is a coordinator of two research groups on NLP and RTS games in Ain Shams University.

Magdy Zakaria is an assistant professor and Chairman of Computer Science Department in Mansoura University. He is the Decision Support Systems unit coordinator at faculty of computers & Information in Mansoura University. He has supervised over 10 PhDs and 15 masters mostly specialized in Artificial Intelligence and its applications related to real life. As a result of his work he has published over 40 papers. Current project is grid computing.

Shahenda Sarhan is a PHD student and an assistant lecturer at the Computer Science Department in Mansoura University. Her subject is in Real-Time Strategy games.

# Sensitivity Analysis of TSEB Model

## by One-Factor-At-A-Time in irrigated olive orchard

Abdelhaq Mouida<sup>1</sup> and Nouredine Alaa<sup>2</sup>

<sup>1</sup> National Weather Service, Marocmeteo  
Rabat, Morocco

<sup>2</sup> Department of Applied Mathematics and Informatics, University of Cadi Ayad, Faculty of Science and Techniques  
Marrakech, Morocco

### Abstract

The aim objective of this present study is to identify the most influencing constant parameters of Two Source Energy Balance (TSEB) Model over irrigated olive orchard in semi-arid area. TSEB (Norman et al. 1995) has been based on surface radiometric temperature, Priestley-Taylor estimation of canopy latent heat, climatic forcing and partitioning energy to double sources (canopy and soil) according parallel resistances network. Sensitivity analysis by approach One-Factor-At-A-Time (OAT) was been studied using Eddy Covariance ground measurements data collected during SUDMED Project in Agdal site, Marrakech, Morocco (2003). Data include surface energy fluxes, meteorological inputs and vegetation parameters related to olive orchard. OAT consists in modifying each input parameters of the model by  $\pm 10\%$  around its initial value. The effect of each operated modification is analyzed on four outputs of the model (i.e: Net radiation, latent heat, sensible heat and soil heat), using variation rate and sensitivity index. The input parameters data such as Leaf Area Index (LAI), Priestley-Taylor constant ( $\alpha_p$ ), and fraction of LAI that is green ( $f_g$ ) have successively a percentage variation of 18.4%, 15.1%, and 15.1% shown to have the greatest impact on the TSEB estimate of the fluxes.

Thus, the results obtained give a fairly clear idea of the most important entrances of TSEB. They can guide the user through the calibration process and also in collecting experimental data.

**Keywords:** TSEB Model, Sensitivity analysis, One-Factor-At-A-Time, Sensitivity index, percentage of variation.

### 1. Introduction

All models describing biophysical phenomenon depending on two kind of uncertainty: First one is due to a system description and second is due to model parameters which estimated through experimental data (Ratto et al., 1996). The values of parameters influence seriously prediction

even correct biophysical description (Ratto et al., 1996). The coherence between model and its biophysical system is essential and is evaluated by sensitivity analysis (Saltelli et al., 1999). The sensitivity and the variation level of output versus constant uncertainties are must be known. Sensitivity analysis permit us to evaluate all constant parameter effect on model result to classify them according to their sensitivity level (Saltelli et al., 2000), and to tune parameters at the time of determination on experiment (Jolicoeur, 2002). This paper highlight the model description used for this study in section 2, while section 3 describe the sensitivity analysis method, and section 4 presents results of sensitivity analysis. Conclusion and perspectives are presented in section 5.

### 2. Brief description of TSEB Model

TSEB Model is based on energy balance closure using surface radiometric temperature, vegetation parameters and climatic data. TSEB outputs surface turbulent fluxes, and temperatures of canopy and soil. The version implemented in this study basically follows what is described in appendix A as the “parallel resistance network”. As such, the model implemented is described in detail in (Norman et al. 1995, Kustas et al. 1999).

### 3. Sensitivity analysis Method

The main goal of this study is to identify among input parameters the most sensitive to model outputs; (i.e: those for which a little variation may involve a great change in model result, (Saltelli et al., 2000b). Screening Designs method of sensitive analysis is utilized here under technique of OAT (Rody Félix, Dimitri Xanthoulis; 2005),



which identify among input parameters whose contribute more to variability of 4 output model: Net radiation, latent heat, sensible and soil heat.

### 3.1 One-Factor-At-A-Time (OAT) method

OAT is the simple technique of Screening Designs (SD) method to carry out a sensitivity analysis. It consists to identify most sensitive parameter among those may be affecting model output (Nearing et al., 1990). SD is efficient when a model has several input parameter (Jolicoeur, 2002). To assess the impact of errors or variation

$\pm 10\%$  around base input value, a sensitivity analysis of TSEB model was performed by computing relative variation rate  $Vr(p)$  and sensitivity index  $SI(p)$ . The effect of each operated modification is analyzed on 4 outputs of the model (i.e: Net radiation, latent heat, sensible heat and soil heat), using variation rate and sensitivity index.

The relative variation rate  $Vr(p)$ , and sensitivity index,  $SI(p)$  of a model flux estimate, in a parameter  $p$ , can be expressed as

$$Vr(p) = \left| \frac{S_2 - S_1}{S_1 - S_1} \right| \cdot 100$$

$$SI(p) = \frac{S_{moy}}{\frac{E_2 - E_1}{E_{moy}}}$$

where  $SI$  is the sensitivity index of model output ;  $E_1$  the initial input parameter ;  $E_2$  the tested input value (e.g  $\pm 10\%$  modification lag);  $E_{moy}$  average between  $E_1$  and  $E_2$ ;  $S_1$ ,  $S_2$  are respectively the outputs corresponding to  $E_1$  and  $E_2$ ;  $S_{moy}$  is the average between  $S_1$  and  $S_2$ .

This index provides a quantitative basis for expressing the sensitivity of model outputs versus the input variables. A sensitivity index equal to unity indicates that the rate of variation of a given parameter causes the same rate at the outputs, but a negative value indicates that the inputs and outputs vary in opposite directions. The index in absolute value is greater then its impact of a given parameter which might have on a specific output.

The model outputs are treated as follows:

- 1- In fact, the change of each input variable by  $\pm 10\%$  produces two values for each selected outputs. From these two introduced input values, the greatest variation at a given output is used to calculate its sensitivity index ( $SI$ ).
- 2- A percentage change (Favis-Mortlock, Smith, 1990) and a sensitivity index (Jolicoeur, 2002) are calculated for each output selected above by previous formulas:

Generally, factors screening may be useful as a first step when dealing with a model containing several no identified parameters. These parameters have often a significant effect on the model output. Screening experiment are used to identify the subset of factors that controls most of the output variability with a relatively low computational effort. This economical method tends to provide qualitative sensitivity measures, (i.e: it ranks the input factors in order of importance, but do not quantify how much a given factor is more important than another.

## 4. Results and discussion

### 4.1 Overview

The input parameters used in this sensitivity analysis are the Priestly-Taylor constant ( $\alpha_p$ ), the leaf area index (LAI), the fraction of the LAI that is green ( $fg$ ), the fraction of the soil net radiation ( $cg$ ), the canopy height ( $h$ ), the mean leaf size ( $s$ ) is given by four times the leaf area divided by the perimeter, the surface emissivity ( $\epsilon$ ), and the surface albedo ( $\alpha$ ). After modifying alternately each model input of datasets mentioned above by  $-10\%$  and  $+10\%$  around its initial value, we analysis only percentage greater than  $0.5\%$ . Such inaccuracies can be derived either from some variability inherent in any consideration or measurement on field. A total of 6983 simulation is performed on the semi-hourly data set obtained from SUDMED Project (The fall year 2003). Each simulation performed here takes into account the change only one input relative to the overall model parameters. The effect of each change made is analyzed in the four model outputs (i.e: Sensible heat (H), Latent heat (LE), Net radiation (Rn) and Ground conduction heat (G)).

### 4.2 Sensitivity of sensible heat (H)

Input parameters modification produce variation rate from  $0.7\%$  to  $32.6\%$  on sensible heat. LAI,  $\alpha_p$  and  $fg$  are the most sensitive parameter on this output (fig.1). They produce variation respectively of  $32.59\%$ ,  $23.55\%$  and  $23.55\%$ . Sensible heat accuse sensitivity index respectively of  $-3.4$  to  $-2$ . It is most sensitive to LAI with  $-3.4$  as negative sensitivity index. This analysis indicates that high uncertainties on these inputs may falsify seriously results of sensible heat. Indeed, it's clear that when vegetation is developing then LAI is increasing and the sensible heat is decreasing (i.e: negative sensitivity index) because vegetation play a role of shock-absorber. Therefore vegetation play a role of shock-absorber, then reduce considerably soil sensible heat with variation rate  $100\%$  ( $SI=-21$ ) and also soil heat stock ( $14.4\%$  with  $SI=-1.28$  (fig.1)). However, this case is occurred during

development phase of olive trees (e.g. during July, August, and September). That is why LAI is related strongly to development phase and has an important influencing in sensible heat especially its soil component.

For the case of the olive, LAI don't vary too much during seasons. Sensible heat is also sensitive to  $fg$  and  $\alpha p$  with 23.59% of variation ( $SI=-2$ ). These parameters reduce considerably canopy sensible heat.  $fg$  represents the green fraction of vegetation and it's increasing play in the opposite direction to total sensible heat especially in the soil contribution.

#### 4.3 Sensitivity of sensible heat (LE)

Figure 2 indicate that LAI,  $fg$ , and  $\alpha p$  are the important input for latent heat. LAI produce a variation rate of 8.13%,  $fg$  and  $\alpha p$  are 6.67% with sensitivity index respectively of 0.74 and 0.65 for input. We observe that sensitivity index is negative for emissivity, albedo,  $cg$  and  $s$ . It means that these parameters vary inversely to total latent heat input. Note well that LAI is also the most sensitive factor on output. We have the same ascertainment then for total sensible heat varies inversely. On TSEB, LAI play an important role in fractional cover vegetation. It's sensitivity index is positive then it confirm a good influence in evapotranspiration and evolves both in the same direction. However, any doubt measurements or uncertainties in LAI index cause some errors in latent heat. Moreover,  $fg$  and  $\alpha p$  are the same influencing in evapotranspiration like LAI.

#### 4.4 Sensitivity of net radiation (Rn)

Net radiation undergoes only the both influence of surface emissivity and albedo having variation rate respectively of 2.9% and 1.6% with negative sensitivity as -0.29 and -0.15. It indicates that these parameters evolve inversely effect to net radiation. Net radiation depends also on climatic variables as long wave, short wave and radiometric temperature. However, inaccuracies intricate always on this output, cause errors can occur on these two parameters. In effect an uncertainty of 10% on albedo and emissivity cause only a variation of 1 to 3% at the outlet (Fig.3).

#### 4.5 Sensitivity of soil conduction heat (G)

Entries LAI,  $\alpha$  and  $\epsilon$  affect  $G$  respectively with a variation rate of 14.4%, 2.9% and 1.6% with negative sensitivity indices as respectively -1.28, -0.29 and -0.16 (Fig.4). LAI is the most influential parameter on  $G$  as it is normal and consistent with what we saw previously, because the index indicates the leaf area cover and play a role of shock-absorber. The sensitivity is negative, then it means more vegetation is growing the radiation received by the ground

is lower and the higher the ground stock heat decreases. In fact, it seems natural that the LAI has this influence on the stock to heat in the soil because it is one of the main parameters that control the level of heat storage in the soil. Uncertainty on this entry could have some imprecision on  $G$  which unfortunately is poorly estimated by the model.

#### 4.6 Comparison of changes in TSEB surface fluxes

An average variation determined for the 4 outputs considered and for each entry shows that LAI is the most important parameter with an average change produced approximately 18.4%. It is followed by  $\alpha p$  and  $fg$  whose variations are 15.1%. Globally changes in other inputs have little influence on model outputs (Fig. 5). Comparing the results of the sensitivity analysis obtained shows a certain similarity in the sensitivity of the four outputs selected with the variation of model inputs of  $\pm 10\%$  from their initial value.

### 5. Conclusions and perspectives

The sensitivity analysis of TSEB model has been applied using One-Factor-At-A-Time (OAT) which is a typical screening designs to assess all constant parameter effect on model result and to classify them according to their sensitivity level. Although simple, easy to implement and computationally cheap, the OAT methods have a limitation in that they do not enable estimation of interactions among factors and usually provide a sensitivity measure that is local. Input parameters used in this sensitivity analysis are the Priestly-Taylor constant ( $\alpha p$ ), the leaf area index (LAI), the fraction of the LAI that is green ( $fg$ ), the fraction of the soil net radiation ( $cg$ ), the canopy height ( $h$ ), the mean leaf size ( $s$ ), the surface emissivity ( $\epsilon$ ), and the surface albedo ( $\alpha$ ). The input parameters data such as LAI,  $\alpha p$ , and  $fg$  are successively (18.4%, 15.1%, and 15.1%) shown to have the greatest impact on the TSEB estimate of the fluxes.

As a result, the sensitivity of the TSEB model output in  $H$  to uncertainties in LAI,  $\alpha p$  and  $fg$  don't exceeded 33% of its reference value. On the other hand, sensitivity of the TSEB model output in  $LE$  to these parameters uncertainties was generally less than 8% and not influencing  $Rn$  and  $G$  except for LAI which have 14% of uncertainties to  $G$ .

The results of a sensitivity analysis should be handled with care, since the apparent sensitivity of a model for a given parameter depends on the importance, during the chosen period, the process that affects this parameter, itself linked

to environmental constraints and to the initial conditions. Thus, in this study, the results obtained give a fairly clear idea of the most important entrances of TSEB. They can guide the user through the calibration process and also in collecting experimental data.

## Appendix A

### TSEB Equations

Soil and vegetation temperature contribute to the radiometric surface temperature in proportion to the fraction of the radiometer view that is occupied by each component along with the component temperature. In particular, assuming that the observed radiometric temperature, ( $T_{rad}$ ) is the combination of soil and canopy temperatures, the TSEB model adds the following relationship (Becker and Li, 1990) to the set of (Eqs 12 and 13):

$$T_{rad}(\theta) = [f(\theta) \cdot T_c^4 + (1-f(\theta)) \cdot T_s^4]^{1/4} \quad (A.1)$$

where  $T_c$  and  $T_s$  are vegetation and soil surface temperatures, and  $f(\theta)$  is the vegetation directional fractional cover (Campbell and Norman, 1998).

$$f(\theta) = 1 - \exp(-0.5 \text{ LAI} / \cos(\theta)) \quad (A.2)$$

The simple fractional cover ( $f_c$ ) is as follows:

$$f_c = 1 - \exp(-0.5 \text{ LAI}) \quad (A.3)$$

LAI is the leaf area index, and the fraction of LAI that is green ( $f_g$ ) is required as an input and may be obtained from knowledge of the phenology of the vegetation.

The total net radiation  $R_n$  ( $\text{Wm}^{-2}$ ) is

$$R_n = H + LE + G \quad (A.4)$$

where  $H$  ( $\text{Wm}^{-2}$ ) is the sensible heat flux,  $LE$  ( $\text{Wm}^{-2}$ ) is the latent heat, and  $G$  ( $\text{Wm}^{-2}$ ) is the soil heat flux. The estimation of total net radiation,  $R_n$  can be obtained by computing the net available energy considering the rate lost by surface reflection in the short wave ( $0.3/2.5\mu\text{m}$ ) and emitted in the long wave ( $6/100\mu\text{m}$ ):

$$R_n = (1 - \alpha_s) \cdot SW + \epsilon_s \cdot LW - \epsilon_s \cdot \sigma \cdot T_{rad}^4 \quad (A.5)$$

where  $SW$  ( $\text{Wm}^{-2}$ ) is the global incoming solar radiation,  $LW$  ( $\text{Wm}^{-2}$ ) is the terrestrial infrared radiation,  $\alpha_s$  is the surface albedo,  $\epsilon_s$  is the surface emissivity,  $\sigma$  is the Stefan-Boltzmann constant,  $T_{rad}$  ( $^{\circ}\text{K}$ ) is the radiometric surface temperature.

The estimation of soil net radiation,  $R_{ns}$  can be obtained by

$$R_{ns} = R_n \exp(-K_s \text{ LAI} / \sqrt{2 \cdot \cos(\theta)}) \quad (A.6)$$

where  $k_s$  is a constant ranging between 0.4 to 0.6 and is the zenithal solar angle.

The  $R_{nc}$  is the canopy net radiation as

$$R_{nc} = R_n - R_{ns} \quad (A.7)$$

where  $R_n$  is obtained using (A.4-5) and is the solar zenith angle. The soil heat flux,  $G$  ( $\text{Wm}^{-2}$ ) can be expressed as a constant fraction  $c_g$  ( $\approx 0.35$ ) of the net radiation at the soil surface by

$$G = c_g R_{ns} \quad (A.8)$$

The constant of  $c_g$  ( $\approx 0.35$ ) is midway between its likely limits of 0.2 and 0.5 (Choudhury et al 1987). The canopy latent heat  $LE_c$  is given by Priestly-Taylor approximation (Priestly-Taylor, 1972).

$$LE_c = R_{nc} \cdot \alpha_p \cdot f_g \cdot \frac{\Delta}{\Delta + \Gamma} \quad (A.9)$$

where  $\alpha_p$  is the Priestly-Taylor constant, which is initially set to 1.26 (Norman et al 1995; Agam et al 2010),  $f_g$  is the fraction of the LAI that is green,  $\Delta$  is the slope of saturation vapor pressure versus temperature curve,  $\Gamma$  is the psychrometer constant (e.g:  $0.066 \text{ kPa } ^{\circ}\text{C}^{-1}$ ). If no information is available on  $f_g$ , then it is assumed to be near unity. As will become apparent later (A.9) is only an initial approximation of canopy latent heat.

If in any case  $LE_c \leq 0$ , then  $LE_c$  is set to zero (i.e: no condensation under daytime convective conditions)

The sum of the contribution of the soil and canopy net radiation, total latent and sensible heat is according to the following equations

$$R_{ns} = H_s + LE_s + G \quad (A.10)$$

$$R_{nc} = H_c + LE_c \quad (A.11)$$

$$LE_t = LE_c + LE_s \quad (A.12)$$

Where the subscript s and c designs soil and canopy.

The TSEB model considers also the contributions from the soil and canopy separately and it uses a few additional parameters to solve for the total sensible heat  $H_t$  which is the sum of the contribution of the soil  $H_s$  and of the canopy  $H_c$  according to the following equations

$$H_t = H_s + H_c \quad (A.13)$$

$$H_c = \rho C_p \left[ \frac{T_c - T_a}{R_a} \right] \quad (A.14)$$

$$H_s = \rho C_p \left[ \frac{T_s - T_a}{R_s + R_a} \right] \quad (A.15)$$

Where  $\rho$  ( $\text{Kg.m}^{-3}$ ) is the air density,  $C_p$  is the specific heat of air ( $\text{JKg}^{-1} \text{K}^{-1}$ ),  $T_a$  ( $^{\circ}\text{K}$ ) is the air temperature at certain reference height, which satisfies the bulk resistance formulation for sensible heat transport (Kustas et al, 2007).  $R_a$  ( $\text{sm}^{-1}$ ) is the aerodynamic resistance to heat transport across the temperature difference that can be evaluated by the following equation (Brutsaert, 1982):

$$R_a = \frac{\ln \left[ \frac{(z_u - d_0)}{z_0, H} - \Psi_H \right]}{k U_*} \quad (A.16)$$

Where  $z_u$  is the height of air wind measurements,  $U_*$  is the wind friction velocity,  $d_0$  (m) is the displacement height,  $z_0, H$  is a roughness parameter (m) that can be evaluated as function of the canopy height (Shuttleworth and Wallace, 1985),  $k$  is the von Karman's constant ( $\approx 0.4$ ),  $\Psi_H$  is the diabatic correction factor for heat is computed (Paulson, 1970):

$$\Psi_H = 2. \ln \left[ \frac{1 + 0.5 \xi}{a} \right] \quad (A.17)$$

Where  $\xi$  is a universal function for heat defined by: (Brutsaert, 1982; Paulson, 1970)

$$\xi = (1 - 16. \xi)^{1/4} \quad (A.18)$$

The term  $\xi$  is dimensionless variable relating observation height  $Z$ , to Monin-Obukhov stability  $L_{mo}$ .  $L_{mo}$  is approximately the height at which aerodynamic shear, or mechanical, energy is equal to buoyancy energy (i.e: convection caused by an air density gradient). It is determined from

$$L_{mo} = -\rho \frac{v_*^3}{k \rho \left( \frac{H}{\rho T_m} + 0.61 \frac{LE}{\lambda} \right)} \quad (A.19)$$

Where  $\rho$  ( $\text{Kg.m}^{-3}$ ) is the air density,  $C_p$  is the specific heat of air ( $\text{JKg}^{-1} \text{K}^{-1}$ ),  $T_a$  ( $^{\circ}\text{K}$ ) is the air temperature at certain reference height,  $H$  is a sensible heat flux,  $LE$  is a latent heat flux, and  $\lambda$  is the latent heat.

Friction velocity is a measure of shear stress at the surface, and can be found from the logarithmic wind profile relationship:

$$U_* = \frac{k U_a}{\ln \left[ \frac{(z_u - d_0)}{z_0, M} - \Psi_M \right]} \quad (A.20)$$

Where  $U_a$  is the wind speed and  $\Psi_M$  is the diabatic correction for momentum.

The  $R_s$  ( $\text{sm}^{-1}$ ) is the soil resistance to the heat transfer (Goudriaan, 1977; Norman et al 1995; Sauer et al 1995; Kustas et al, 1999), between the soil surface and a height representing the canopy, and then a reasonable simplified equation is:

$$R_s = \frac{1}{a' + b' U_s} \quad (A.21)$$

Where  $a' = 0.004$  ( $\text{ms}^{-1}$ ),  $b' = 0.012$  and  $U_s$  is the wind speed in ( $\text{ms}^{-1}$ ) at a height above the soil surface where the effect of the soil surface roughness is minimal; typically 0.05 to 0.2 m. These coefficients depend on turbulent length scale in the canopy, soil surface roughness and turbulence intensity in the canopy and are discussed by (Sauer et al. 1995). If soil temperature is great than air temperature the constant  $a'$  becomes  $a' = c \cdot (T_s - T_c)^{1/3}$  with  $c = 0.004$

$U_s$  is the wind speed just above the soil surface as described by (Goudriaan 1977):

$$U_s = U_c \cdot \exp \left[ -a \left( 1 - \frac{0.05}{h_c} \right) \right] \quad (A.22)$$

Where the factor (a) is given by (Goudriaan 1977) as

$$a = 0.28 \cdot P^{2/3} \cdot h_c^{1/3} \cdot s^{-2/3} \quad (A.23)$$

The mean leaf size (s) is given by four times the leaf area divided by the perimeter.

$U_c$  is the wind speed at the top of the canopy, given by:

$$U_c = U_a \frac{\ln \left( \frac{h_c - d}{z_0, M} \right)}{\ln \left( \frac{z_u - d}{z_0, M} \right) - \Psi_M} \quad (A.24)$$

Where  $U_a$  is the wind speed above the canopy at height  $z_u$  and the stability correction at the top of the canopy is assumed negligible due to roughness sublayer effects (Garratt, 1980; Cellier et al, 1992).

TSEB implementation and algorithm



The TSEB model is run with the use of ground thermal remote sensing and meteorological data of Agdal site during 2003. Some model constant parameters are supposed invariable along time such as the Priestly-Taylor constant  $\alpha_p$ , albedo, emissivity, leaf area index (LAI), the fraction of the LAI that is green (fg), leaf size (s), the vegetation height and a constant fraction (cg) of the net radiation at the soil surface. These considerations are certainly some consequences on model results according to seasons. The Priestly-Taylor constant  $\alpha_p$  is fixed to 1.26 (McNaughton and Spriggs 1987). The albedo, value of 0.11 is an annual averaged measured with CNR1, and a surface emissivity of 0.98, the leaf area index (LAI) is equal to 3 (Ezzahar et al, 2007). The fraction of LAI (fg) that is green is fixed to 90% of vegetation (i.e: 10% of vegetation could be considered no active). The mean leaf size (s), is given by four times the leaf area divided by the perimeter ( $s=0.01$ ). The average height of the olive trees is 6 meters. The fraction of the net radiation at the soil surface is fixed to  $cg=0.35$ .

Sensible and latent heat flux components for soil and vegetation are computed by TSEB, only in the atmospheric surface layer instability. Note that the storage of heat within the canopy and energy for photosynthesis are considered negligible for the instantaneous measurements. The total computed heat flux components are then from equations (A.5-8).

The canopy heat fluxes are solved by first estimating the canopy latent heat flux from the Priestley-Taylor relation (A.9), which provides an initial estimation of the canopy fluxes, and can be overridden if vegetation is under stress (Norman et al., 1995). Outside the positive latent heat situation, two cases of stress occur, when the computed value for canopy ( $LE_c$ ) or soil ( $LE_s$ ) latent heat become negative which are an unrealistic conditions.

In the first case, the normal evaluation procedure is overridden by setting ( $LE_c$ ) to zero and the remaining flux components are balanced by (A. 1-10-11-13-15). But in the second case, ( $LE_s$ ) is recomputed by using specific soil Bowen Ratio determined by  $\beta = H_s/LE_s$  and flux components are next balanced by (A.1-10-11-13-15).

In order to solve (A.15) additional computations are needed to determine soil temperature, and the resistance terms  $R_{ah}$  and  $R_s$  but as will become apparent, they must be solved iteratively. Soil temperature is determined from two equations: one to relate the observed radiometric temperature to the soil and vegetation canopy temperature, and another to determine the vegetation canopy temperature. The composite temperature is related to soil and canopy temperatures by (A.1). The resistance components are determined from (A.16), for  $R_{ah}$  and the following equation (Sauer et al., 1995) for  $R_s$  (A.18).

To complete the solution of the soil heat flux components, the ground stock heat flux can be computed as a fraction of net radiation at the soil surface (A.8).

Applying energy balance for the two source flux components resolves the surface fluxes, which cannot be reached directly because of the interdependence between atmospheric stability corrections, near surface wind speeds, and surface resistances (A.16-17). In these equations, the stability correction factors  $\Psi_M$  and  $\Psi_H$  depend upon the surface energy flux components  $H$  and  $LE$  via the Monin-Obukhov roughness length  $L_{mo}$ .

TSEB computation for solving the surface energy balance by ten primary unknowns and ten associated equations (Table.1), needs an iterative solution process by setting a large negative value to  $L_{mo}$  (i.e: in highly unstable atmospheric conditions). This permits an initial set of stability correction factors  $\Psi_M$  and  $\Psi_H$  to be computed. Computed iteration is repeated until  $L_{mo}$  converges.

### Acknowledgments

This study is considered within the framework of research between the University of Cadi Ayad Gueliz, Marrakech, Morocco, and the Department of National Service of Meteorology, Morocco (DMN, Morocco). The first author is very grateful for encouragement to all his family especially to Mrs F. Bent Ahmed his mother, Mrs K.Aglou his wife and Mr Mustapha.Mouida. Finally the authors gratefully acknowledge evaluation and judgments by reviewers, and the editor.

### References

- [1] Agam et al, "Application of the Priestley-Taylor Approach in Two Source Surface Energy Balance Model", Am Meteo Soc, Journal of Hydrometeorology, Volume 11, 2010, pp. 185-198.
- [2] Becker. F, and Li. Z.L, "Temperature independent spectral indices in thermal infrared bands" Remote Sensing of Environment, vol. 32, 1990, pp. 17-33.
- [3] Brutsaert, W, Evaporation Into The Atmosphere, D. Reidel, Dordrecht, 1982.
- [4] Choudhury, B.J, Idso, S.B, and Reginato, R.J, " Analysis of an empirical model for soil heat flux under a growing wheat crop for estimating evaporation by an infrared-temperature based energy balance equation", Agric. For. Meteorol, Vol. 39, pp. 283-297.
- [5] Campbell, G. S, and Norman, J. M, An Introduction to Environmental Biophysics, (2nd ed.): New York: Springer-Verlag, 286 pp. 1998.
- [6] Ezzahar.J, " Spatialisation des flux d'énergie et de masse à l'interface Biosphère-Atmosphère dans les régions semi-arides en utilisant la méthode de scintillation ", Ph.D. thesis, University of Cadi Ayyad. Marrakech, Morocco, 2007.



[7] Favis, Mortlock DT, Smith FR, "A sensitivity analysis of EPIC ", Documentation. U.S. Department of Agriculture. Agric. Tech. Bull. 1768, 1990, pp. 178–190.

[8] Garratt et al, "Momentum, heat and water vapor transfer to and from natural and artificial surfaces ", Q. J. R. Meteorol. Sot., 99, pp. 680-687.

[9] Goudriaan, J, "Crop Micrometeorology: A Simulation Study ", Center for Agricultural Publications and Documentation, Wageningen, 1977.

[10] Jacob. F et al, "Using airborne vis-NIR-TIR data and a surface energy balance model to map evapotranspiration at high spatial resolution", In Remote sensing and hydrology IAHS-AISH, 2000.

[11] Jolicoeur, "Screening designs sensitivity of a nitrate leaching model (ANIMO) using a one-at-a-time method", USA: State University of New York at Binghamton, 14 p. 2002.

[12] Kustas et al, "A Two-Source Energy Balance Approach Using Directional Radiometric Temperature Observations for Sparse Canopy Covered Surfaces", Agronomy Journal, 92, 1999, pp. 847-854.

[13] Kustas et al, "Utility of radiometric-aerodynamic temperature relations for heat flux estimation", Bound.-Lay. Meteorol, 122, pp.167–187, 2007.

[14] McNaughton. K. G, and T. W. Spriggs, "An evaluation of the Priestley and Taylor equation and the complimentary relationship using results from a mixed-layer model of the convective boundary layer", T. A. Black, D. L, 1987, pp.89-104.

[15] Nearing AM, Deer- Ascough LA, Laflen JM, "Sensitivity analysis of the WEPP hillslope profile erosion model". Trans. ASAE 33 (3), 1990, p p. 839–849.

[16] Norman L, J. M, Kustas, W. P, and Humes, K. S. "A two-source approach for estimating soil and vegetation energy fluxes in observations of directional radiometric surface temperature", Agric. For. Meteorol, pp.77, 263-293.

[17] Norman et al, "Source approach for estimating soil and vegetation energy fluxes in observations of directional radiometric surface temperature", Agricultural and Forest Meteorology 77, 1995, pp. 263-293

[18] Paulson, C.A, "The mathematical representation of wind speed and temperature profiles in the unstable atmospheric surface layer", J. Appl. Meteorol, 9, 1970, pp. 857-861.

[19] Priestley, C. H. B, and Taylor. R. J, "On the assessment of surface heat flux and evaporation using large-scale parameters", Mon. Weather Rev, 100, 1972, pp. 81-92.

[20] Rody Félix, Dimitri Xanthoulis, "Analyse de sensibilité du modèle mathématique "Erosion Productivity Impact Calculator" (EPIC) par l'approche One-Factor-At-A-Time (OAT) " 2005.

[21] Ratto M, Lodi G, Costa P, "Sensitivity analysis of a fixed bed gas-solid TSA: the problem of design with uncertain models", Sep. Technol, 6, 1996, pp. 235–245.

[22] Saltelli et al, "Sensitivity Analysis", New York, John Wiley & Sons publishers, 2000.

[23] Sauer et al, "Measurement of heat and vapor transfer at the soil surface beneath a maize canopy using source plates", Agric. For. Meteorol, 75, 1995, pp. 161-189.

[24] Shuttleworth. W.J, and Wallace. J.S, "Evaporation from sparse canopies-an energy combination theory", Q. J. R. Meteorol. Sot., 111, 1985, pp. 839-855.

**First Author** Engineer in meteorology since 1986-2004, Chief Engineer in meteorology 2004-2011, and Chief Operating Meteorological Service 2000-2011, current research is about estimation of fire forest risk using water stress mapping and meteorological data.

**Second Author** received his Master of Science and his Ph.D. degrees from the University of Nancy France respectively in 1986 and 1989. In 2006, he received the HDR in Applied Mathematics from the University of Cadi Ayyad, Morocco. He is currently Professor of modeling and scientific computing at the Faculty of Sciences and Technology of Marrakech. His research is geared towards non-linear mathematical models and their analysis and digital processing applications.

## Figures

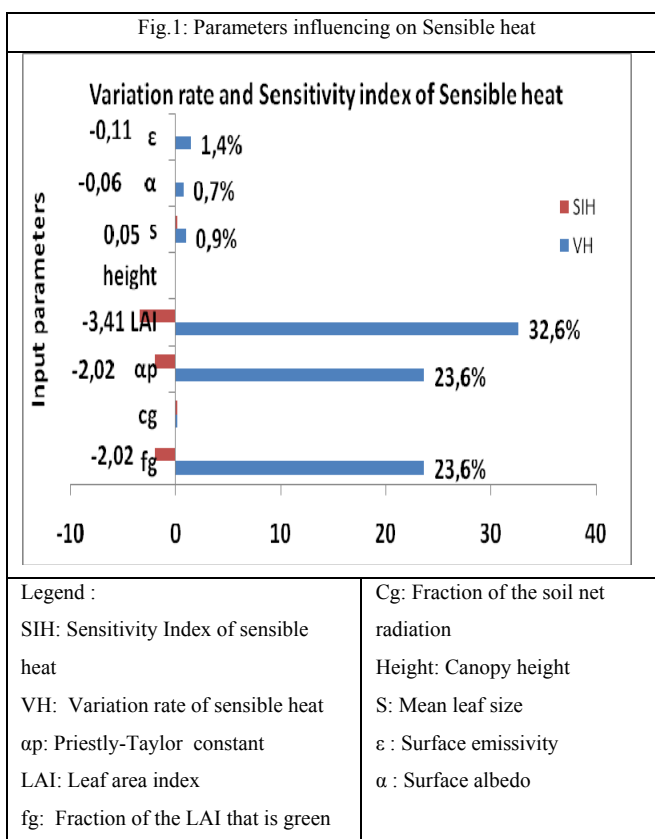
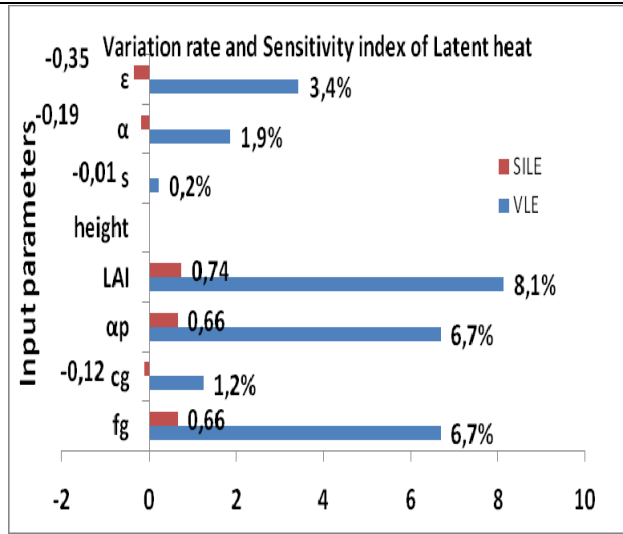


Fig.2: Parameters influencing on Latent heat

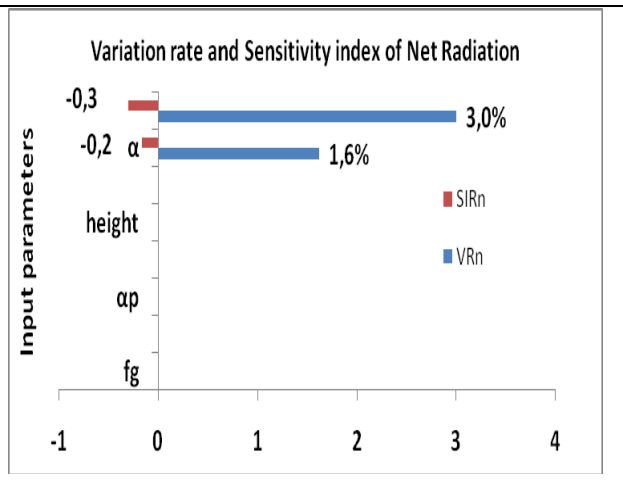


Legend :

SILE: Sensitivity Index on Latent heat  
 VLE: Variation rate on Latent heat  
 ap: Priestly-Taylor constant  
 LAI: Leaf area index  
 fg: Fraction of the LAI that is green

Cg: Fraction of the soil net radiation  
 Height: Canopy height  
 S: Mean leaf size  
 $\epsilon$  : Surface emissivity  
 $\alpha$  : Surface albedo

Fig.3: Parameters influencing on Net Radiation

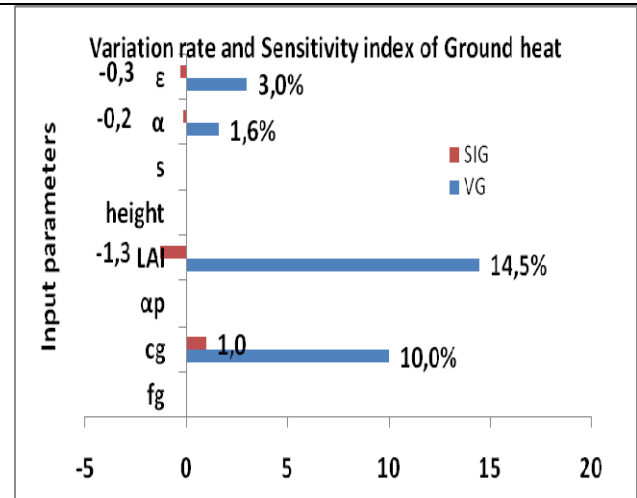


Legend :

SIRn: Sensitivity Index on Net Radiation  
 VRn: Variation rate on Net Radiation  
 ap: Priestly-Taylor constant  
 LAI: Leaf area index  
 fg: Fraction of the LAI that is green

Cg: Fraction of the soil net radiation  
 Height: Canopy height  
 S: Mean leaf size  
 $\epsilon$  : Surface emissivity  
 $\alpha$  : Surface albedo

Fig.4: Parameters influencing on soil conduction heat

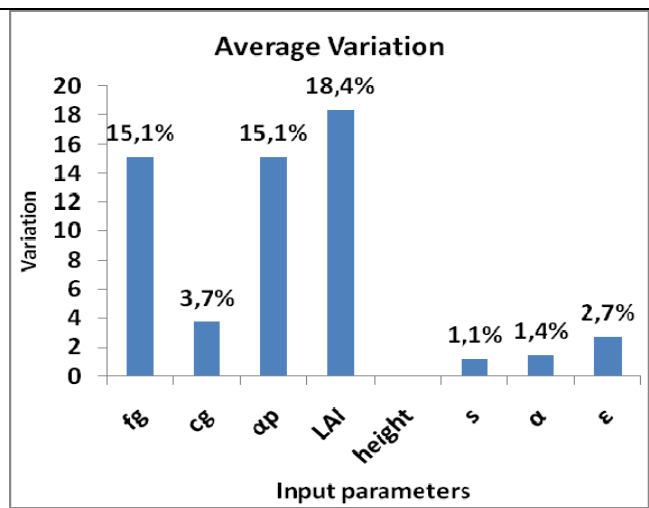


Legend :

SIG: Sensitivity Index on soil conduction heat  
 VG: Variation rate on soil conduction heat  
 ap: Priestly-Taylor constant  
 LAI: Leaf area index  
 fg: Fraction of the LAI that is green

Cg: Fraction of the soil net radiation  
 Height: Canopy height  
 S: Mean leaf size  
 $\epsilon$  : Surface emissivity  
 $\alpha$  : Surface albedo

Fig.5: Global average variation of TSEB outputs



Legend :

ap: Priestly-Taylor constant  
 LAI: Leaf area index  
 fg: Fraction of the LAI that is green

Cg: Fraction of the soil net radiation  
 Height: Canopy height  
 S: Mean leaf size  
 ε : Surface emissivity  
 α : Surface albedo

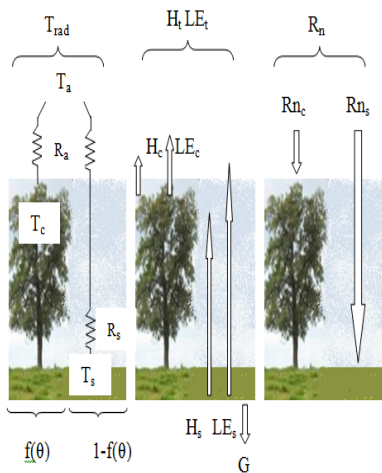


Fig.6: Scheme of resistances and flux partitioning between soil and canopy, corresponding to the TSEB parallel network

### TSEB Algorithm

(Instable Conditions)

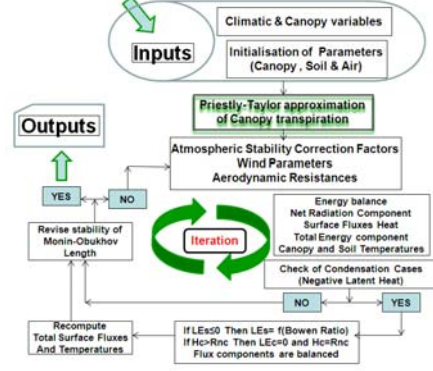


Fig.7: Algorithm of TSEB Model

Table.1: 11 Unknowns Variables of TSEB Model and associated formulae

Unknown variable	Formula
Rn	$R_n = (1 - \alpha_s) \cdot SW + \epsilon_s \cdot LW - \epsilon_s \cdot \sigma \cdot Trad_4$
Rns	$R_{ns} = R_n \exp(0.9 \ln(1 - f_c))$
Rnc	$R_{nc} = R_n - R_{ns}$
G	$G = c_g R_{ns}$
Hc	$H_c = R_{nc} - LE_c$
Hs	$H_s = \rho C_p \left[ \frac{T_c - T_a}{R_a + R_s} \right]$
LEc	$LE_c = R_{nc} \cdot C_p \cdot f_g \cdot \frac{H_c}{\Delta T}$
LEs	$LE_s = R_{ns} - H_s - G$
Tc	$H_c = \rho C_p \left[ \frac{T_c - T_a}{R_a} \right]$
Ts	$Trad(\theta) = [f(\theta) \cdot T_{c4} + (1 - f(\theta)) \cdot T_{s4}]^{1/4}$
fc	$f_c = 1 - \exp(-0.5 LAI)$

# Power Efficient Higher Order Sliding Mode Control of SR Motor for Speed Control Applications

Muhammad Rafiq<sup>1</sup>, Saeed-ur-Rehman<sup>2</sup>, Fazal-ur-Rehman<sup>1</sup>, Qarab Raza<sup>2</sup>

<sup>1</sup>Muhammad Ali Jinnah University,  
Islamabad, Pakistan.

<sup>2</sup>Centre for Advanced Studies in Engineering (CASE),  
Islamabad, Pakistan.

## Abstract.

This paper presents a novel scheme for speed regulation/tracking of Switched Reluctance (SR) motors based on Higher-Order Sliding-Mode technique. In particular, a Second-Order Sliding-Mode Controller (SOSMC) based on Super Twisting algorithm is developed. Owing to the peculiar structural properties of SRM, torque produced by each motor phase is a function of phase current as well as rotor position. More importantly, unlike many other motors the polarity of the phase torque in SR motors is solely determined by the rotor position and is independent of the polarity of the applied voltage or phase current. The proposed controller takes advantage of this property and incorporates a commutation scheme which, at any time instant, selects only those motor phases for the computation of control law, which can contribute torque of the desired polarity at that instant. This feature helps in achieving the desired speed regulation/tracking objective in a power efficient manner as control efforts are applied through selective phases and counterproductive phases are left un-energized. This approach also minimizes the power loss in the motor windings thus reducing the heat generation within the motor. In order to highlight the advantages of Higher-Order Sliding-Mode controllers, a classical First-Order Sliding-Mode controller (FOSMC) is also developed and applied to the same system. The comparison of the two schemes shows much reduced chattering in case of SOSMC. The performance of the proposed SOSMC controller for speed regulation is also compared with that of another sliding mode speed controller published in the literature.

**Keywords**— *SR motor, sliding mode control, higher order sliding mode control, commutation, speed regulation/tracking control*

## 1. Introduction

Switched reluctance motors have received considerable attention among the researchers due to its simple construction, rugged mechanical structure, and low cost driver electronics. Because of the absence of any windings on the rotor, SR motor is very suitable for operations at high speed and/or at high temperatures [2]. SR motor is doubly salient machine, i.e. both stator and rotor have salient poles on their laminations. Torque is developed in the motor when

rotor poles align with the excited stator poles. Due to this particular nature of torque production, the phase torque is independent of the polarity of phase current and depends only upon the relative position of the rotor poles with respect to the excited phase poles. For this reason, low cost unipolar power converters are used to drive SR motors. This fact also leads to a very important feature peculiar to this motor, i.e. unlike most of the other types of electrical motors, not all the phases of SR motor can produce the torque of the same polarity at any given rotor position. For example, in a 3-phase SR motor, there are certain rotor positions where only one phase can contribute the torque of the desired polarity whereas the torques produced by the other two phases are of opposite polarity. Thus energizing all 3-phases would lead to reduction in the net motor torque because of the cancellation among the phase torques.

SR motors are usually operated in magnetic saturation to increase its output torque. Magnetic saturation and mechanical saliencies in SR motors make phase torque a highly non-linear function of phase current and rotor position. Due to advancements in control theory, many nonlinear control techniques such as artificial neural network, feedback linearization, sliding mode, back stepping, fuzzy logic, etc. have been explored in the literature for the control of SR motors. Hajatipour and Farrokhi [3] developed an adaptive intelligent control based on Lyapunov functions. The proposed technique consists of two components; the first one approximates the load-torque, error in the moment of inertia and the coefficient of friction, the second component drives the system output to track the desired value. The speed controller does not require exact motor parameters and is shown to be robust against disturbances and uncertainties. Neural network torque estimator is used as a second controller in the proposed technique for torque ripple reduction. In [4], artificial neural network technique was also adopted in designing the speed controller of SR motor for regulation problem. The performance of the pro-

posed controller was shown better than fuzzy logic and fuzzy logic PI controllers.

Sliding-Mode control has been gaining popularity in control application due to its simple structure, inherent robustness and capability to control nonlinear systems [5]. John and Eastham [6], and Forrai et al [7] have used sliding-mode control for SR motor to control speed but their research did not account for magnetic saturation of the motor. Sahoo et al. [8] has applied sliding-mode technique for direct torque control of SR motor. Nihat Inanc and Veysel Ozbular [9] proposed sliding mode control to minimize torque ripples in SR motor. The proposed controller was then used for speed regulation problem and its performance was compared with conventional PI and fuzzy controllers. It was shown that the proposed controller works well for reduction of low frequency oscillations. Dynamic sliding mode controller (DSMC) has also been developed for SR motors [10]. The performance of DSMC has been compared with the conventional sliding mode controller. Both these controllers were shown to be robust against parameter and load torque variations; DSMC, however, had the advantage of reduced chattering. Chiag et al. [11] has applied sliding-mode control on synchronous reluctance motor for speed regulation problem. Tahour et al. [12] used the same technique for SR motor and compared its performance with conventional PI controller. It was shown that the proposed controller outperformed the conventional one. This performance was further improved in [13] by introducing fuzzy sliding mode control in order to remove chattering. The proposed scheme provides good transient response. Chen et al. [14] used the idea of Gaussian radial basis function neural network and developed sliding mode controller for synchronous reluctance motor. The proposed technique was based on Lyapunov approach and steepest descent rule. With this technique, the chattering problem can be reduced.

The conventional sliding mode technique has a chattering problem that can be evaded by introducing higher order sliding mode (HOSM) control [15]. HOSM has been successfully applied for various engineering problems (see [16]-[19]; for example). Rain et al. [20] developed and implemented a novel current controller for SR motor using HOSM technique for position control problem. The proposed algorithm shows good dynamic response in handling parametric uncertainties and external disturbance. A similar work was also reported in ([21]-[22]) for stepper motor. To compensate uncertainties and modeling inaccuracies, an integral term was also augmented in the proposed scheme that was shown to be robust against unknown disturbances and parametric variations. Rashed et al. [23] applied HOSM on induction motor to achieve chattering free and decoupled control over motor speed and flux by incorporat-

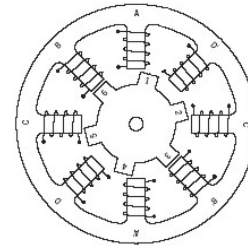


Fig. 1 Four-Phase SR motor with 8-Stator and 6-Rotor poles.

ing stator phase resistance, rotor speed and load torque estimators. To see the performance of HOSM controller clearly, a test was conducted in [1] with conventional sliding mode and PI controller on SR motor at different system parameters and unknown disturbances. It was found that HOSM has outperformed the conventional controllers with respect to chattering reduction and robustness.

Two sliding-mode controllers are developed here for speed control applications, which incorporate a commutation scheme for selectively energizing motor phases in order to achieve power efficiency. The proposed designs use less power as compared to the conventional schemes where all phases are energized. Comparison of power losses is carried out with [24] and [8].

The paper is organized as follows: Section -2 represents the dynamic model of the system, Section-3 discusses HOSM technique, Section-4 describes the important steps in controller design for regulation and tracking speed problems and Section-5 introduces the commutation scheme used in the proposed designs. Simulation results are addressed in Section-6 and finally Section-7 concludes this paper.

## 2. Mathematical model of the system

Before describing the details of controller design, the electro-mechanical model of an SR motor is described in a form suitable for the purpose. Although the proposed controllers can be developed for any SR motor with arbitrary number of phases (a four phase motor schematic is shown in Fig. 1), for clarity of presentation and subsequent simulations we consider a specific 3-phase commercial SR motor whose parameters are listed in Table-1 and its dynamic model is given by the following set of equations (see [1], [24] for a detailed explanation and derivation of the model):

$$\frac{d\theta}{dt} = \omega \quad (1)$$

$$\frac{d\omega}{dt} = \frac{1}{J} (T_e - B \omega - T_L) \quad (2)$$



$$\frac{di_j}{dt} = \left( \frac{\partial \lambda_j(\theta, i_j)}{\partial i_j} \right)^{-1} \left( u_j - R_j i_j - \omega \frac{\partial \lambda_j(\theta, i_j)}{\partial \theta} \right) \quad (3)$$

$j = 1, 2, 3$

where

$\theta$	Rotor position
$\omega$	Angular velocity of rotor
$J$	Moment of inertia (rotor)
$T_e$	Total electromagnetic torque
$B$	Coefficient of friction
$T_L$	Load torque
$i_j$	Current in the $j^{\text{th}}$ phase
$\lambda_j$	Flux linkages in $j^{\text{th}}$ phase
$u_j$	Voltages of $j^{\text{th}}$ phase.
$R_j$	Resistance to the $j^{\text{th}}$ phase

### 3. Higher Order Sliding Mode (HOSM)

The basic idea behind sliding-mode control is to define an observable function of the system states, also called switching surface, and then to design a controller in such a way that trajectories in the state space are directed towards the switching surface or the hyper plane. Once the system states reach the hyper plane, it slides along the surface towards the equilibrium point. In this technique the system's behavior remains robust to certain parameter variations and unknown disturbances [25].

The higher-order sliding-mode (HOSM) technique extends the basic idea of sliding-mode by incorporating higher order derivatives of the sliding variable. The addition of higher-order derivatives leads to a reduction in the undesirable chattering issue inherent in the sliding-mode technique while keeping the same robustness and performance as that of traditional sliding mode [26]. HOSM technique attains this quality due to the knowledge of the higher-order derivative terms of sliding variable. For example, for an  $n^{\text{th}}$  order SMC;  $s, \dot{s}, \ddot{s} \dots s^{(n-1)}$  should be known to make  $s = 0$ . To get the information about all these variables is a problem. However, this problem can be resolved with the help of super twisting algorithm.

Super-Twisting algorithm has been used for chattering reduction with systems having relative degree one. This algorithm does not demand any extra information about sliding variable and ensures that system trajectories twist around the origin in the phase portrait as shown in Fig. 2. This property makes it prominent to the other algorithms. Super twisting algorithm has been successfully applied and implemented on various engineering applications. Derafa et al. [27] used super twisting algorithm for altitude tracking of four rotors helicopter. The simulation results show that

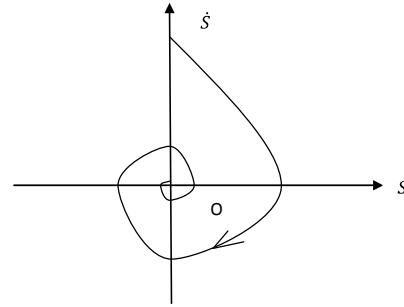


Fig. 2 Evolution of Switching Surface during the super twisting controller action; minimizing the error between reference signal and desired output signal.

the proposed scheme performs well for stabilization, robustness and tracking, in addition of chattering reduction. In [28] super twisting algorithm was employed for fault detection and isolation problem of three tank system. It is verified through simulation results that the proposed algorithm can significantly reduce the estimation error as well as chattering. In [29], the performance of super twisting algorithm was tested with other algorithm, so called twisting algorithm on dc drive under uncertain parameters and load conditions. It was claimed that super twisting algorithm, is best suited for real experiments subject to certain conditions. The super twisting algorithm based sliding mode observer was also investigated in [30] for sensorless speed control of permanent magnet synchronous motor. The simulation results show that the proposed scheme behaves well at low speed and is robust. The further detail of super twisting algorithm design is given in [1].

### 4. Controllers Design

A speed controller is designed to minimize the speed error, i.e.

$$e(t) = \omega(t) - \omega_{ref}(t) \quad (4)$$

where  $\omega_{ref}(t)$  is the desired speed. In this section, we develop speed controllers based on sliding-mode technique. The starting point is to define a sliding surface, which in our case is taken to be

$$s = \dot{e} + \lambda e \quad (5)$$

An appropriate candidate for Lyapunov function is taken as  $V = \frac{1}{2} s^2$  which would yield  $\dot{V} = s \dot{s}$  on differentiation where

$$\dot{s} = \ddot{e} + \lambda \dot{e} \quad (6)$$

$$\dot{s} = \ddot{\omega}(t) + \lambda \dot{\omega}(t) - (\ddot{\omega}_{ref}(t) + \lambda \dot{\omega}_{ref}(t)) \quad (7)$$

$$\dot{V} = s \left( \ddot{\omega}(t) + \lambda \dot{\omega}(t) - (\ddot{\omega}_{ref}(t) + \lambda \dot{\omega}_{ref}(t)) \right) \quad (8)$$

In order to find conditions which would guarantee  $\dot{V} < 0$ , we begin by differentiating Eq. (2)

$$\ddot{\omega} = \frac{1}{J} \left( \frac{dT_e}{dt} - B \frac{d\omega}{dt} - \frac{dT_L}{dt} \right) \quad (9)$$

$$\ddot{\omega} = \frac{1}{J} \left( \sum_{j=1}^3 \frac{dT_j(\theta, i_j)}{dt} - B \dot{\omega} - \frac{dT_L}{dt} \right) \quad (10)$$

$$\dot{\omega} = \frac{1}{J} \left( \sum_{j=1}^3 \frac{\partial T_j(\theta, i_j)}{\partial i_j} \frac{di_j}{dt} + \omega \sum_{j=1}^3 \frac{\partial T_j(\theta, i_j)}{\partial \theta} - B \dot{\omega} - \frac{dT_L}{dt} \right) \quad (11)$$

Substituting (3) into (11) leads to:

$$\begin{aligned} \ddot{\omega} = & \frac{1}{J} \left( \sum_{j=1}^3 \frac{\partial T_j(\theta, i_j)}{\partial i_j} \left( \frac{\partial \lambda_j(\theta, i_j)}{\partial i_j} \right)^{-1} \left( u_j - \right. \right. \\ & \left. \left. R_j i_j - \omega \frac{\partial \lambda_j(\theta, i_j)}{\partial \theta} \right) + \omega \sum_{j=1}^3 \frac{\partial T_j(\theta, i_j)}{\partial \theta} - B \dot{\omega} - \frac{dT_L}{dt} \right) \end{aligned} \quad (12)$$

Which can be written in the following form suitable for the design of our proposed controllers discussed in the following sections:

$$\begin{aligned} \ddot{\omega} = & \left( \sum_{j=1}^N \frac{\partial T_j(\theta, i_j)}{\partial i_j} \left( \frac{\partial \lambda_j(\theta, i_j)}{\partial i_j} \right)^{-1} \left( -R_j i_j - \omega \frac{\partial \lambda_j(\theta, i_j)}{\partial \theta} \right) + \right. \\ & \left. \omega \sum_{j=1}^N \frac{\partial T_j(\theta, i_j)}{\partial \theta} - B \dot{\omega} - \frac{dT_L}{dt} \right) + \\ & \frac{1}{J} \left( \left( \sum_{j=1}^N \frac{\partial T_j(\theta, i_j)}{\partial i_j} \left( \frac{\partial \lambda_j(\theta, i_j)}{\partial i_j} \right)^{-1} \right) u_j \right) \end{aligned} \quad (13)$$

Which can simply be written in a compact form as:

$$\ddot{\omega} = f(\theta, \omega, i, B, T_L) + g(\theta, i, \omega) u \quad (14)$$

where  $u$  represents the input vector comprising of  $N$  phase voltages,  $N$  represents the number of phases which are being energized at a particular instant to produce net torque and will be determined through the commutation scheme described in the next section. The scalar function  $f$  and vector function  $g$  are defined as:

$$\begin{aligned} f = & \frac{1}{J} \left( \sum_{j=1}^3 \frac{\partial T_j(\theta, i_j)}{\partial i_j} \left( \frac{\partial \lambda_j(\theta, i_j)}{\partial i_j} \right)^{-1} \left( -R_j i_j - \omega \frac{\partial \lambda_j(\theta, i_j)}{\partial \theta} \right) \right. \\ & \left. + \omega \sum_{j=1}^3 \frac{\partial T_j(\theta, i_j)}{\partial \theta} - B \dot{\omega} - \frac{dT_L}{dt} \right) \end{aligned} \quad (15)$$

$$\text{and } g = \frac{1}{J} \sum_{j=1}^3 \frac{\partial T_j(\theta, i_j)}{\partial i_j} \left( \frac{\partial \lambda_j(\theta, i_j)}{\partial i_j} \right)^{-1} \quad (16)$$

For simplicity, the explicit dependence of  $u$  on time  $t$  and  $f$  &  $g$  vectors on  $\theta, \omega, i, B, T_L$  will be omitted in the following sections. Now we consider the speed regulation and tracking problem one by one for the design of FOSMC.

#### 4.1 Case-1: Regulation Problem

The objective of the regulation problem is to stabilize the motor speed at a desired constant value. i.e.  $\omega_{ref}(t) = \omega_{ref}$  and  $\dot{\omega}_{ref}(t) = 0$ . For proving that the proposed control law guarantees the constant speed requirement, first consider the Lemma 1.

**Lemma 1:** The following control law will stabilize the motor speed to its desired value when  $t \rightarrow \infty$ .

$$u = -\frac{1}{g} (f + \lambda \dot{\omega} + K \text{sign}(s)) \quad (17)$$

**Proof:** Substituting Eq. (14) in Eq. (8), the following expression is obtained.

$$\dot{V} = s(f + gu + \lambda \dot{\omega}) \quad (18)$$

Now plugging in Eq. (17) in Eq. (18), we get

$$\dot{V} = s(f - f - \lambda \dot{\omega} - K \text{sign}(s) + \lambda \dot{\omega}) \quad (19)$$

$$\text{Then } \dot{V} = -K s \text{sign}(s) < 0 \quad (20)$$

As it is clear from Eq. (20) that  $\dot{V} = 0$  only when  $s = 0$ . This ensures that the control law as defined in Eq. (17) would guarantee that  $\omega(t) \rightarrow \omega_{ref}$  when  $t \rightarrow \infty$

#### 4.2 Case-2: Tracking Problem

The aim of tracking problem is to follow the time varying reference signal minimizing the tracking error. To prove that the proposed control law will track the reference signal, consider the Lemma 2.

**Lemma 2:** The following control law will ensure that the speed will follow a time varying reference signal as  $t \rightarrow \infty$ .

$$u = -\frac{1}{g} (f + \lambda \dot{\omega}(t) + K \text{sign}(s) - (\ddot{\omega}_{ref}(t) + \lambda \dot{\omega}_{ref}(t))) \quad (21)$$

**Proof:** Combining Eq. (8) and Eq. (14), the following is obtained.

$$\dot{V} = s(f + gu + \lambda \dot{\omega}(t) - (\ddot{\omega}_{ref}(t) + \lambda \dot{\omega}_{ref}(t))) \quad (22)$$

Substituting Eq. (21) in Eq. (22) yields

$$\dot{V} = -K s \text{sign}(s) < 0 \quad (23)$$

From Eq. (23), it is obvious that,  $\dot{V} = 0$  would be zero only when  $s = 0$ . This ensures that the control law defined in Eq. (21) would guarantee that the motor speed follows the time-varying reference signal in the limit. In both the above cases, it is shown that the Lyapunov function  $V$  is positive definite and its time derivative  $\dot{V}$  is negative definite, hence decaying and therefore the control law  $u$  will guarantee that  $\omega(t) \rightarrow \omega_{ref}(t)$  as  $t \rightarrow \infty$ .

Now we proceed with the design of SOSMC based on super-twisting algorithm as discussed in section-3. Finally, the control law after incorporating the super-twisting algorithm takes the following form for speed regulation case:

$$u = -\frac{1}{g} (f + \lambda \dot{\omega}(t) + K |s|^{0.5} \text{sign}(s)) + u_a \quad (24)$$

$$\dot{u}_a = -K \text{sign}(s) \quad (25)$$

and for speed tracking, it becomes

$$u = -\frac{1}{g} \left( f + \lambda \dot{\omega}(t) + K |s|^{0.5} \text{sign}(s) - (\ddot{\omega}_{ref}(t) + \lambda \dot{\omega}_{ref}(t)) \right) + u_a \quad (26)$$

$$\dot{u}_a = -K \text{sign}(s) \quad (27)$$

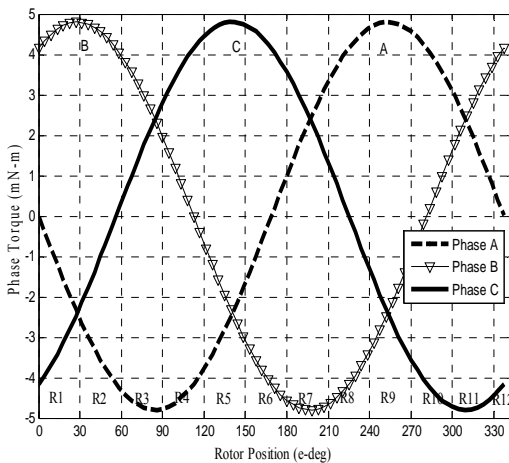


Fig. 3 Division of one electrical cycle into 12 distinct regions for commutation purposes.

## 5. Commutation Scheme

While applying Eq. (17), and Eq. (21) for FOSMC and Eq. (24), Eq. (25) and Eq. (26), Eq. (27) for SOSMC in order to compute the control laws at each instant, in terms of phase voltages, the elements of control vector  $u$  will be selected so as to engage only those phases which can contribute torque of the desired polarity. The selection of the appropriate motor phases at each instant depends upon the rotor position and the sign of the speed error at that instant. The commutation scheme which does this phase selection is now being explained and also shown in Fig. 3.

Although the phase torque of SR motor is a complex non-linear function of rotor position and phase current as explained earlier, for ease in explaining the commutation scheme we would refer to Fig. 3 which shows the phase torques of a 3-phase SR motor at a specific value of phase currents, ignoring the effects of magnetic saturation and

spatial harmonics. The figure shows the variation of phase torques as a function of rotor position within one electrical cycle. The complete electrical cycle is divided into 12 distinct regions R1-R12 such that in each region, specific phase(s) can only produce positive torque whereas the remaining phase(s) can contribute only negative torque. For example, in region R1, positive torque is only produced via phase B, while phase C and phase A when energized provide only negative torques. Similarly in region R3, phases B and C provide positive torque while negative torque can only be produced by energizing phase A.

Thus if the rotor position lies in R3 at a specific time instant and positive torque is required to reduce the speed error, only phases B and C should be used to compute the control law using Eq. (24)-Eq. (27). This would lead to achieving the desired net torque using lower voltage levels and with reduced copper losses in the motor windings. Had we energized all the motor phases, not only that phase A would have generated counterproductive torque, other phases would had to produce higher than the required values of torques to cancel the opposing torque produced by phase A. This would have led to applying higher phase voltages and an increase in copper losses too. On the other hand, if negative torque is required in this region to reduce the speed error, then only phase A should be energized. There is no need to energize phases B and C because they can only produce positive torques in this region, which would be counterproductive.

A similar approach is adopted in all these regions, which suggests that for a 3-phase SR motor, only one or at the most two phases can produce the desired polarity torque at any instant depending upon the current rotor position. Thus a judicious choice of the phases to be used in computing the control law as in Eq. (17), Eq. (21) and Eq. (24)-(27) would result in saving net power leading to increased system efficiency.

## 6. Simulation Results and Discussion

The effectiveness of the proposed controllers is evaluated by simulations carried out using MATLAB/SIMULINK software. The parameters of SR motor used for simulations are given as:

No of phases=3, No. of stator poles= 6,  
 No. of rotor poles=8, Rotor inertia (J) =0.1 N.ms<sup>2</sup>  
 Phase Resistance =4.7  $\Omega$  DC Voltage Supply =250 V  
 Coefficient of friction (B) =0.1 N.ms

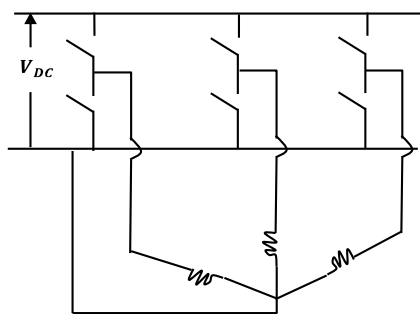


Fig. 4 Driver circuit used for energizing motor phases.

A schematic of driver electronics used to drive the motor phases is shown in Fig. 4, which uses only one leg of the H-bridge as our proposed controllers require only positive phase voltage. The FOSMC and SOSMC designed in Section-4 along with the commutation scheme developed in Section-5 are applied for speed regulation as well as speed tracking problems.

For comparison purposes, the sliding-mode controller of [10] is also implemented. Simulation results are presented in Fig 5 to Fig. 16. A number of advantageous features of FOSMC and SOSMC with the designed commutation

scheme are elaborated and compared to the conventional control; the latter can also be seen in [10] and [23].

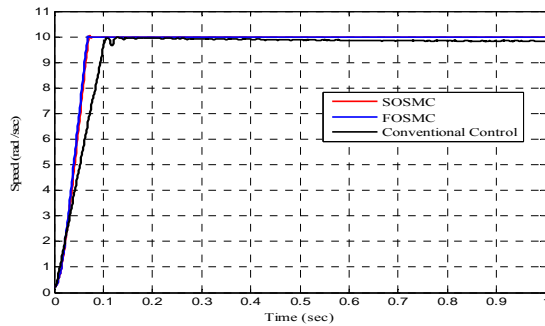


Figure 5: Speed Response of FOSMC and SOSMC to a step command.

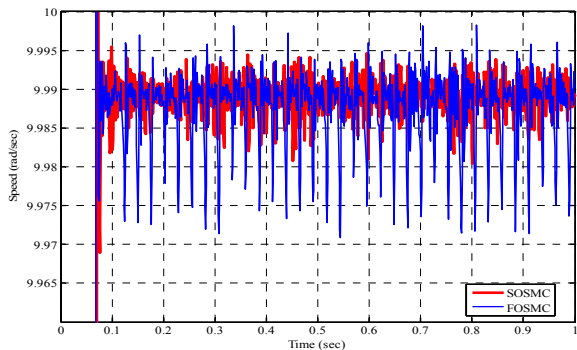


Fig 6: A Close-up View of Response of both FOSMC and SOSMC to a step command. The high magnitude of chattering signal of FOSMC is clearly noticeable.

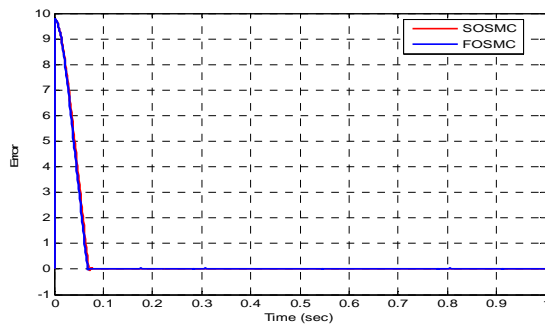


Fig 7: Error plot of Speed Response of FOSMC and SOSMC to a step command

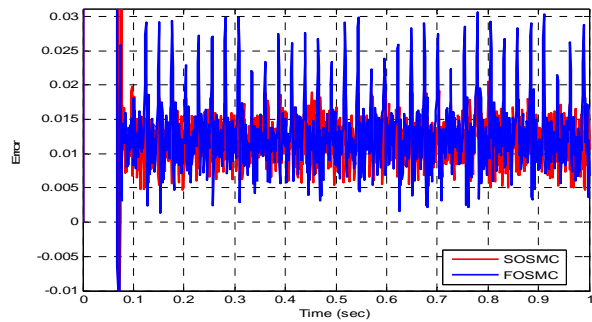


Fig 8: A Close-up View of Error plot of Speed Response for FOSMC and SOSMC to a step command. The reduced amount of error magnitude is clearly visible

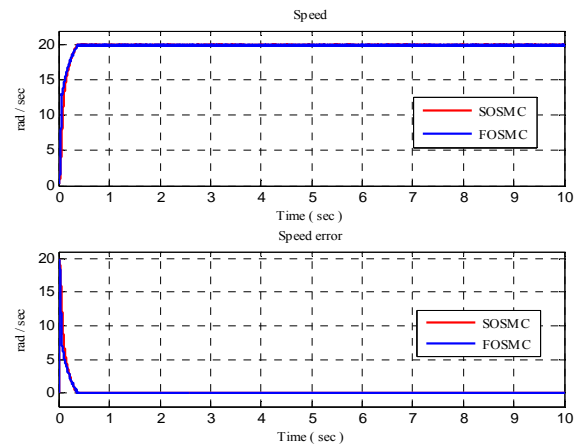


Fig 9: Speed Response and error plot of FOSMC and SOSMC to a step command for a reference speed of 20 rad/s.

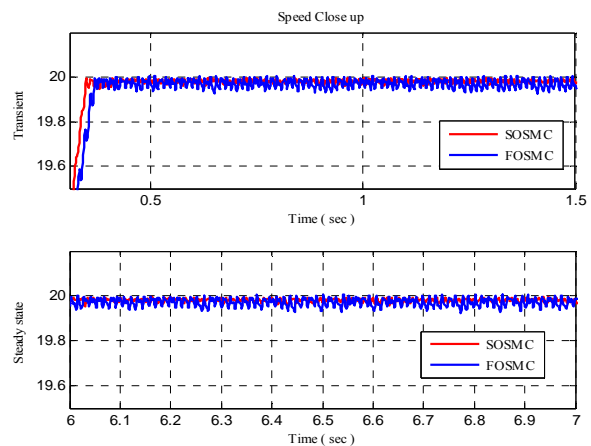


Fig 10: A Close-up View of Response of both FOSMC and SOSMC to a step command in the starting and steady state regions.. The high magnitude of chattering signal of FOSMC is clearly noticeable.

Fig. 5 and Fig. 6 compare the outputs of FOSM & SOSM controllers developed in this paper with that of another sliding mode controller reported in [10] for the case when motor is commanded to move at 10 rad/s from rest. As it is clear, the commutation based controllers converge the motor speed more quickly to the desired value. Fig. 7 and Fig. 8 show a performance test for FOSMC and SOSMC for a reference speed of 10 rad/s. It can be visualized that FOSMC shows higher magnitude of chattering than SOSMC. The same results are verified for a motor speed of 20 rad/s in next simulation tests shown in Fig. 9 and Fig 10.

A comparison of power loss in motor phases during operation is shown in Fig. 11. Power loss in conventional design is about 85 kW whereas even the FOSMC which suffers from the same level of chattering has much reduced power loss, i.e. only 19 KW.

This power saving can be mainly attributed to the commutation scheme employed in FOSMC. SOSMC with its reduced level of chattering reduces the power loss further to 15 kW which confirms the effectiveness of the proposed design. Fig. 12 and Fig. 13 highlight the main reason behind this power savings (area under the curve, which is less for SOSMC).

Fig. 12 shows the three phase voltages during initial stage of steady state operation. It is well clear from these figures that in commutation based controllers; only one or two motor phases are selected for generation of control efforts at any given instant of time. The conventional design, on the other hand, energizes all the three phases simultaneously and applies bipolar voltages to motor phases. A closer focus on the time interval 0.44 - 0.45 sec is of particular interest. It shows that even in those cases where apparently only two of the three phases are being energized by the conventional sliding-mode controller, the controller has selected wrong phases for the generation of control efforts. Despite that maximum voltages are being applied to the two phases resulting in large phase currents, the torques produced by the two phases are cancelling each other. This results in much reduced net motor torque as compared to the torques produced by each phase independently. This amounts to wastage of efforts and also results in increased power loss in motor windings.

The commutation based FOSMC and SOSMC use only unipolar voltages with reduced voltage levels thus resulting in lower phase currents. As a result, proposed controllers (FOSMC and HOSMC) produce lesser individual torques of the same polarities which add up to give a higher net torque. The torques produced by three individual phases and net torque are shown in Fig. 14 which verify this.

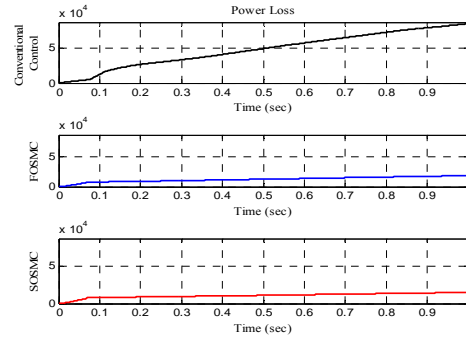


Figure 11: Power loss in Conventional Design is about 85 kW. Using FOSMC it is about 19 KW. Using SOSMC it even lowers to about 15 kW.

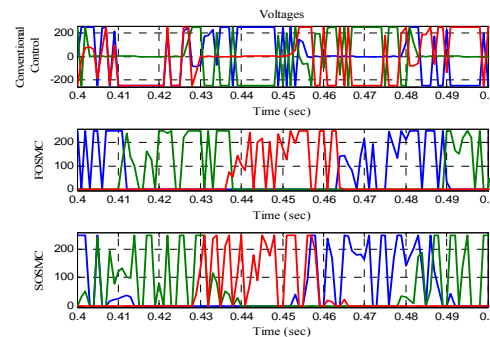


Figure 12: 3-phase voltages during initial stage of steady state response.

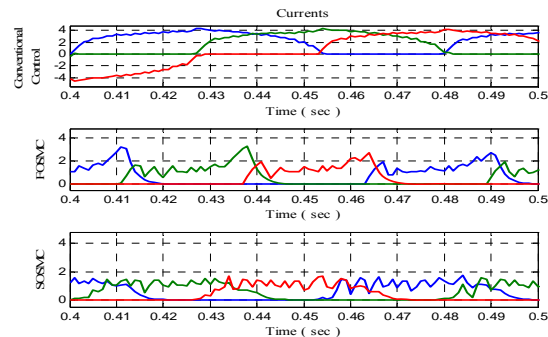


Figure 13: 3-phase currents during initial stage of steady state response.

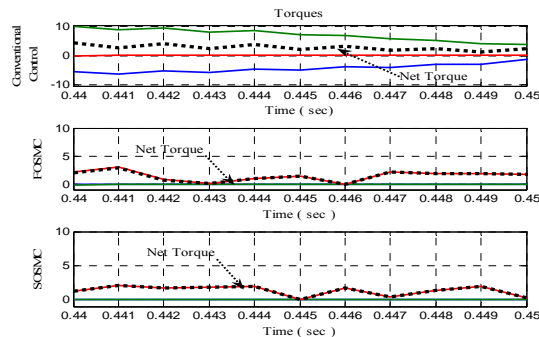


Fig. 14 Torques during initial stage of steady state response.



Tracking performance of FOSMC and SOSMC can be reflected from Fig. 15 where sinusoidal signal is selected for comparison test. It can be seen that SOSMC is exhibiting less amount of chattering and smaller spikes than FOSMC. Another good performance of SOSMC can be shown from Fig. 16 when SR motor experiences a sudden change in external load driven by the motor. The external load varies from 0 to 1.5 Nm, 0 to 2 Nm and 0 to 2.5 Nm during the intervals  $t=3$  to  $t=3.1$  second,  $t=5$  to  $t=5.1$  second and  $t=7$  to  $t=7.1$  second respectively. It can be seen that despite a sudden change in external load, the SOSMC does not allow a bigger dip and keeps the motor closer to its desired speed. The results of these simulations clearly indicate that commutation scheme based sliding-mode controllers developed in this paper show promising results. These results are good enough to establish the fidelity of both designs in tracking as well as regulation applications. A selection out of these two schemes would depend upon a number of factors, some of which are highlighted below:

- The magnitude of error a designer can safely tolerate.
- The effect of chattering on the actuator action.
- The actuator safety while dealing with chattering in the actuation signal.

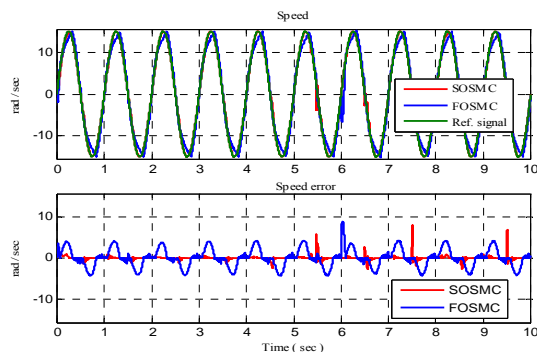


Fig. 15 Speed response of proposed controllers while tracking a reference signal given by  $\omega_{ref}(t) = 15 \sin 2\pi t$ . The lower plot shows a close up to elaborate the performance of both controllers.

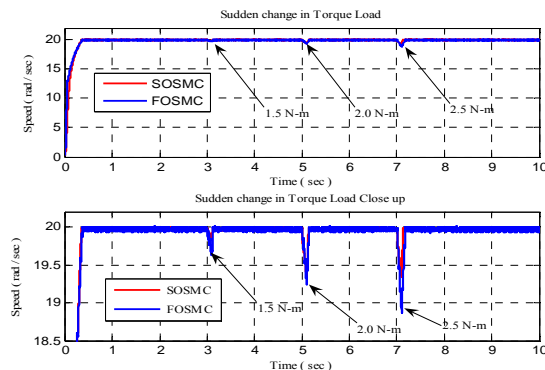


Fig. 16 Speed response of proposed controllers against sudden change in torque load.

- The natural frequency of actuator and the frequency and magnitude of chattering, etc.

## 7. Conclusion

First-order and Second-order sliding-mode controllers have been developed for speed (regulation/tracking) control of SR motors. Contrary to the conventional sliding-mode controllers developed for SR motors, the proposed controllers use a designed commutation scheme which uses only those motor phases for the computation of control law, at any given instant, which can produce torque of the desired polarity. Second-order sliding-mode controller (SOSMC) is shown to be more effective in terms of accuracy and reduced amount of chattering than First-order sliding-mode controller (FOSMC). Both the controllers are shown to be power efficient and also result in reduced power loss in motor windings leading to reduced heat generation.

## Acknowledgement

This research was financially supported by Higher Education Commission (HEC) Pakistan.

## References

- [1] M. Rafiq, S.U Rehman, F. R. Rehman, Q.A. Butt, "Performance Comparison of PI and Sliding Mode for Speed Control Applications of SR Motor", *European Journal of Scientific Research*, Vol. 50, No. 3, pp. 368-384, 2011.
- [2] R. Krishnan, "Switched Reluctance Motor Drives: Modeling, Simulation, Analysis, Design, and Applications", *Industrial Electronic Series*, CRC Press LLC, 2001.
- [3] M. Hajatipour, M. Farrokhi, "Adaptive intelligent speed control of switched reluctance motors with torque ripple reduction", *Energy Conversion and Management*, Vol. 49, No. 5, pp. 1028-1038, 2008.
- [4] E. Karakas, S. Vardarbasi, "Speed control of SR motor by self-tuning fuzzy PI controller with artificial neural network" in *S-adhan Academy Proceedings in Engineering Sciences*, Vol. 32, No. 5, pp. 587-596, 2007.
- [5] V. I. Utkin, H. C. Chang, "Sliding Mode Control on Electro-Mechanical Systems", *Mathematical Problems in Engineering*, Vol. 5, No. (4-5), pp. 451-473, 2002.
- [6] G. John, A.R. Eastham, "Speed control of switched reluctance motor using sliding mode control strategy" in *Proc. IEEE, 13<sup>th</sup> IAS, Industry Application Conference, 1995*, Vol. 1, 1995, pp. 263 - 270.
- [7] A. Forrai, Z. Biro, V. Chiorean, "Sliding Mode Control of Switched Reluctance Motor Drive" in *Proc. IEEE, 6th International Conference on Optimization of Electrical and Electronic equipments*, Vol. 2, 1998, pp. 467 - 472.
- [8] S. K. Sahoo, S. K. Panda, J. X. Xu, "Direct Torque Controller for Switched Reluctance Motor Drive using Sliding Mode Control", in *Proc. IEEE, International Conference on Power Electronic and Drive System*, Vol. 2, 2005, pp.1129-1134.

- [9] I. Nihat, O. Veysel, "Torque ripple minimization of a switched reluctance motor by using continuous sliding mode control technique", *Electric Power Systems Research*, Vol. 66, No. 3, pp. 241-251, 2003.
- [10] M. T. Alrifai, M. Zribi, H. S. Ramirez, "Static and dynamic sliding mode control of variable reluctance motors", *International Journal of Control*, Vol. 77, pp. 1171-1188, 2004.
- [11] H. K. Chiang, C. H. Tseng, W. L. Hsu, "Implementation Of A Sliding Mode Controller For Synchronous Reluctance Motor Drive Considering Core Losses", *Journal of the Chinese Institute of Engineers*, Vol. 26, No. 1, pp. 81-86, 2003.
- [12] A. Tahour, A. Meroufel, H. Abid, A. A. Ghani, "Sliding Controller of Switched Reluctance Motor", *Leonardo Electronic Journal of Practices and Technologies*, Vol. 12, pp. 151-162, 2008.
- [13] A. Tahour, H. Abid, A. A. Ghani, "Speed Control of Switched Reluctance Motor Using Fuzzy Sliding Mode", *Advances in Electrical and . & Computer Engineering*, Vol. 8, No.1, pp.21-25, 2008.
- [14] C. A. Chen, H. K. Chiang, W. B. Lin, C.H. Tseng, "Synchronous reluctance motor speed drive using sliding mode controller based on Gaussian radial basis function neural network", in *Proc. 14<sup>th</sup> International Symposium on Artificial Life and Robotics*, Vol. 14, 2009, pp. 53-57.
- [15] A. Levant, "Higher order sliding modes and their application for controlling uncertain processes", PhD Dissertation, 1987
- [16] Q. R. Butt, A. I. Bhatti, "Estimation of Gasoline-Engine Parameters Using Higher Order Sliding Mode" *IEEE Transaction on Ind. Electronics*, Vol.55, No.11, pp.3891-3898, 2008.
- [17] S.H. Qaiser, A.I. Bhatti, Masood Iqbal, R. Samar, J. Qadir, "Model validation and higher order sliding mode controller design for a research reactor", *Annal of Nuclear Energy*, Vol. 36, pp. 37-45, 2009.
- [18] Q. R. Butt, A. I. Bhatti, M. A. Iqbal, M. A. Rizvi, R. Mufti, I. H. Kazmi, "Estimation of Automotive Engine Parameters Part I: Discharge coefficient of throttle body", in *Proc. IEEE, 6<sup>th</sup> International Bhurban Conference on Applied Sciences and Technology*, 2009, pp. 275-280.
- [19] M. Iqbal, A. I. Bhatti, S. I. Ayubi, Q. Khan, "Robust Parameter Estimation of Nonlinear Systems Using Sliding-Mode Differentiator Observer", *IEEE Transactions on Industrial Electronics*, Vol. 58, No. 2, pp. 680-689, 2011.
- [20] X. Rain, M. Hilairet, R. Talj, "Second order sliding mode current controller for the switched reluctance machine", in *Proc. IEEE, 36<sup>th</sup> Annual Conference on Industrial Electronics Society*, 2010, pp. 3301-3306.
- [21] M. Defoort, F. Nollet, T. Floquet, W. Perruquetti, "Higher order sliding mode control of a stepper motor" *Proc. IEEE Conference on Decision & Control*, Vol. 1, 2006, pp.4002-4007.
- [22] M. Defoort, F. Nollet, T. Floquet, W. Perruquetti, "A Third Order Sliding Mode Control of a Stepper Motor" *IEEE transactions on Ind. Electronics*, Vol. 56, No. 9, pp.3337-3346, 2009.
- [23] M. Rashed, K.B. Goh, M.W. Dunnigan, P.F.A. McConnell, A.F. Stronach, B.W. Williams, "Sensorless second-order sliding-mode speed control of a voltage-fed induction-motor drive using nonlinear state feedback", in *Proc. IEE, Electric Power Application*, Vol. 152, No. 6. 2005, pp. 1127-1136.
- [24] M. Rafiq, S. A. Rehman, Q. R. Butt, A. I. Bhatti, "Power Efficient Sliding Mode Control of SR Motor for Speed Control Applications," in *Proc. IEEE, 13<sup>th</sup> INMIC*. pp. 1-6, 2009.
- [25] J. J. E. Slotine, L. Weiping 1991, "Applied Nonlinear Control", Prentice Hall, Englewood Cliffs, New Jersey 07632
- [26] A. Levant, "Chattering Analysis", *IEEE Transactions on Automatic Control*, Vol. 55, No.6, pp. 1380-1389, 2010.
- [27] L. Derafa, L. Fridman, A. Benallegue and A. Ouldali, "Super Twisting Control Algorithm for the Four Rotors Helicopter Attitude Tracking Problem", in *Proc. IEEE, 11<sup>th</sup> International Workshop on Variable Structure Systems*, 2010, pp. 62-67.
- [28] M. Rolink, T. Boukhobza, D. Sauter, "High Order Sliding Mode Observer For Fault Actuator Estimation And Its Application to The Three Tanks Benchmark", in *Proc. HAL*, Vol. 1, 2006, pp. 1-7.
- [29] H. Chaal, M. Jovanovic, "Second Order Sliding Mode Control of a DC Drive with Uncertain Parameters and Load Conditions", in *Proc. IEEE, Conference on Control and Decision*, 2010, pp. 3204-3208.
- [30] M. Ezzat, J. D. Leon, N. Gonzalez, A. Glumineau, "Sensorless Speed Control of Permanent Magnet Synchronous Motor using Sliding Mode Observer", *Proc. IEEE, 11<sup>th</sup> International Workshop on Variable Structure Systems*, 2010, pp. 227-232.



**Muhammad Rafiq Mufti** received M.Sc. degree in computer science from Bahauddin Zakariya University, Multan, Pakistan and M. Sc. degree in computer engineering from Centre for Advanced Studies in Engineering (CASE) Islamabad in 1994 and 2007, respectively. Currently, he is working towards his PhD degree from Mohammad Ali Jinnah University (MAJU) Islamabad. His research interests include sliding mode control, fractional control, and neural network.



**Dr. Saeed-ur-Rehman** got his PhD from GA. Tech. Atlanta, Georgia, USA. He specializes in digital control systems and power electronics. Dr. Rehman has more than 15 years of industrial and academic experience. Currently he is a professor at Centre for Advanced Studies in Engineering (CASE) Pakistan. He is also associated with CARE from where he has developed several embedded systems for ruggedized industrial/military applications. He has authored several papers and holds a US patent on sensorless motor control.



**Dr. Fazal-ur-Rehman** received his M. Sc. & M. Phil. degrees in Mathematics from B. Z. University Multan, Pakistan in 1986 and 1990, and M. Eng. & Ph.D. degrees in Control Systems from Department of Electrical Engineering, McGill University, Montreal, Canada in 1993 and 1997 respectively. He joined the Faculty of Electronic Engineering, GIK Institute of Engineering, Pakistan as an Assistant Professor in January 1998 and worked there as an Assistant Professor 1998-2002 and Associate Professor 2003-2005. Presently he is working as a Professor in Department of Electronic Engineering, Mohammad Ali Jinnah University (MAJU), Islamabad, Pakistan. Dr. Fazal's research interests are primarily in a particular area of Nonlinear Control Systems, called Nonholonomic Control Systems. He has also interest in Optimal Control and Digital Signal Processing.



**Qarab Raza** got his Bachelor's degree in Mechanical Engineering from University College of Engineering, Taxila in 1989. He received a post graduate diploma in computer system software and hardware in 1990 from Computer Training Center, Islamabad. Since 1990, he has been working in industry as an installation, fabrication and design engineer for nearly sixteen years. He got his Masters Degree in Control Engineering from Center for Advanced Studies in Engineering, Islamabad in 2004. He is first author and co-author of 10 Conference and 4 journal publications. He is currently a PhD candidate at Center for Advanced Studies in Engineering, Islamabad. His research interests are sliding mode control, fractional control, mathematical modeling of dynamic systems for control and fault diagnosis.

# Semantic Search in Wiki using HTML5 Microdata for Semantic Annotation

Pabitha P<sup>1</sup>, Vignesh Nandha Kumar K R<sup>2</sup>, Pandurangan N<sup>2</sup>, Vijayakumar R<sup>2</sup> and Rajaram M<sup>3</sup>

<sup>1</sup> Assistant Professor, Dept of Computer Technology, MIT Campus, Anna University  
Chennai 600 044, Tamilnadu, India.

<sup>2</sup> Student, Computer Science and Engineering, Anna University  
Chennai 600 044, Tamilnadu, India.

<sup>3</sup> Professor, Anna University of Technology, Tirunelveli

## Abstract

Wiki, the collaborative web authoring system makes Web a huge collection of information, as the Wiki pages are authored by anybody all over the world. These Wiki pages, if annotated semantically, will serve as a universal pool of intellectual resources that can be read by machines too. This paper presents an analytical study and implementation of making the Wiki pages semantic by using HTML5 semantic elements and annotating with microdata. And using the semantics the search module is enhanced to provide accurate results.

**Keywords:** HTML5, Microdata, Search, Semantics, Annotation, Wiki

## 1. Introduction

Wikipedia contains vast amount of information and resources. Though it provides vast amount of information, they can be only understandable only by humans. We can make them machine understandable by including the semantic contents in the wiki engine. Thereby, we can make search efficient and optimization.

## 2. Literature survey

### 2.1 Semantic web

The term semantic web coined by Tim Berners-Lee, is not a separate web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation.

Conventional web contains a large pool of information that is human readable but not interpretable by computers. Semantic web extends it by annotating the web pages with semantic description. This allows computers to retrieve information from the web automatically and to manipulate them.

### 2.2 Ontology

An ontology is the formal explicit specification of shared conceptualization. A conceptualization refers to an abstract model of some phenomenon in the world that identifies the relevant concepts of that phenomenon. Explicit means that the type of concepts used and the constraints on their use are explicitly defined. Formal refers to the fact that the ontology should be machine understandable.

### 2.3 Wiki

A wiki is a Web-based system that enables collaborative editing of Web pages. The most important properties of wikis are their openness and flexibility. Their openness lets each user participate in content creation, and their flexibility supports different users' working styles without imposing technological constraints. Wikis provide a Web-based text editor with a simple mark-up language to create content and to link easily between pages as well as a versioning system to track content changes and full-text search for querying the wiki pages.

### 2.4 Semantic wiki

A semantic wiki tries to extend a normal wiki's flexibility to address structured data. To this end, it supports metadata in the form of semantic annotations of the wiki pages themselves, they can and of the link relations between wiki pages. The annotations usually correspond to an ontology that defines the properties that can be associated with different object types.

Semantic Wiki offers:

- a simple formalism for semantically annotating links and wiki articles or other kinds of content.
- a semantic search for querying by not only keyword but also semantic relations between objects and

- possibly an additional automatic or semi-automatic extraction of metadata from wiki articles to simplify the annotation process – for example, by topic(EUprojects)or even indirectly (meeting minutes of EU projects)

### 3. HTML5

HTML5 is the 5th major revision of the core language of the World Wide Web: the Hypertext Mark-up Language (HTML), initiated and developed mainly by WHATWG (Web Hypertext Applications Technology Working Group).Started with the aim to improve HTML in the area of Web Applications, HTML5 introduces a number of semantic elementswhich include: <section>, <nav>, <article>, <aside>, <hgroup>, <header>, <footer>, <time> and <mark>.

These are some of the tags that have been introduced just to bring semantics in web pages, with no effect on the way it is displayed. They behave much like a grouping element such as <div> as far as displaying them is concerned. This means if an old browser cannot recognize these tags it will handle them much similar to the way a grouping element is handled. The semantic elements tell the browsers and web crawlers clearly the type of content contained within the element. For instance states explicitly that the figures within the element represent a time.

### 4. Microdata

Apart from the semantic elements HTML5 introduces Microdata – the way of annotating web pages with semantic metadata using just DOM attributes, rather than separate XML documents. Microdata annotates the DOM with scoped name/value pairs from custom vocabularies. Anyone can define a microdata vocabulary and start embedding custom properties in their own web pages. Every microdata vocabulary defines a set of named properties. For example, a Person vocabulary could define properties like name and photo. To include a specific microdata property on your web page, you provide the property name in a specific place. Depending on where you declare the property name, microdata has rules about how to extract the property value. Defining your own microdata vocabulary is easy. First, you need a namespace, which is just a URL. The namespace URL could actually point to a working web page, although that's not strictly required. Let's say I want to create a microdata vocabulary that describes a person. If I own the data- vocabulary.org domain, I'll use the URL <http://data-vocabulary.org/Person> as the namespace for my microdata vocabulary. That's an easy way to create a globally unique identifier: pick a URL on a domain that you control. In this vocabulary, I need to define some named properties.

Let's start with three basic properties:

- name (your full name)
- photo (a link to a picture of you)
- url (a link to a site associated with you, like a weblog or a Google profile)

Some of these properties are URLs, others are plain text. Each of them lends itself to a natural form of markup, even before you start thinking about microdata or vocabularies or whatnot. Imagine that you have a profile page or an —about page. Your name is probably marked up as a heading, like an <h1> element. Your photo is probably an <img> element, since you want people to see it. And any URLs associated your profile are probably already marked up as hyperlinks, because you want people to be able to click them. For the sake of discussion, let's say your entire profile is also wrapped in a <section> element to separate it from the rest of the page content. Thus:

```
<section itemscope itemtype= "http://data-  
vocabulary.org/Person">  
  <div itemprop="title" class="title">      President  
  </div>  
  <div itemprop="name" class="name">  
    Mark Pilgrim  
  </div>  
</section>
```

The major advantage of Microdata is its interoperability, i.e any RDF representation of an ontology can be mapped to HTML5 microdata.

### 5. Existing System

MediaWiki is a free software wiki package written in PHP, originally for use on Wikipedia. It is now used by several other projects of the non-profit Wikimedia Foundation and by many other wikis. MediaWiki is an extremely powerful, scalable software and a feature-rich wiki implementation, that uses PHP to process and display data stored in its MySQL database. Pages use MediaWiki's wiki-text format, so that users without knowledge of HTML or CSS can edit them easily.

#### 5.1 MediaWiki Architecture

In the architecture of MediaWiki as shown in Fig.1 the top two layers hardly have anything to do with semantic annotation. The layers of concern are the Logic Layer and the Data Layer; the major part lies in Logic Layer. The following figure shows the architecture of MediaWiki:



<b>User layer</b>	Web browser	
<b>Network layer</b>	Squid	
	Apache web-server	
<b>Logic layer</b>	MediaWiki's PHP scripts	
	PHP	
<b>Data layer</b>	File system	MySQL Database (program and content)

Fig. 1 Architecture of Mediawiki [12]

**Logic Layer:** This is the core part of MediaWiki that accomplishes the above said tasks. The PHP scripts of MediaWiki are to be edited to carry out these tasks. The parser module (Fig. 6) is to be enhanced to convert between Wiki and HTML markups. Also the data-vocabulary referred in the pages must be validated and appropriate flags must be set.

**Data Layer:** The MySQL database layout of Mediawiki is so normalized that adding a new table needs no alterations in any table [13]. The metadata about each page is stored in the page table, whose layout is given in Fig.2

Field Name	Field Type
page_id	INTEGER(8)
page_namespace	INTEGER(11)
page_title	VARCHAR(255)
page_restrictions	TINYBLOB
page_counter	BIGINT(20)
page_js_redirect	TINYINT(1)
page_js_new	TINYINT(1)
page_random	DOUBLE
page_touched	CHAR(14)
page_latest	INTEGER(8)
page_len	INTEGER(8)

Fig. 2 Layout of the page table

The actual content of the page is stored in a separate table named text whose layout is given in Fig.3

Field Name	Field Type
page_id	INTEGER(8)
page_namespace	INTEGER(11)
page_title	VARCHAR(255)
page_restrictions	TINYBLOB
page_counter	BIGINT(20)
page_js_redirect	TINYINT(1)
page_js_new	TINYINT(1)
page_random	DOUBLE
page_touched	CHAR(14)
page_latest	INTEGER(8)
page_len	INTEGER(8)

Fig. 3 Layout of the text table

**Wiki Parser:** Here is a sample Wiki markup:

The "[Wikimedia Foundation, Inc.](#)" is a [\[\[Non-profit organization|nonprofit\]\]](#) [\[\[Foundation \(nonprofit\)|charitable organization\]\]](#) For the Internal Revenue Service (the IRS) to recognize an organization's exemption, the organization must be organized as a trust, a corporation, or an association.

The original HTML syntax markup corresponding to this shown below:

```
<p>The <b>Wikimedia Foundation, Inc.</b> is a <a href="/wiki/Nonprofit_organization" title="Non-profit organization">non-profit</a> <a href="/wiki/Foundation_(non-profit)" title="Foundation (nonprofit)">charitable organization</a> For the Internal Revenue Service (the IRS) to recognize an organization's exemption, the organization must be organized as a trust, a corporation, or an association. </p>
```

Here, for instance, [\[\[Non-profit organization|non-profit\]\]](#) corresponds to `<a href="/wiki/Nonprofit_organization" title="Nonprofit organization">non-profit</a>`. That means the Wiki engine parses the Wiki markup entered by the author and generates the corresponding HTML markup.

## 6. Proposed system

The Wiki pages, if annotated semantically, will serve as a universal pool of intellectual resources that can be read by machines too.

Mediawiki follows a standard template for its web pages. Thus a search engine or any other software that needs data to be extracted from Wiki pages need not search the entire web page; instead it is enough to search the variable data, i.e. the contents excluding the fixed (template) part [2]. This project is to define a way of annotating the wiki pages using a simple markup similar to that already available for editing conventional wiki pages and to define a set of vocabularies to represent the relationship among Wiki pages. This involves developing a parser to parse the markup and to replace it with actual HTML5 microdata for storing and the vice-versa while editing.

A parser to recognize the Semantic Wiki mark-up and to generate the corresponding HTML5 markup has been developed [1]. Thus the project includes:

- Defining a Wiki mark-up for representing ontology
- Extending the parser for translating this to corresponding HTML5 mark-up
- Defining vocabularies that define entities related to Wiki pages

Enhancing the search engine to take advantage of the

Semantic definitions is being implemented.

To account for the semantic annotations in the pages, we add a new table **microdataobject** whose layout is:



Fig. 4 Layout of the new table **microdataobject**

**Block diagram:**

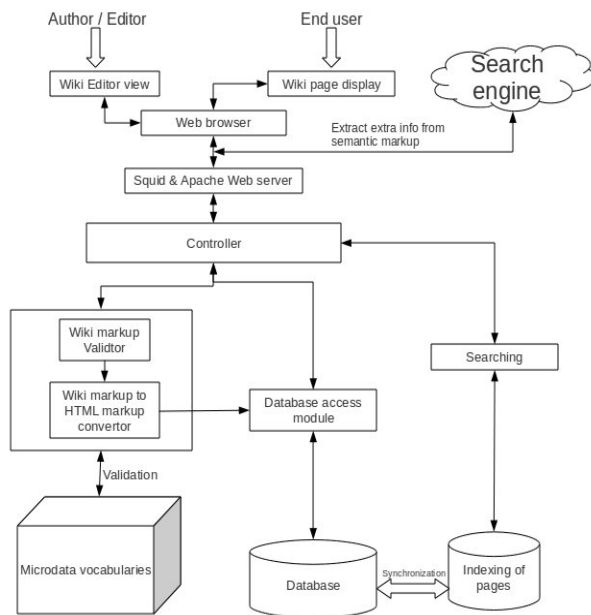


Fig. 5 Block diagram of Semantic Mediawiki

**Controller** is the module that despatches the requests from the user to the corresponding module. However the Squid (proxy server) may serve the user with cached results from previous requests.

**Microdata vocabulary** is the actual definition of the class to which the object described in the page belongs to. InHTML5 microdata this is referred to by the value of `itemtype` attribute.

**Editor** module provides the interface through which a user can edit or create wiki pages. If the user edits an already existing page, the corresponding page is fetched from the database and the HTML markup is converted into wiki markup and is displayed in the editor interface. After the user edits the contents and clicks *Save page* the modified contents are given to the parser to be converted to HTML markup.

**6.1 Parser**

Parser is the core module that is responsible for validating the wiki markup and converting it to HTML markup to be rendered as a web page. The Wiki markup may be a control markup that does not affect the content of the page – the one which updates the metadata alone, like minor edits. For such cases the parser asks the database access module to update the associated entries in the database.

A part of the newly added modules in Parser.php file:

```
function addSemantics( $text ) {
    wfProfileIn( __METHOD__ );
    $satParaStart = preg_match('/^<p>\{__:\}', $text);
    $satParaEnd = preg_match('/\}_<\p>/', $text);
    $spos = strpos($text, '{__:');
    if($spos == false)
        return $text;
    $spattern = array(
        '/(?<=\{__:\})(w+)' =>
            'http://data-vocabulary.org/'. '\1.' . "'",
        '/\{__:\}' => '<span itemscope itemtype=""',
        '/_}\}' => '</span>',
        '/(?<=@)(\w+):("[^"]*"|' =>
            '\1.' . "'>'\3'</span>',
        '/@(?=(\w+))' => '<span itemprop=""');
    if($satParaStart==1) {
        $text = preg_replace('/^<p>\{__:\}', '{__:', $text);
        $spattern['/(?<=\{__:\})(w+)' ] = 'http://data-
        vocabulary.org/'. '\1.' . "'><p>";
    }
    if($satParaEnd==1) {
        $text = preg_replace('/_}\}'<\p>/', '_}\}', $text);
        $spattern['/_}\}'] = '<p></span>'; } $text = preg_replace(
        array_keys($spattern), array_values($spattern), $text);
    wfProfileOut( __METHOD__ );
    return $text; }
```

The wiki markup to include microdata annotation is:

```
{__:ItemType
    ... @itempropName:"value" ...
__}
```

For instance, to include microdata annotation about a person, the Wiki markup is as follows:

```
{__:Person
    ... @name:"Richard Stallman" ...
```

```
... @title:"President" ...
... @nickname:"RMS"
```

```
}_
```

Here, the ellipsis are used to represent some arbitrary content, just as placeholder; not part of the syntax. This Wiki markup on passing the Parser module becomes:

```
<span itemscope
  itemprop="http://data-vocabulary/Person">
  ...<span itemprop="name">Richard Stallman</span>...
  ...<span itemprop="title">President</span>...
  ...<span itemprop="nickname">RMS</span>...
</span>
```

This approach differs from the earlier proposals of semantic wiki using RDF (such as KawaWiki [4] and Rhizome [5]) in simplicity. The user's effort to annotate a web page is reduced drastically as semantic HTML elements and attributes serve the purpose of their XML counterparts. Thus to make the e-resources most updated as well as semantic without much strain HTML5 microdata suits best.

### 6.2 Mediawiki Search module

The search module of Mediawiki is organised as one base class named SearchEngine and 6 subclasses. SearchUpdate, one of the subclasses, is to update the search index in the database whereas database specific operations are carried out by the other 5 classes, one for each of MySQL, MySQL4, PostgreSQL, SQLite, Oracle and IBM-DB2.

In the base class, some functions are just declared as stub and their actual implementation is done in the database-specific subclasses.

The class diagram is as shown below:

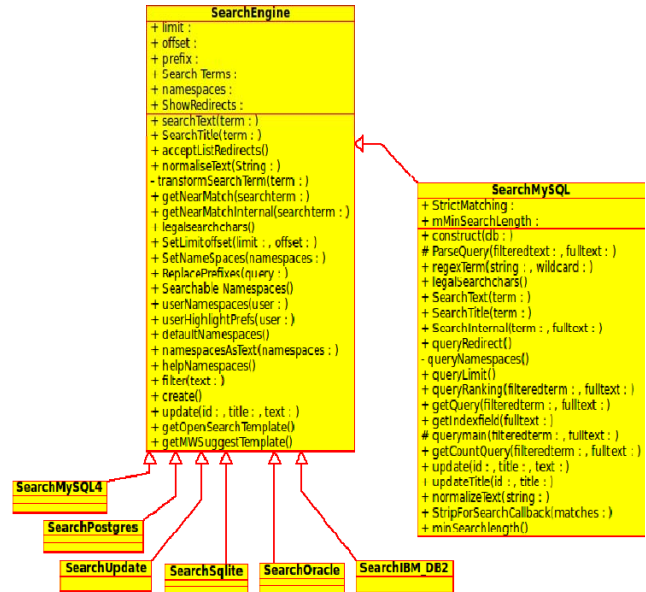


Fig. 6 Class diagram of the search implementation

### Flowchart:

The control flow of the search module in Mediawiki is depicted in the following figure. It involves tasks such as preprocessing and normalizing the search text, replacing get arguments with corresponding prefixes, resolving namespaces and so on.

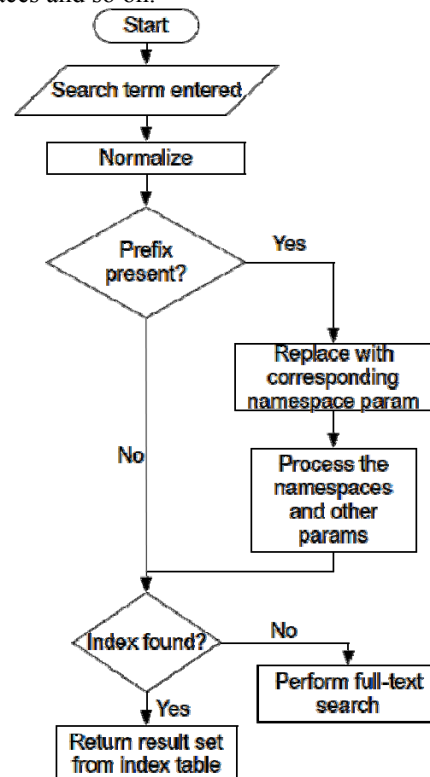


Fig. 7 Flowchart of the search process

A snippet of the search code implemented is as follows:

```
function replacePrefixes( $query ){
    global $wgContLang;
    $parsed = $query;
    if( strpos($query,':') === false )
    { // nothing to do
        wfRunHooks(
        'SearchEngineReplacePrefixesComplete', array(
        $this, $query, &$parsed ) );
        return $parsed;
    }
    $allkeyword = wfMsgForContent('searchall').":";
    if( strncmp($query, $allkeyword,
    strlen($allkeyword)) == 0 ){
        $this->namespaces = null;
        $parsed = substr($query,strlen($allkeyword));
    } else if( strpos($query,':') !== false ) {
        $prefix =
        substr($query,0, strpos($query,':'));
        $index = $wgContLang->getNsIndex($prefix);
        if($index !== false){
            $this->namespaces = array($index);
            $parsed = substr($query,strlen($prefix)+1);
        }
        else {
            $prefix =
            substr($query,0, strpos($query,':')-1)
            $parsed = '{_:'.$prefix;
        } }
        if(trim($parsed) == '')
            $parsed = $query; // prefix was the whole
            query
            wfRunHooks(
            'SearchEngineReplacePrefixesComplete', array(
            $this,
            $query, &$parsed ) );
            return $parsed;
        }
    public static function userNamespaces( $user ) {
        global $wgSearchEverythingOnlyLoggedIn;
        // get search everything preference, that can
        // be set to be read for logged-in users
        $searcheverything = false;
        if( ( $wgSearchEverythingOnlyLoggedIn && $user-
        >isLoggedIn() ) ||
        !$wgSearchEverythingOnlyLoggedIn )
            $searcheverything = $user ->
            getOption('searcheverything');
        // searcheverything overrides other options
        if( $searcheverything )
```

```
        return
        array_keys( SearchEngine::searchableNamespaces() );
        $sarr = Preferences::loadOldSearchNs( $user );
        $searchableNamespaces =
        SearchEngine::searchableNamespaces();
        $sarr = array_intersect( $sarr,
        array_keys($searchableNamespaces) ); // Filter
        return $sarr;
    }
}
```

## APPLICATIONS AND SCOPE

There are two major classes of applications that consume, and by extension, microdata:

- Web browsers
- Search engines

Browsers can provide enhanced features by detecting the annotated elements. For instance it can provide to add an event marked up as Event data-vocabulary directly to the user's Google calendar or export it to ICS format.

The other major consumer of is search engines. Instead of simply displaying the page title and an excerpt of text, the search engine could integrate some of that structured information and display it. Full name, job title, employer, address, may be even a little thumbnail of a profile photo. It would definitely catch the attention of everyone.

Google supports microdata as part of their Rich Snippets program [10]. When Google's web crawler parses your page and finds microdata properties that conform to the <http://data-vocabulary.org/Person> vocabulary, it parses out those properties and stores them alongside the rest of the page data. Google even provides a handy tool to see how Google – sees your microdataproperties.

### Google search preview

#### Richard Stallman - CSMIT

president - Free Software Foundation

The excerpt from the page will show up here. The reason we can't show text from your webpage is because the text depends on the query the user types.

[csmit.org/wiki/index.php?title=Richard\\_Stallman](http://csmit.org/wiki/index.php?title=Richard_Stallman) - Cached - Similar

Note that there is no guarantee that a Rich Snippet will be shown for this search results. For more details, see the [FAQ](#).

### Extracted rich snippet data from the page

Item

**Type:** <http://data-vocabulary.org/person>  
name = Richard Matthew Stallman  
nickname = rms  
role = software freedom activist  
role = computer programmer  
role = lead architect  
affiliation = Free Software Foundation  
title = president  
role = author

Fig. 8 Screen-shot of Output from Google Rich Snippets tool

And how does Google use all of this information? That depends. There are no hard and fast rules about how microdata properties should be displayed, which ones should be displayed, or whether they should be displayed at all. If someone searches for —Mark Pilgrim, and

Google determines that this – about page should rank in the results, and Google decides that the microdata properties it originally found on that page are worth displaying, then the search result listing might look something like the one shown in the screen-shot below.

The output shown above can be tested at <http://www.google.com/webmasters/tools/richsnippets> by entering the URL [http://csmit.org/wiki/index.php?title=Richard\\_Stallman](http://csmit.org/wiki/index.php?title=Richard_Stallman) in the input field.

### CONCLUSION AND FUTURE WORK

The project enhances Mediawiki to recognize the new Semantic Wiki markup developed and to produce microdata annotations accordingly. Thus the huge collection of Wiki pages can be made to serve as a pool of various information, for not only human beings, but also machines.

This can be further extended by making the entire output to be in HTML5, making use of the semantic elements. The search module of Mediawiki is to be enhanced to take advantage of the semantic annotations to provide accurate results with more helpful information than just excerpt of text.

### REFERENCES

- [1] Vignesh Nandha Kumar K R, Pandurangan N, Vijayakumar R and Pabitha P, *Semantic Annotation of Wiki using Wiki markup for HTML5 Microdata*, International Journal of Engineering Science and Technology, Vol. 2, Issue 12, pp. 7866-7873, 2010.
- [2] Mohammed Kayed and Chia-Hui Chang, Member, IEEE, —FiVaTech: *Page-Level Web Data Extraction from Template Pages*, IEEE Transactions on Knowledge and Data Engineering, Vol. 22, No.2, pp. 249-263, 2009.
- [3] Amal Zouaq and Roger Nkambou, Member, IEEE, *Evaluating the Generation of Domain Ontologies in the Knowledge Puzzle Project*, IEEE Transactions on Knowledge and Data Engineering, Vol. 21, No.11, pp. 15591572, 2008.
- [4] Jinhyun Ahn, Jason J. Jung, Key-Sun Choi, *Interleaving Ontology Mapping for Online Semantic Annotation on Semantic Wiki*, IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 2008.
- [5] Kensaku Kawamoto, Yasuhiko Kitamura, and Yuri Tijerino Kwasei, Gakuin University, *KawaWiki: A Semantic Wiki Based on RDF Templates*, Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT 2006 Workshops)(WI-IATW'06 , 2006.
- [6] Adam Souzis, *Building a Semantic Wiki*, IEEE Intelligent Systems, Vol. 20, No. 5 September/October 2005.
- [7] *Spinning the Semantic Web*, Edited by Dieter Fensel, James A. Hendler, Henry Lieberman and Wolfgang Wahlster, Foreword by Tim Berners-Lee.
- [8] Sebastian Schaffert, Salzburg Research Forschungsgesellschaft , François Bry, Ludwig-Maximilian University of Munich , Joachim Baumeister, University of Würzburg , Malte Kiesel, DFKI GmbH , *Semantic Wikis*, IEEE Software, 2008.

- [9] Tim Berners-Lee, James Hendler and Ora Lassila, *The Semantic Web*, Scientific American, May 2001.
- [10] Mark Pilgrim, Developer advocate at Google, Inc. Apex, NC, <http://diveintohtml5.org/extensibility.html>, 2010.
- [11] Web Hypertext Applications Technology Working Group, <http://whatwg.org/specs/web-apps/current-work/multipage>, September 2010.
- [12] Mediawiki manual [http://www.mediawiki.org/wiki/Manual:MediaWiki\\_architecture](http://www.mediawiki.org/wiki/Manual:MediaWiki_architecture), June 2010.
- [13] [http://www.mediawiki.org/wiki/Manual:Database\\_layout](http://www.mediawiki.org/wiki/Manual:Database_layout)



# Formal Verification of Finger Print ATM Transaction through Real Time Constraint Notation (RTCN)

Vivek K. Singh<sup>1</sup>, Tripathi S. P.<sup>2</sup>, R. P. Agarwal<sup>3</sup> and Singh J. B.<sup>4</sup>

<sup>1</sup> School of Computer Engineering & Information Technology, Shobhit University, Research Scholar  
Meerut, U.P. 250110, India

<sup>2</sup> Department of Computer Science & Engineering, Gautam Buddha Technical University, IET  
Lucknow, U.P. 226021, India

<sup>3</sup> School of Computer Engineering & Information Technology, Shobhit University, Professor  
Meerut, U.P. 250110, India

<sup>4</sup> School of Computer Engineering & Information Technology, Shobhit University, Professor  
Meerut, U.P. 250110, India

## Abstract

In this paper we propose the Formal Verification of existing models like in banking sector ie ATM Transaction through biometric (Finger Print) with the help of Real Time Constraint Notation. Finger print recognition is most popular and commonest method of using the biometrics. In the finger print technology, the uniqueness of epidermis of fingers is utilized for identification of user. The user has to keep its finger on a sensory pad, which reads the ridges of epidermis of finger and try to match it with available data of the finger with the bank.

Sequence Diagrams (SDs), Finite State Machine (FSM) have proven useful for describing transaction-oriented systems, and can form a basis for creating state charts. However, Finger Print ATM system require special support for branching, state information, and composing SDs.

**Keywords:** *SD\_Sequence Diagram, FPI\_Finger Print Impression, DB\_Data Base, OCL\_Object Constraint Language, FSM\_Finite State Machine*

## 1. Introduction

Traditionally, access to secure areas or sensitive information has been controlled by possession of a particular artifact (such as a card or key) and/or knowledge of a specific piece of information such as a Personal Identification Number (PIN) or a password. Today, many people have PINs and passwords for a multitude of devices, from the car radio and mobile phone, to the computer, web-based

services and their bank information. Herein lies a major difficulty involving the trade-off between usability, memorability and security[17]. Methods for increasing security, such as regularly changing PINs and passwords, increasing their length, ensuring they do not form words and ensuring all are different, makes them more difficult to remember and, therefore, error-prone. Alternatives to the traditional Personal Identification Number (PIN) have also been investigated for instance using pictures (finger print) instead of numbers [18]. Of course, traditional methods rely upon the assumption that the artifact (such as key or card) will be in the possession of the rightful owner and that the information to activate it will be kept secret. Unfortunately, neither of these assumptions can be wholly relied upon. If people are permitted to choose their own passwords they tend to select ones which are easily guessed. People tend to choose ones that are related to their everyday life [17]. They choose passwords which are easy to remember, and, typically, easily predicted, or they

change all PINs to be the same. Also, people are often lax about the security of this information and may deliberately share the information, say with a spouse or family member, or write the PIN down and even keep it with the card itself. Biometric techniques [19] may ease many of these problems: they can confirm that a person is actually present (rather than their token or passwords) without requiring the user to remember anything. In this paper, we explore how to use UML sequence diagrams to support the needs of finger print ATM verification system. First, we review methods for composing sequence diagrams that support flexible finger print ATM modeling. Then, we show how determining required information content can be represented as finite state machine to guarantee correct, cohesive diagrams. A generic approach is described; with supporting finger print ATM verification system incorporating data, state, and timing information. Finally, the more commonly discussed transaction processing model is revisited to illustrate system differences.

## 2. Biometric Approach – ATM transaction through Finger Print recognition

A *biometric system* is essentially a pattern recognition system that recognizes a person by determining the authenticity of a specific physiological and / or behavioral characteristic possessed by that person. An important issue in designing a practical biometric system is to determine how an individual is recognized. Depending on the application context, a biometric system may be called either a *verification* system or an *identification* system [16]:

1. A verification system authenticates a person's identity by comparing the captured biometric characteristic with her own biometric template(s) pre-stored in the system. It conducts one-to-one comparison to determine whether the identity claimed by the individual is true. A verification system either rejects or accepts the submitted claim of identity.
2. An identification system recognizes an individual by searching the entire template database for a match. It conducts one-to-many comparisons to establish the identity of the individual. In an identification system, the

system establishes a subject's identity (or fails if the subject is not enrolled in the system database) without the subject having to claim an identity.

The term *authentication* is also frequently used in the biometric field, sometimes as a synonym for verification; actually, in the information technology language, authenticating a user means to let the system know the user identity regardless of the mode (verification or identification).

The banking and financial sector has adopted this system wholeheartedly because of its robustness and the advantages it provides in cutting costs and making processes more streamlined. The technology started out as a novelty however due exigencies in the banking sector characterized by decreasing profits it became a necessity. The use of Biometric ATM's based on finger print recognition technology has gone a long way in improving customer service by providing a safe and paperless banking environment.

Identification of right user by the use of face recognition technology is the latest form of biometric ATMs. Identification based on the different walk style while entering in ATMs is used in gait based ATMs.

Benefits of biometric technology: Since biometric technology can be used in the place of PIN codes in ATMs, its benefits mostly accrues to rural and illiterate masses who find it difficult to use the keypad of ATMs. Such people can easily put their thumbs on the pad available at ATMs machines and proceed for their transactions.

Biometric technology provides strong authentication, as it uses the unique features of body parts. This helps reduce the chances of occurring frauds in ATM usage.

Though use of biometric technology has its high cost implications to banks, several other costs of conventional ATMs like re-issuance of password, helpdesk etc will be reduced, which will be a positive factor for banks to go for biometric ATMs.

## 3. Terminology Used

### 3.1 Scenario, Sequence Diagrams, State Chart & Message Sequence Charts:

A scenario is a sequence of events that occurs during one particular execution of a system. A scenario describes a way to use a system to accomplish some function [5]. Scenarios can be expressed in many forms, textual and graphical, informal and formal. Sequence diagrams emphasize temporal ordering of events, whereas collaboration diagrams focus on the structure of interactions between objects. Each may be readily translated into the other. State chart diagrams represent the behavior of entities capable of dynamic behavior by specifying its response to the receipt of event in stances.

Typically, it is used for describing the behavior of classes, but state charts may also describe the behavior of other model entities such as use cases, actors, subsystems, operations, or methods. Message sequence charts constitute an attractive visual formalism that is widely used to capture system requirements during the early design stages in domains such as ATM transaction via fingerprint recognition [15].

### 3.2 Composition of Scenarios:

A crucial challenge in describing formal verification of fingerprint ATM recognition is the composition of scenarios. In order to be adequately expressive, sequence diagrams must reflect the structures of the programs they represent. In this paper, we survey approaches to modeling execution structures and transfer of control, and select a method that lends itself to Fingerprint Verification System.

Our objective is to refine a model that utilizes sequential, conditional, iterative, and concurrent execution. As many ideas exist, our task is to determine which are appropriate for Fingerprint Verification System. Hsia et al. [5] discusses a process for scenario analysis that includes conditional branching. Glinz [2] includes iteration as well. Koskimies et al.[8] and Systa [13] present a tool that handles “algorithmic scenario diagrams” - sequence diagrams with sequential, iterative, conditional and concurrent behavior. We use elements of each, for a combined model that allows sequential, conditional, iterative, and concurrent behavior.

Another objective is to model transfer of control through sequence diagram composition. The main decision to make is where to annotate control information. One approach is to include composition information in individual diagrams.

### 3.3 Finite State Machines (FSM):

By Dr. Matt Stallmann & Suzanne Balik: “A *finite-state machine (FSM)* is an abstract model of a system (physical, biological, mechanical, electronic, or software)”.

A *finite state machine (FSM)* is a mathematical model of a system that attempts to reduce the model complexity by making simplifying assumptions. Specifically, it assumes

1. The system being modeled can assume only a finite number of conditions, called *states*.
2. The system behavior within a given state is essentially identical.
3. The system resides in states for significant periods of time.
4. The system may change these conditions only in a finite number of well defined ways, called *transitions*.
5. Transitions are the response of the system to *events*.
6. Transitions take (approximately) zero time.

### 3.4 Object Constraint Language (OCL):

The Object Constraint Language (OCL) is an expression language that enables one to describe constraints on object – oriented models and other object modeling artifacts.

A constraint is a restriction on one or more values of (part of) of an object oriented model or system. OCL is the part of the Unified Modeling Language (UML), the OMG (Object Management Group, a consortium with a membership of more than 700 companies. The organization's goal is to provide a common framework for developing applications using object-oriented programming techniques) standard for object – oriented analysis and design.

OCL has been used in a wide variety of domains, and this has led to the identification of some under – specified areas in the relationship between OCL and UML.

OCL can be used for a number of different purposes:

1. to specify invariants on classes and types in the class model
2. to specify type invariants for Stereotypes
3. to describe pre- and post- conditions on Operations and Methods
4. to describe guard
5. as a navigation language
6. to specify constraint on operations

In OCL, UML operation semantics can be expressed using pre and post condition constraints – The pre condition says what must be true for the operation to meaningfully execute – The post condition expresses what is guaranteed to be true after execution completes

1. About the return value
2. About any state changes (e.g. instance variables)

## 4. Proposed Formal Verification of Finger Print ATM Transaction through Real Time Constraint Notation (RTCN) :-

Now we are going to demonstrate the formal verification of ATM transaction through Fingerprint Verification Model with the help of Sequence Diagram their corresponding Finite State Machine and their corresponding Real Time Constraint Notation with the help of Object Constraint Language (OCL). There are four objects exchanging messages: the user, the ATM, the consortium, and the bank. In this example, State charts are generated for the ATM object only. The scenarios share the same initial condition.

### 1. Through Sequence Diagrams (SD's):

Case 1: Transaction Fail due to mismatch Finger Print Impression (FPI) at server site database:

In this case, ATM transaction fails due to mismatch of finger print through finger print database (DB) file server site:

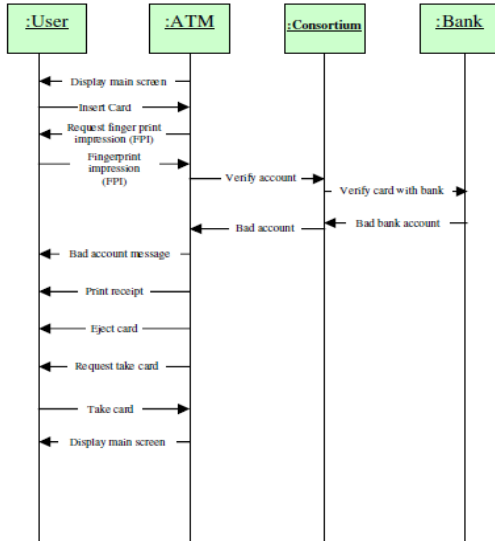


Figure 1 Sequence Diagram for Finger Print Verification ATM, Case1: Mismatch Fingerprint(FPI)

Case 2: Transaction success due to match of Finger Print Impression (FPI) at server database:  
 In this case, ATM transaction is successfully processed due to accurate match of finger print through finger print database (DB) file server site:

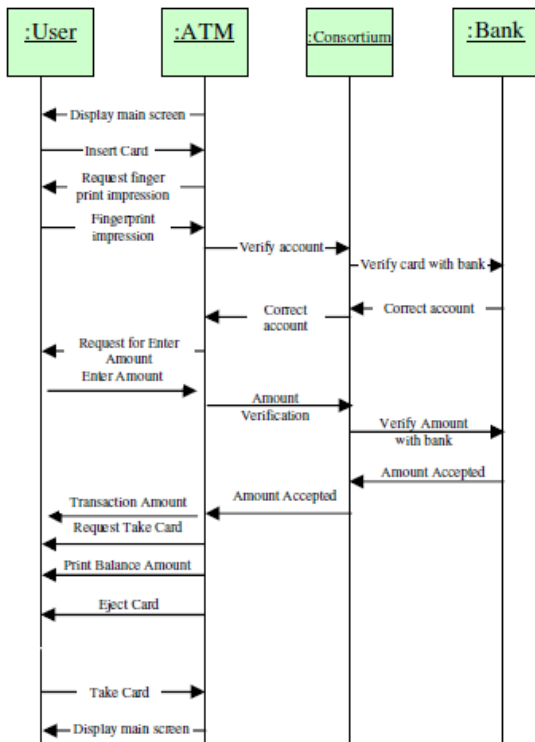


Figure 2 Sequence Diagram for Fingerprint Verification ATM, Case2: Correct FPI, Successful Transaction  
 2. Finite State Machine corresponding to Sequence Diagrams (SD's):  
 The above two cases can be represented with help of their corresponding Finite State machine as :

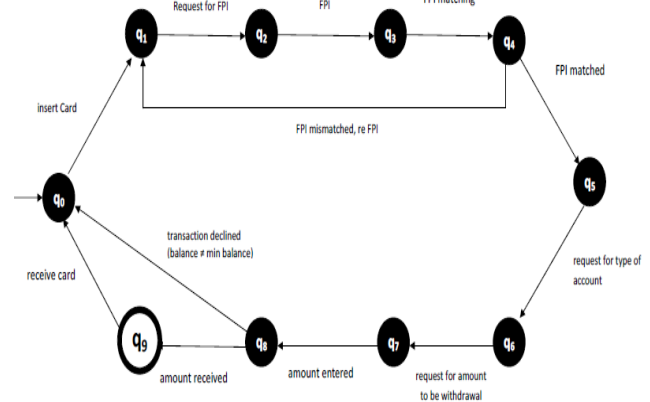


Figure 3 : Finite State Machine for ATM Transaction by FPI

Fig 3 representing the Finite State Machine (FSM) for the transaction through ATM after the verification of Finger Print Impression(FPI). FSM has ten states represented as q0, q1, q2, q3, q4, q5, q6, q7, q8, q9, where q0 is an initial state and q9 is the final state of the FSM. We can define the transition function ( $\delta$ ) for the proper working FSM as shown in Fig 3 as follows:

1.  $\delta(q_0, \text{insert card}) = q_1$
2.  $\delta(q_1, \text{request for FPI}) = q_2$
3.  $\delta(q_2, \text{FPI}) = q_3$
4.  $\delta(q_3, \text{FPI matching}) = q_4$
- 5.1  $\delta(q_4, \text{FPI matched}) = q_5$
- 5.2  $\delta(q_4, \text{FPI mismatched, re FPI}) = q_1$
6.  $\delta(q_5, \text{request for type of account}) = q_6$
7.  $\delta(q_6, \text{request for amount to be withdrawal}) = q_7$
8.  $\delta(q_7, \text{amount entered}) = q_8$
9. 9.1  $\delta(q_8, \text{transaction declined}) = q_0$
- 9.2  $\delta(q_8, \text{amount received}) = q_9$  - Final State
10.  $\delta(q_9, \text{receive card}) = q_0$  - Initial State

3. Real Time Constraint Notation corresponding to Sequence Diagrams (SD's) (Case 1 & 2) with the help of Object Constraint Language(OCL):

We are proposing the Object Constraint Language (OCL) notation for user's transaction through ATM after finger print verification as :

**User :: Finger Print Impression(FPI)**  
**pre : Finger Print Identification**

**post :**  
**Request Transaction.Saving Account**  
**->reject ( not ( FPI = User ))**  
**or**  
**Request Transaction.Saving Account**  
**->reject((FPI = User))**  
**and**  
**(Request Transaction and Saving Account @ pre**  
**+ Account Balance >= Minimum Balance)**  
**and**  
**(Request Transaction and Saving Account @ post**  
**+ Account Balance >= Account Balance -**  
**Withdrawal Amount)**

## 5. Conclusion:

We presented a methodology that guarantees sufficient sequence diagram information to generate correct Statecharts. We converted sequence diagrams to the respective Finite State Machine (FSM) and also give the Object Constraint Language (OCL), pre / post conditions for transaction process through ATM. When state, message preconditions, and timing information are included in the FSM & OCL notations seems to be sufficient to guarantee determinism for the Fingerprint ATM Verification we discussed. We have also examined diagram composition and information content to assess adequacy for Fingerprint ATM Verification.

## References

- [1] Douglass, B. Doing Hard Time. Addison-Wesley, 99.
- [2] Glinz, M. An Integrated Formal Model of Scenarios Based on Statecharts. In *Proceedings of the 5th European Software Engineering Conference (ESEC 95)*, Sitges, Spain, 1995, pp.254-271.
- [3] Harel, D. Statecharts: A Visual Formalism for Complex Systems. *Science of Computer Programming*, vol.8, no.3, 1987, pp. 231-274.
- [4] Hitz, M., and G. Kappel. Developing with UML - Some Pitfalls and Workarounds. *UML '98 - The Unified Modeling Language*, Lecture Notes in Computer Science 1618, Springer-Verlag, 1999, pp. 9-20.
- [5] Hsia, P. et al. Formal Approach to Scenario Analysis. *IEEE Software*, vol.11, no.2, 1994, pp.33-41.
- [6]ITU-T. Recommendation Z.120. ITU Telecommunication Standardization Sector, Geneva, Switzerland, May 1996.
- [7] Khriiss, I., M. Elkoutbi, and R. Keller. Automating the Synthesis of UML StateChart Diagrams from Multiple Collaboration Diagrams. *UML '98 - The Unified Modeling Language*, Lecture Notes in Computer Science 1618, Springer-Verlag, 1999, pp. 132-147.

[8] Koskimies, K., T. Systa, J. Tuomi, and T. Mannisto. Automated Support for Modeling Software. *IEEE Software*, vol.15, no.1, 1998, pp. 87-94.

[9] Leue, S., L. Mehrmann, and M. Rezaei. Synthesizing Software Architecture Descriptions from Message Sequence Chart Specifications. In *Proceedings of the 13th IEEE International Conference on Automated Software Engineering*, Honolulu, Hawaii, 1998, pp. 192-195.

[10] Li, X. and J. Lilius. Checking Compositions of UML Sequence Diagrams for Timing Inconsistency. In *Proceedings of the Seventh Asia-Pacific Software Engineering Conference (APSEC 2000)*, Singapore, 2000, pp. 154-161.

[11] Louden, K. Compiler Construction : Principles and Practice. PWS Publishing Company, 1997.

[12] Some, S., R. Dssouli, and J. Vaucher. From Scenarios to Timed Automata: Building Specifications from User Requirements. In *Proceedings of the 1995 Asia-Pacific Software Engineering Conference*, Australia, 1995, pp. 48-57.

[13] Systa, T. Incremental Construction of Dynamic Models for Object-Oriented Software Systems. *Journal of Object-Oriented Programming*, vol.13, no.5, 2000, pp. 18-27.

[14] Unified Modeling Language Specification, Version 1.3, 1999. Available from the Object Management Group. <http://www.omg.com>.

[15] Elizabeth Latronico and Philip Koopman: Representing Embedded System Sequence Diagrams As A Formal Language

[16] D. Maltoni, D. Maio, Handbook of Fingerprint Recognition, Springer, 2003

[17] Adams, A. and Chang, S.Y. An investigation of keypad interface security. *Information & Management*, 24, 53-59, 1993.

[18] De Angeli, A., Coutts, M. Coventry, L., Johnson, G.I, Cameron D., and Fischer M. VIP: a visual approach to user authentication. *Proceedings of the Working Conference on Advanced Visual Interfaces AVI 2002*, ACM Press, pp. 316-323, 2002

[19] Ashbourn, J. *Biometrics. Advanced Identity Verification*. Springer Verlag, London, 2000.



### **Acknowledgments**

Singh, Vivek thanks Darbari, Manuj without his essential guidance this research paper would not have been possible .

**Singh, Vivek** is currently working as Assistant Professor in the Department of I. T. at BBDNITM, Lucknow. He has over 10 years of teaching & experience. Having done his B.Tech in Computer Science & Engineering from Purvanchal University in 2001, M.Tech from U.P.Technical University, Lucknow in 2006, he is pursuing his Ph.D. from Shobhit University, Meerut.

**Tripathi, S.P. (Dr.)** is currently working as Associate Professor in Department of Computer Science & Engineering at I.E.T. Lucknow. He has over 30 years of experience. He has published numbers of papers in referred Journals.

**Agarwal, R. P. (Dr.)** is currently working as Professor & Director in School of CE&IT at Shobhit University, Meerut. He has 40 years of teaching experience and has published number of papers in referred Journals.

**Singh, J.B. (Dr.)** is currently working as Dean Students Welfare at Shobhit University, Meerut. He has 38 years of teaching experience and has published number of papers in referred Journals.

# Self-Destructible Concentrated P2P Botnet

Mukesh Kumar, Pothula Sujatha, P. Manikandan, Madarapu Naresh Kumar, Chetana Sidige and Sunil Kumar Verma\*

School of Engineering and Technology, Department of Computer Science  
Pondicherry University-605014  
Puducherry, INDIA  
Indian Institute of Information Technology Allahabad, INDIA\*

## Abstract

Small botnets are tough to detect and easy to control by the botmaster. Having a small botnet with high speed internet connectivity than large but slow connection is more effective and dangerous in nature. According to diurnal dynamics studies only about 20 percent of computers are always online, to maximize a botnet attack power, botmaster should know diurnal dynamics of her botnet. In our project we are designing a peer-to-peer bot. This bot after infecting any of the system first check the internet connection speed of the interface, if it is not up to the desired speed i.e. 2 Mbps the bot will kill itself because slow speed bots are not desired. In another scenario bot will sense is it in a honeypot trap? If so it will kill itself so that the whole botnet could not be exposed to the defender. We will suggest the mitigation techniques to defend bots with these types of properties.

**Keywords-** Peer-to-Peer, Botnet, Honey pot, Firepower

## 1. Introduction

As technology for internet security matured Internet malware, and Ransom ware domination also increased. Users and organizations are suffering a lot by these attack emerging trend[1]. These hackers became equipped with more advanced technologies and planning their attack in better-organized manner which is more dangerous than earlier years. The botnet crime results E-mail spam, extortion through denial-of-service attacks, identity theft, data theft and click fraud resource consumption etc. A “botnet” is a network of systems affected by malwares known as “bots”. These bots has one specific property that distinguish them with other malwares, they can be remotely operated and controlled. This specific property of bots makes them weapons for various denials of service attacks. These bots are distributed over the internet having

enormous cumulative bandwidth if controlled by the Botmaster, to attack any target on the internet. The concept of botnets is evolved from just the last decade, due to open source communities day by day new variants of bots with new stealthy protocols and infection capability are attacking and affecting the victim.

Botnet-based attacks are becoming more powerful and dangerous in such case security professionals needs to understand the newly developed bots. For understanding and study of the bots various works has been done by the researchers across the world [4], [7],[8],[9],[10],[11]. Internet Relay Chat(IRC) based botnets are the first kind of bots using C&C (Command & Control) architecture as a centralized systems. Recent years are more prominent with new technology based bots for their Command & Control. A new type of the bot using Peer-to-Peer topology for the spreading of command and control by the botmaster is more prominent. Various works have been done to understand and create detection frameworks and systems to detected and dismantle the botnets. Various detection mechanism for IRC based botnets are proposed[12],[15],[16]. As now a days Peer-to-Peer botnets are more dangerous in nature detection framework is proposed for them [12],[13],[14]. As per our understanding new kind of bots can be generated easily for creating and developing a mitigation system for the botnets we have to understand their capability and activity. For this purpose development framework for new bots should be created.



## 2. Related Works

Bots and Botnets are very hot topics for last few years [10], [1]. The first ever Peer-to-Peer bot Storm Bot had control over a million systems. In 2003 first ever bots and botnets properties and overview is discussed by Puri and McCarty. Today the main concentration of bots researcher are on Peer-to-peer bots because of their sustainability and robust network topology formation makes tough to detect and dismantle. Various authors proposed different types of Peer-to-Peer bots. [3] developed a Stochastic Model of Peer-to-Peer botnet to understand different factors and impact of the growth of the botnet. The botnet stochastic model was constructed in the Mobius software tool, which was designed to perform discrete event simulation and compute analytical/numerical solution of models by inputting various input parameters. This kind of research helps to understand the behavior of botnets and it became easy to create mitigation systems and framework for these bots. In botnet technology various works are going on for the detection and mitigation of the Peer-to-Peer botnets.

Authors has proposed an advanced hybrid peer-to-peer botnet [19] which concentrated on the problem of are using the liability constraint of the security professional to detect installed honeypot, because honeypots are not allowed to participate in the real attack scenario. But still some probability remains for the capture of the bots and reverse engineered to understand their strength. This lack of security in bots capture by the defender make the whole botnet susceptible to get exposed.

Significant exposure of the network topology when one of the bot is captured, making easy for the botmaster for the overall control of the botnet. They also included some concept of Honey Pot awareness in their bot system. But still few problems with communication channel and the capture and re engineering of the bot is remains.[7] Predicting a new botnet from the framework and comparing its performance with known ones. Loosely Coupled peer-to-Peer botnet lcbot, which is stealthy and can be considered as a combination of existing P2P botnet structure. Their botnet architecture still follows the idea of “Buddy list” or routing information of the infected host or

friend bots. Which keep the whole botnet easy to exposed if one of the bot got captured by the defender. Peer list construction is the main concept behind any P2P botnet which also leave the complete bot exposed any time to the Defender. [5],[6],[17] Authors giving an idea of Honey pot aware bots, and botnets. Honeypots are the only way to observe and understand the activities of a bot. That also makes botnet prone to be exposed to the defender and help them to create a mitigation system for the botnet. Bot masters.

## 3. Proposed P2P Botnet Architecture

### 3.1 Classification of Our Bots

We classified our bots very extensively so that it becomes easy to control and operate the botnet by the botmaster. This classification is mainly to refine the bots used in the attack for an effective firepower and less prone for the exposure to the defender. First of all we will group our bots on the basis of their bandwidth if the infected system has a internet connectivity to the outside world equal or greater than our specified bandwidth then only we will consider them to build our botnet these kind of bots we call as Live bots, otherwise we will discard the further infection and these kind of bots will be called as Dead bots and they will not participate in further creation of the botnet. Further we will classify Live bots in two groups one Peer bots which will have global IP addresses without firewall or proxy servers in between, and rest all bots including 1) bots with global IP addresses with firewall or proxy 2) bots with dynamically allocated global IP addresses 3) bots with private IP addresses. We will call second group of bots as Non-peer bots. Further bots are dedicated for the purpose of either infecting other victims or only for attack purpose. If the bot is dedicated for infection of other victims then the code module will send the existing peer list to newly infected bot. In case of attack bots the code module will be spam emails, DDoS command and control handling.

We will mainly concentrate to prevent detection of the Peer bots because they are security bottle neck for our botnet to get exposed to the defender as they only contain peer list or seed list information of other Peer bots. The Peer bots will be able to act as a server for other Peer and Non-Peer bots and client for other Peer



bots. Non-Peer bots will be able to act only as a client, they will have entry of other Peer bots only in their peer list.

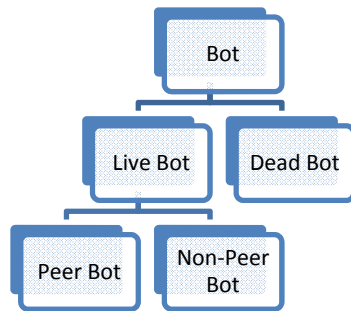


Fig:1 Classification of Proposed Bot

According to the following properties suicide module will delete the peer list information.

- 1- Bot master intentionally wants to kill the bot, sends command through the communication channel.
- 2- It will delete peer list in the peer bot to save whole botnet.
- 3- If average availability of bot in a week is less than desired bot will kill itself.
- 4- In case of any threat or danger like honey pot trap it will execute suicide module to delete peer list information.

### 3.2 Botnet Infection

The infection and propagation of the bot is a very important for the purpose of robust bot network and control . The bot will use the backdoor created by other worms for infection. It will first infect the system with first stage small infection code, after wards it will execute various commands and modules to check the bandwidth of the infected system interface, if it is upto desired speed then it will become as live bot and next step of the infection will proceed otherwise it will be declared as a dead bot which further not participate in the bot formation. After the bandwidth check it will perform the IP type checking to confirm weather that infected system can work as a Peer bot or Non-Peer bot. These bots will further propagate to infect other systems.

Stage 1: Initial Infection (Compromising the system)

- I. Install the initial Infection files
- II. Check the connection speed of victim
- III. Decide whether the compromised host is Live or Dead Bot

Stage 2: Participating in Peer network (Creation of Botnet)

- I. Connect to the Peers
- II. Update Peers list
- III. Search the network for encrypted URL

Stage 3: Secondary Injection ( Code for attack purpose)

- I. Connect to the encrypted URL
- II. Download the secondary injection code
- III. Execute the code to enhance Bots Power

### 3.3 Peer Network Creation

Newly infected bots will communicate with the existing bots to update their peer list and other information's. The bot cannot participate in the attack until it connects to the other peers in to the existing botnet, and become ready to share the command and control given by the Botmaster. For every newly connected bot in the botnet we'll use hashing of IP addresses, and a key for the identification by the botmaster. A Bot's identifier is chosen by hashing the bot's IP address, while a key identifier is produced by hashing the key. The identifier length must be large enough to make the probability of two Bots or keys hashing to the same identifier negligible. Identifiers are ordered in an *identifier circle*. Key is assigned to the first node whose identifier is equal to or follows (the identifier of) in the identifier space. This node is called the *successor node* of key , denoted by  $successor(k)$ . Consistent hashing is designed to let bots enter and leave the network with minimal disruption. To maintain the consistent hashing mapping when a bot  $n$  joins the network, certain keys previously assigned to  $n$ 's successor now become assigned to  $n$ . When bot  $n$  leaves the network, all of its assigned keys are reassigned to  $n$ 's successor. No other changes in assignment of keys to nodes need occur.

In a dynamic bot network, bots can join (and leave) at any time. The main challenge in implementing these operations is preserving the ability to locate every key in the bot network. In order for lookups to be fast, it is also desirable for the finger tables to be correct. We'll maintain a table in each bot for the storing the peer information and the identifier key generated by the hash function using the bots IP address. This finger

table is important to maintain the robust connectivity of the bot to the network. Our emphasis will be here to maintain the correct finger table as accurate as possible.

### 3.4 Botnet Communication Channel Architecture

Each Peer Bot will contain list of its next two peer bots and other two non-peer bot information in seed list. The non-peer bot will have only two entries of peer bot information with the condition that they both peer bot will contain the information of each other. The bot master will pass the command to any one of the Peer bot depending upon the diurnal dynamics that particular bot will be selected for the first command passing to the whole botnetwork. After getting the command by the botmaster the peer bot will share this command to its next neighbor peer bot as will connected non-peer bot that will ensure the effective communication, for the purpose of command passing priority will be given to the peer bot. Because peer bot can work as client as well as server too, and connected to other peer bots . On the basis of this topology the communication will be handled.

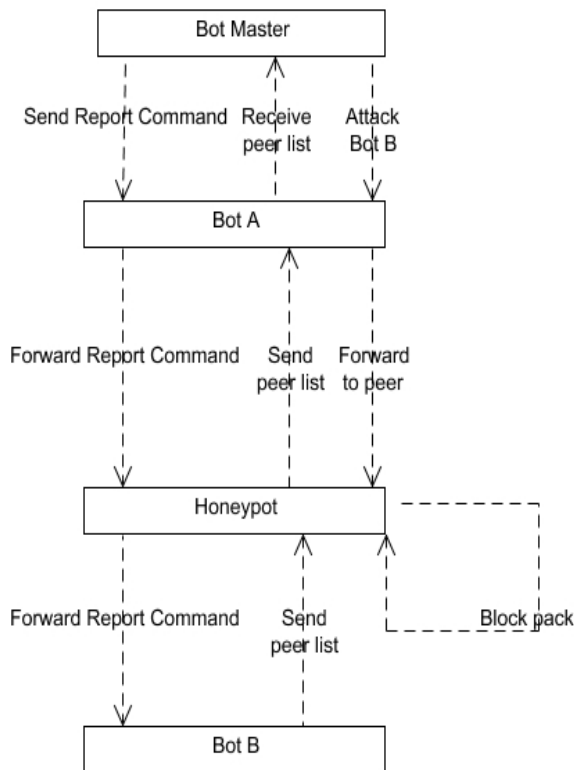


Fig 2: Bot Communication

## 4. Simulation and Experimental Results

We are presenting simulation results and snapshots of our proposed peer to peer suicide botnet. Our experiment is still in its inception stage, in its current scenario we became successful to implement and execute few properties of our proposed model of botnet. In the simulation model implemented using java technology the botmaster is able to command bots through listing their IP addresses. If necessity arises botmaster sends kill command to the bots to destruct itself.

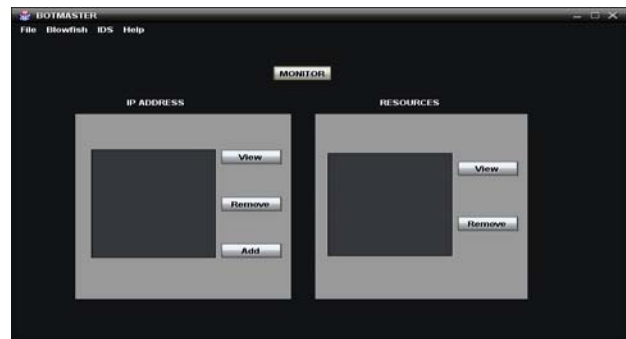


Fig:3 Bot Master Control Interface

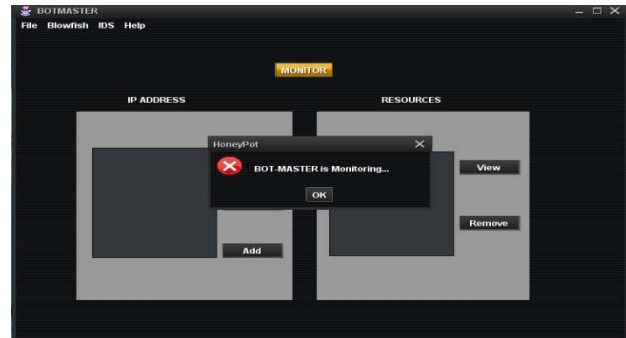


Fig:4 Bot Monitoring by Bot Master

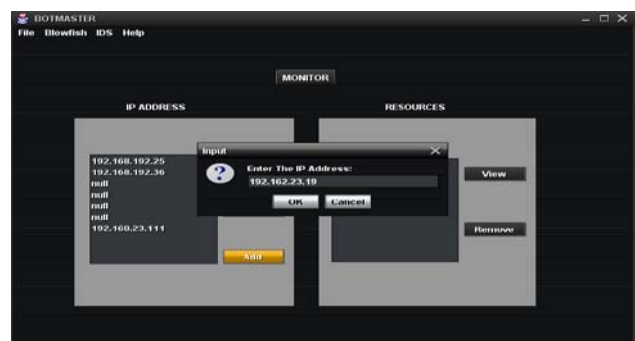


Fig:5 Selecting Bot IP Address by Bot Master to send Kill Command





Fig:6 IP address listing of Bot by Bot Master

## 5. Conclusion

Implementation of new types of bots will facilitates to understand the future bots which can be created by the attackers. Study and simulation results of our bot provide framework to understand the bot working and there communication channel architecture. This bot is tough to control because of the peer network topology but harder to reverse engineered or trapped by the honey pots. It provide small but high fire power bot network to the bot master which is tough to shut down.

## References

- [1] B. McCarty, "Botnets: Big and Bigger," IEEE Security & Privacy Magazine, vol. 1, no. 4, pp. 87-90, July-Aug. 2003.
- [2] Elizabeth Van Ruitenbeek and William H. Sanders, "Modeling Peer-to-Peer Botnets", Quantitative Evaluation of Systems 2008 IEEEComputerSociety, DOI 10.1109/QEST.2008.43, 5<sup>th</sup> International Conference on Quantitative Evaluation of SysTem Palais du Grand Large at Saint Malo, France 14th-17th September, 2008
- [3] Julian B. Grizzard, Vikram Sharma, Chris Nunnery, and Brent ByungHoon Kang, "Peer-to-Peer Botnets: Overview and Case Study", In USENIX Workshop on Hot Topics in Understanding Botnets (HotBots'07) April 10 2007, Cambridge, MA, USA
- [4] Justin Leonard, Shouhuai Xu and Ravi Sandhu, "A Framework for Understanding Botnets", 2009 International Conference on Availability, Reliability and Security Fukuoka Institute of Technology, Fukuoka, Japan March 16-March 19.
- [5] Ping Wang, Lei Wu, Ryan Cunningham, Cliff C. Zou, "Honey-pot Detection in Advanced Botnet Attacks", Int. J. Information and Computer Security, Vol. 4, Issue 1, 2010, DOI: 10.1504/IJICS.2010.031858
- [6] Simon Innes, Craig Valli, "Honeypots: How do you know when you are inside one?", the 4th Australian

- Digital Forensics Conference, Edith Cowan University, Perth Western Australia, December 4<sup>th</sup> 2006.
- [7] A Framework for P2P Botnet, Su Chang, Linfeng Zhang, Yong Guan, Thomas E. Daniel, 2009 International Conference on Communications and Mobile Computing
- [8] The Zombie Roundup: Understanding, Detecting and Disrupting Botnets, Evan Cooke, Rarnam Jahanian, Danny McPherson Electrical Engineering and Computer Science Department Arbor Networks.
- [9] wide-scale Botnet Detection and Characterization Anestis Karasaridis, Brain Rexroud, David Hoefflin
- [10] A Survey of Bots used for Distributed Denial of Service Attack, Vrizzlynn L. L. Thing, Morris Solman, Naranker Dulay, <http://www.doc.ic.ac.uk>.
- [11] Criminology of Botnets and their detection and Defense Methods, Jivesh Govil, Jivika Govil, IEEE EIT 2007 Proceedings, IEEE 2007
- [12] Automatic Discovery of Botnet Communities of Large Scale Communication Network, Wei Lu, Mahbood Tavallae and Ali A Ghorbani, ASIACCS'09, March 10-12, 2009, Sydney, NSW, Australia ACM 2009.
- [13] P2P botnet detection using behavior clustering & Statistical Tests, Su Chang, Thomas E. Daniels, AISec'09, November 9, 2009 ACM 978-1-60558-781-3/09/11
- [14] A Proposed Framework for P2P Botnet Detection, Hossein Rouhani Zeidanloo, Azizah Bt Abdul Manaf, Rabiah Bt Ahmad, Mazdak Zamani, Saman Shojae Chaeikar, IACSIT International Journal of Engineering And Technology, Vol, No. 2, April 2010,
- [15] Detecting Botnets with Tight Command and Control, W. Timothy Strayer, Robert Walsh, Carl livadsa, David Lapsley,
- [16] A Novel Approach to Detect IRC-based Botnets, Wei WANG, Binxing FANG, Zhaoxin ZHANG, Chao LI, " 2009 International Conference On Network Security, Wireless Communication and Trusted Computing, 2009 IEEE.
- [17] Honey-pot-Aware Advanced Botnet Construction and Maintenance Cliff C. Zou Ryan Cunningham , Proceedings of the 2006 International Conference on Dependable Systems and Networks (DSN'06) IEEE.
- [18] VMM-Based Framework for P2P Botnets Tracking and Detection LingYun Zhou , 2009 International Conference on Information Technology and Computer Science
- [19] Ping Wang, Sherri Sparks, and Cliff C. Zou Member IEEE, "An Advanced Hybrid Peer-to-Peer Botnet", IEEE Transactions On Dependable and Secure Computing, Vol. 7, No. 2, April-June 2010 page





**Mukesh Kumar** received his Bachelor of Technology degree in Computer Science and Engineering from Uttar Pradesh Technical University Lucknow, India, in 2009. He is currently pursuing his master's degree in Network and Internet Engineering in the School of Engineering and Technology, Department of Computer Science, Pondicherry University, India. His research interests

include Denial-of Service resilient protocol design, Cloud Computing and Peer to Peer Networks.



**Chetana Sidige** is presently pursuing M.Tech (Final year) in Computer Science of Engineering at Pondicherry University. She did her B.Tech in Computer Science and Information Technology from G. Pulla Reddy Engineering College, affiliated to Sri Krishnadevaraya University. Her research interest includes Network Security, Information retrieval Systems and Software metrics. Currently the author is working on Multilingual Information retrieval evaluation.



**Pothula Sujatha** is currently working as Assistant Professor and pursuing her PhD in Department of Computer Science from Pondicherry University, India. She completed her Master of Technology in Computer Science and Engineering from Pondicherry University and completed her Bachelor of Technology in Computer Science and Engineering from Pondicherry Engineering College,

Pondicherry. Her research interest includes Information Security, Modern Operating Systems, Multimedia Databases, Software Metrics and Information Retrieval. Her PhD research is on performance Evaluation of MLIR systems.



**Sunil Kumar Verma** received his Bachelor of Technology degree in Computer Science and Engineering from Uttar Pradesh Technical University Lucknow, India in 2009. He is currently pursuing his master's degree in Cyber Law & Information Security from Indian Institute of Information Technology Allahabad. His research interests include

Denial-of Service resilient protocol design, cryptography and network security.



**P Manikandan** is presently pursuing Master of Technology in Computer Science with specialization in Network and Internet Engineering from Pondicherry University, India. He has completed his Bachelor of Technology in Computer Science and Engineering from Bharathiyar College of Engineering and Technology affiliated to

Pondicherry University. His research interest includes Wireless Communication, Network Security, distributed systems, Red Hat. Currently he is working on Thread Scheduling in Solaris.



**Madarapu Naresh Kumar** is presently pursuing Master of Technology in Computer Science with specialization in Network and Internet Engineering from Pondicherry University, India. He has completed his Bachelor of Technology in Computer Science and Engineering from JNTU Hyderabad. His research interest includes Cloud Computing, Web

Services, Software Metrics, SOA and Information Retrieval. Currently the author is working on security issues in Cloud Computing

# Fast Overflow Detection in Moduli Set $\{2^n - 1, 2^n, 2^n + 1\}$

Mehrin Rouhifar<sup>1,\*</sup>, Mehdi Hosseinzadeh<sup>2</sup>, Saeid Bahanfar<sup>3</sup> and Mohammad Teshnehlab<sup>4</sup>

<sup>1</sup> Dept. of Computer engineering , Islamic Azad University, Tabriz branch  
Tabriz, Iran

<sup>2</sup> Islamic Azad University, Science and research branch  
Tehran, Iran

<sup>3</sup> Dept. of Computer engineering , Islamic Azad University, Tabriz branch  
Tabriz, Iran

<sup>4</sup> Dept. of Control engineering, K. N. Toosi University of Technology  
Tehran, Iran

## Abstract

The Residue Number System (RNS) is a non weighted system. It supports parallel, high speed, low power and secure arithmetic. Detecting overflow in RNS systems is very important, because if overflow is not detected properly, an incorrect result may be considered as a correct answer. The previously proposed methods or algorithms for detecting overflow need to residue comparison or complete convert of numbers from RNS to binary. We propose a new and fast overflow detection approach for moduli set  $\{2^n-1, 2^n, 2^n+1\}$ , which it is different from previous methods. Our technique implements RNS overflow detection much faster applying a few more hardware than previous methods.

**Keywords:** Residue number system, overflow detection, moduli set  $\{2^n-1, 2^n, 2^n+1\}$ , group number.

## 1. Introduction

Residue number systems (RNS) have been for a long time a topic of intensive research. Their usefulness has been demonstrated, especially for computations where additions, subtractions and multiplications dominate, because such operations can be done independently for each residue digit without carry propagation [1]. Other operations such as overflow detection, sign detection, magnitude comparison and division in RNS are very difficult and time consuming [2, 3]. However, above mentioned operations are essential in certain applications, e.g. in exact arithmetic or computational geometry, where residue arithmetic is applied [4].

The RNS is determined by the set  $m$  of  $n$  positive coprime integers  $m_i > 1$ , which forms the base of the system. The dynamic range  $M$  of that system is given as a product of the moduli  $m_i$  where

$$M = \prod_{i=1}^n m_i. \quad (1)$$

Any integer  $X \in [0, M)$  has a unique representation  $(x_1, x_2, \dots, x_n)$  in RNS  $(m_1, m_2, \dots, m_n)$ . The residues  $x_i = |X|_{m_i}$ , also called residue digits, are defined as

$$x_i = X \bmod m_i, \quad 0 \leq x_i < m_i. \quad (2)$$

To convert a residue number  $(x_1, x_2, \dots, x_n)$  into its binary representation  $X$ , the Chinese Remainder Theory (CRT) is widely used. In CRT, the binary  $X$  is computed by:

$$X = \left\langle \sum_{i=1}^n (x_i N_i)_{m_i} \times M_i \right\rangle_M \quad (3)$$

where  $M_i = M / m_i$  and  $N_i = \langle M_i^{-1} \rangle_{m_i}$  is the multiplicative inverse  $M_i$  modulo  $m_i$  [5].

RNS has numerous applications in Digital Signal Processing (DSP) for filtering, convolutions, correlations, FFT computation [6, 7], fault tolerant computer systems

\* Corresponding Author

[8], communication [9], cryptography and image processing [10, 11].

Overflow detection is one of the fundamental issues in efficient design of RNS systems. In a generic approach, overflow occurs in the addition of two numbers  $X$  and  $Y$ , whenever  $Z = (X + Y) \bmod M$  be less than  $X$ . Thus, the problem of overflow detection in RNS arithmetic is equivalent to the magnitude of the problem of comparison [12, 13]. Another algorithm which proposed for overflow detection in odd dynamic range  $M$  is a ROM-based algorithm and called the parity checking technique. In this method, parity indicates whether an integer number is even or odd. Let operands of  $X$  and  $Y$  have the same parity and  $Z = X + Y$ . So, the addition process is with overflow, if  $Z$  be an odd number [14, 15]. For signed RNS, overflow occurs when the sign of the sum is different from the operands [16].

In this paper, we will propose an algorithm to detect overflow in moduli set  $\{2^n-1, 2^n, 2^{n+1}\}$ . This moduli set is one of the most popular three-module set, and can also be extended to improve the RNS dynamic range [17]. In proposed method, numbers  $[0, M - 1]$  are distributed among several groups. Then, by using their group numbers, is diagnosed in the process of addition of two numbers, whether overflow has occurred or no.

## 2. Proposed Method

To detect overflow in moduli set  $\{2^n-1, 2^n, 2^{n+1}\}$ , we distribute the numbers in dynamic representation range  $M$  into several groups. Since, residue representation of  $X$  in mentioned moduli set is corresponding with  $(x_1, x_2, x_3)$ , so its group number obtained according to Fig.1.

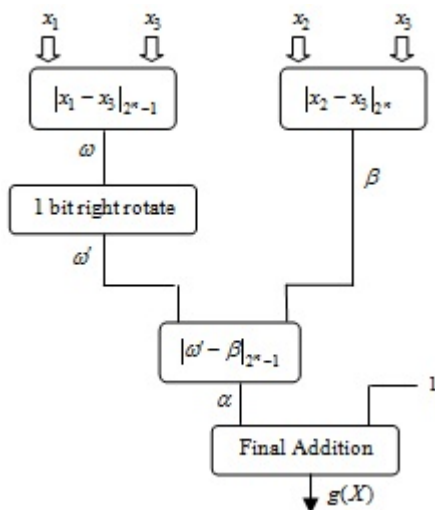


Fig. 1 Group Number Detection.

The number of groups required for this distribution is equal to  $\gamma$  and can be expressed as

$$\gamma = \left\| |x_1 - x_3|_{2^n-1} - |x_2 - x_3|_{2^n} \right\|_{2^n-1} = 2^n - 1. \quad (4)$$

So, we can concluded that length of any group namely  $l$  is given as

$$l = \frac{M}{\gamma} = \frac{(2^n - 1) \cdot 2^n \cdot (2^n + 1)}{2^n - 1} = 2^n \cdot (2^n + 1). \quad (5)$$

In any of these groups there are  $2^n$  subgroups, because

$$\beta = |x_2 - x_3|_{2^n}, \quad 0 \leq \beta \leq 2^n - 1. \quad (6)$$

For example, the value of  $\beta$  for numbers in first group with range  $[0, 2^{2n} + 2^n)$  is shown in the following:

$$\beta = |x_2 - x_3|_{2^n} \Rightarrow \begin{cases} 0 \leq X < 2^{n+1}, & \beta = 0 \\ 2^{n+1} \leq X < 2(2^{n+1}), & \beta = 1 \\ \vdots & \\ (2^n-1)(2^{n+1}) \leq X < 2^n(2^{n+1}), & \beta = 2^n-1. \end{cases} \quad (7)$$

For determination of group number of any residue number, first should be get the value of  $\omega$ . For clarity, we have exhibited it in range  $[0, 2^{2n} + 2^n)$  as follows:

$$\omega = |x_1 - x_3|_{2^n-1} \Rightarrow \begin{cases} 0 \leq X < 2^{n+1}, & \omega = 0 \\ 2^{n+1} \leq X < 2(2^{n+1}), & \omega = 2 \\ 2(2^{n+1}) \leq X < 3(2^{n+1}), & \omega = 4 \\ \vdots & \\ (2^{n-1}-1)(2^{n+1}) \leq X < 2^{n-1}(2^{n+1}), & \omega = 2^n-2 \\ 2^{n-1}(2^{n+1}) \leq X < (2^{n-1}+1)(2^{n+1}), & \omega = 1 \\ \vdots & \\ (2^n-2)(2^{n+1}) \leq X < (2^n-1)(2^{n+1}), & \omega = 2^n-3 \\ (2^n-1)(2^{n+1}) \leq X < 2^n(2^{n+1}), & \omega = 0. \end{cases} \quad (8)$$

According to (8) and with regard to the product result from moduli subtraction in each group be appeared first, odd values and afterward even respectively. Since, in order to accomplishment of arithmetic operations should be arranged the  $\omega$  values increasingly, so it is achievable through one bit right rotate. Therefore, if assume  $\omega = 0, 2, 4, 6, \dots, 2^n - 2, 1, 3, \dots, 2^n - 3$ , after 1-bit right rotate, we get  $\omega' = 0, 1, 2, \dots, 2^n - 3, 2^n - 2$ .

Now by having the values of  $\beta$  and  $\omega'$ , the group number of any residue number in RNS (counting from 0) is defined as

$$\alpha = |\omega' - \beta|_{2^n - 1}, \quad 0 \leq \alpha \leq 2^n - 2. \quad (9)$$

For facility in implementation of proposed algorithm, we add the obtained group number from (9) with one. In this case, if  $X$  be an integer, its group number  $g(X)$  is

$$g(X) = \alpha + 1, \quad 1 \leq g(X) \leq 2^n - 1. \quad (10)$$

Table 1 shows the distribution of numbers in dynamic range  $[0, 2^{3n} - 2^n]$  which is given as a product of the  $m_i$ 's in moduli set  $\{2^n - 1, 2^n, 2^{n+1}\}$ .

Number	Group
$0 \rightarrow 2^n (2^{n+1}) - 1$	1
$2^n (2^{n+1}) \rightarrow 2[2^n (2^{n+1})] - 1$	2
$\vdots$	
$(2^n - 2)[2^n (2^{n+1})] \rightarrow (2^n - 1)[2^n (2^{n+1})] - 1$	$\gamma$

Let  $X$  and  $Y$  are two operands in the process of addition  $Z = X + Y$  and also  $g(X)$  and  $g(Y)$  be the group number of operands, respectively. It can be shown from Table 1 that:

- i) if  $g(X) + g(Y) < 2^n$ , no overflow will occur.
- ii) if  $g(X) + g(Y) > 2^n$ , overflow must occur.
- iii) if  $g(X) + g(Y) = 2^n$ , overflow may or may not occur. So, is required it be checked more.

**Proof:** in case iii, range of the sum  $X + Y$  in binary system is

$$(2^n - 2)[2^n (2^n + 1)] \leq Z \leq 2^n [2^n (2^n + 1)] - 2. \quad (11)$$

Since,  $M$  is exactly located in middle of obtained range from (11), so it can be rewritten as

$$(2^n - 2)[2^n (2^n + 1)] \leq M \leq 2^n [2^n (2^n + 1)] - 2. \quad (12)$$

In order to proof of  $g(X) + g(Y) = 2^n$ , we replace the values of  $(2^n - 2)$  and  $2^n$  in terms of  $g(X) + g(Y)$ .

Therefore, the final form of (12) is

$$\begin{aligned} & ((g(X) - 1) + (g(Y) - 1))[2^n (2^n + 1)] \\ & < (2^n - 1)2^n (2^n + 1) < \\ & (g(X) + g(Y))[2^n (2^n + 1)] \end{aligned} \quad (13)$$

As seen, value of  $2^n (2^n + 1)$  is common in the sides of inequality (13), thus it can be eliminated as follows:

$$g(X) + g(Y) - 2 < 2^n - 1 < g(X) + g(Y) \quad (14)$$

After adding one whit the sides of (14), the resulting inequality be defined as

$$g(X) + g(Y) - 1 < 2^n < g(X) + g(Y) + 1. \quad (15)$$

Finally (15) can be divided by two parts, that is

$$\begin{cases} g(X) + g(Y) < 2^n + 1 \\ g(X) + g(Y) > 2^n - 1 \end{cases} \Rightarrow g(X) + g(Y) = 2^n. \quad (16)$$

Therefore, overflow can be detected by comparing the sum of the groups of operands with  $2^n$ . If the sum exceeds  $2^n$ , overflow must occur. Notice that, overflow probability should be again checked in third mode. For this purpose,  $g(X) + g(Y) = 2^n$  is given 1-bit shift to right as  $2^n / 2 = 2^{n-1}$ . Subsequently, it be compared with group number of sum of operands  $g(Z)$ . In this case, if  $g(Z) > 2^{n-1}$ , then overflow does not exist and otherwise  $g(Z) < 2^{n-1}$  overflow has occurred. Fig. 2 shows the overflow detection circuit in moduli set  $\{2^n - 1, 2^n, 2^{n+1}\}$ .

Table2: Group Number Calculations for RNS  $\{15,16,17\}$

$X$	$X_{RNS}$	$\beta$	$\omega'$	$\alpha =  \omega' - \beta _{15}$	$g(X)$
62	(2, 14, 11)	3	3	0	1
1045	(10, 5, 8)	13	1	3	4
1111	(1, 7, 6)	1	5	4	5
2040	(0, 8, 0)	8	0	7	8
2048	(8, 0, 8)	8	0	7	8
3097	(7, 9, 3)	6	2	11	12
4079	(14, 15, 16)	15	14	14	15

As an example, consider moduli set  $\{15,16,17\}$ . Therefore  $M = 15 \times 16 \times 17 = 4080$  and the number of groups  $\gamma = 15$ . The example calculation for the distribution of a few values of numbers are shown in Table 2. If  $X = 1111$  and  $Y = 2048$ , then  $Z = X + Y = 3159 < 4079$ . In RNS, according to Table 2, groups of operands are equal to 5 and 8 respectively. Based on proposed method, because sum of the group of operands 13 is less than 16, thus no overflow exits. Another instance of overflow consists of:

$$X = 1045 = (10, 5, 8) \rightarrow g(X) = 4$$

$$Y = 3097 = (7, 9, 3) \rightarrow g(Y) = 12$$

Since,  $g(X) + g(Y) = 16 = 2^n$  therefore, is required  $g(Z)$  be compared with  $2^{n-1} = 8$

$$Z = |X + Y|_M = (2, 14, 11) \rightarrow g(Z) = 1$$

According to our algorithm  $1 < 8$  and it denotes that an overflow has occurred. In the other words, we have:  $Z = X + Y = 4142 > 4080$ .



### 3. Hardware Implementation

The group detection function is determined by Eq.(10) as a sum of  $\alpha$  and 1. The value of  $\alpha$  is given by  $\alpha = |\omega' - \beta|_{2^n-1}$ . Since  $\alpha$  is computed as a residue modulo  $2^n-1$  then, instead of subtracting  $|\beta|_{2^n-1}$  we can add its additive inverse modulo  $2^n-1$ . An additive inverse modulo  $2^n-1$  is simply a negation of binary representation. For simplification reasons the additive inverse of  $|\beta|_{2^n-1}$  is denoted as

$$\hat{\beta} = \left| -|\beta|_{2^n-1} \right|_{2^n-1}. \quad (17)$$

So that, the binary form of (17) is  $\hat{\beta} = \bar{\beta}_{n-1}, \dots, \bar{\beta}_1, \bar{\beta}_0$ .

Thus (9) can be rewritten as the sum

$$\alpha = \left| \omega' + \hat{\beta} \right|_{2^n-1}. \quad (18)$$

From [18], an addition modulo  $(2^n - 1)$  with redundant zero elimination can be expressed as

$$\left| a + b \right|_{2^n-1} = \left| a + b + c_{out} + p \right|_{2^n} \quad (19)$$

where  $c_{out}$  is a carry bit of  $a + b$  addition and  $p = 1$  for  $a + b = 11 \dots 1_2$ . The sum  $c_{out} + p$  is 0 for  $a + b < 2^n - 1$  and 1 for  $a + b \geq 2^n - 1$  [1]. By assuming that  $C_{in} = c_{out} + p$ , the final form of (18) is then

$$\alpha = \left| \omega' + \hat{\beta} + C_{in} \right|_{2^n}. \quad (20)$$

Also, the values of  $\omega$  and  $\beta$  is given using this way. Notice that, in computing of  $\omega = |x_1 - x_3|_{2^n-1}$ , because  $x_3$  is a residue number modulo  $2^n + 1$  and  $x_3 \leq 2^n$  then  $|x_3|_{2^n-1}$  is given by OR-ing the least and the most significant bits of  $x_3$ . Therefore, binary form of  $\hat{x}_3$  is  $\overline{x_{3,0} + x_{3,n} + \bar{x}_{3,n-1} + \dots + \bar{x}_{3,1}}$ .

To overflow detection, should be compared  $g(X) + g(Y)$  with  $2^n$ . We know the number comparison in RNS is one of the difficult and time consuming operations, therefore attempted to do this operation whit another way. In this paper, in order to  $g(X) + g(Y)$  addition, we designed a new circuit that just generates the required valves. The outputs of this unit are the most significant bit (MSB) of the sum as ( $M$ ), a carry bit of the sum namely  $C$  and  $P'_1$  where if be equal to one, denotes the all the bits of the sum, except  $M$ , are zero. Hence, mentioned unit is called  $MCP'_1G$ . Consequently, comparison operation performs as following:

$$g(X) + g(Y) \begin{cases} < 2^n, & \text{if } C = 0 \\ = 2^n, & \text{if } C = 1, M = 0, P'_1 = 1 \\ > 2^n, & \text{otherwise.} \end{cases} \quad (21)$$

As mentioned above, whenever  $g(X) + g(Y) = 2^n$ , is required to overflow probability be checked again. For this propose,  $g(Z)$  be compared with  $2^{n-1}$ . In this case, by having the MSB of  $g(Z)$  as  $W = S'_{n-1}$  and its  $P'_2 = P'_{0,n-2}$ , can be said:

$$g(Z) \begin{cases} > 2^{n-1}, & \text{if } W = 1, P'_2 = 0 \\ < 2^{n-1}, & \text{otherwise.} \end{cases} \quad (22)$$

The proposed method to overflow detection is implemented as shown in Fig. 2. The circuit consists of five main blocks: three group detection units, a unit for generation of (MSB), output carry and  $P'_1$  of  $g(X) + g(Y)$  addition and the final post-processing unit to detecting overflow.

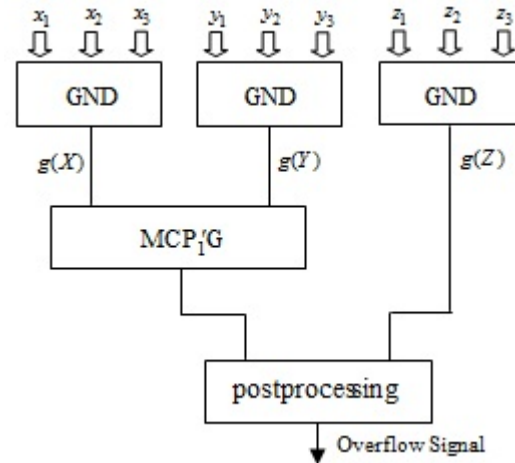


Fig. 2 Overflow detection unit.

The group number detection unit shown in Fig.1 is used for determination of group number of operands and their sum. These values are represented as three vectors  $g(X)$ ,  $g(Y)$  and  $g(Z)$  respectively. The produced vectors are connected to the inputs of the  $MCP'_1G$  unit.

The goal of the  $MCP'_1G$  unit is to determination the  $C = c_n$ ,  $M = P_{n-1,n-1} \oplus G_{0,n-2}$  where  $G_{0,n-2}$  is the carry of the  $(n-1)$ -bit of  $g(X) + g(Y)$  from the position 0 to  $n-2$  and also generation of  $P'_1 = P'_{0,n-2}$  which detects a result in the form of  $X00 \dots 0_2$ . These signals can be

computed in a simplified and new prefix structure proposed in [18]. Hence, we no need to use a full  $n$ -bit adder.

A parallel prefix adder and also parallel prefix adder with end-around-carry are built from elements shown in Fig. 3.

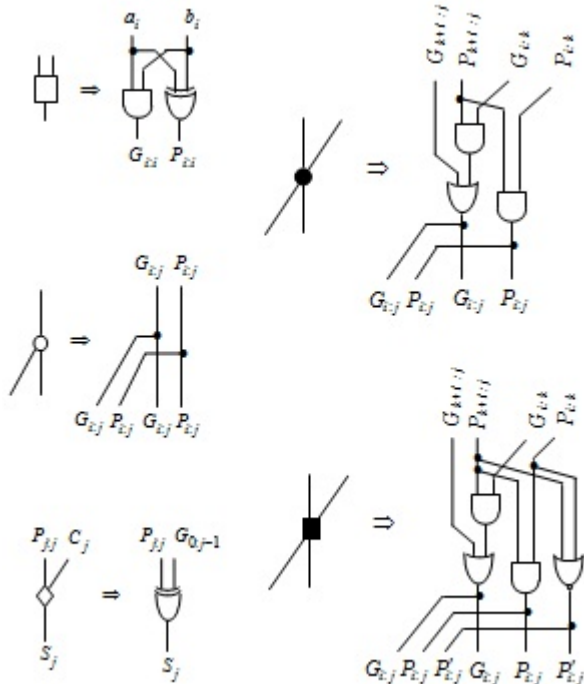


Fig. 3 Blocks of prefix adder.

The signals  $G_{i,j}$  and  $P_{i,j}$  are the carry generation and propagation functions from the position  $i$  to  $j$ . The  $P'_{i,j}$  signal is a function where indicates whether all the bits from the position  $i$  to  $j$  are equal zero or no. For an addition of two binary vectors  $a_{n-1}...a_0$  and  $b_{n-1}...b_0$  and for  $i < k < j$ , these functions can be expressed by logic equations

$$\begin{aligned}
 G_{i,i} &= a_i \cdot b_i \\
 P_{i,i} &= a_i \oplus b_i \\
 G_{i,j} &= G_{i,k} \cdot P_{k+1:j} + G_{k+1:j} \\
 P_{i,j} &= P_{i:k} \cdot P_{k+1:j} \\
 P'_{i,j} &= \overline{P_{i:k} + P_{k+1:j}}
 \end{aligned} \tag{23}$$

The carry signals  $c_j$  are equal to  $G_{0,j-1}$  and the bits  $s_j$  of a final sum are  $s_j = P_{j,j} \oplus c_j$ . An addition advantage of prefix structures is that the end-around carry can be added in the last stage with a delay cost of two logic levels [1]. The detailed description of this idea is presented in [18].

Fig.4. depicts the structure of parallel prefix adder with end-around-carry (PPA with EAC). We applied it for doing addition operations in order to obtain the values of  $\alpha$  and  $\omega$ .

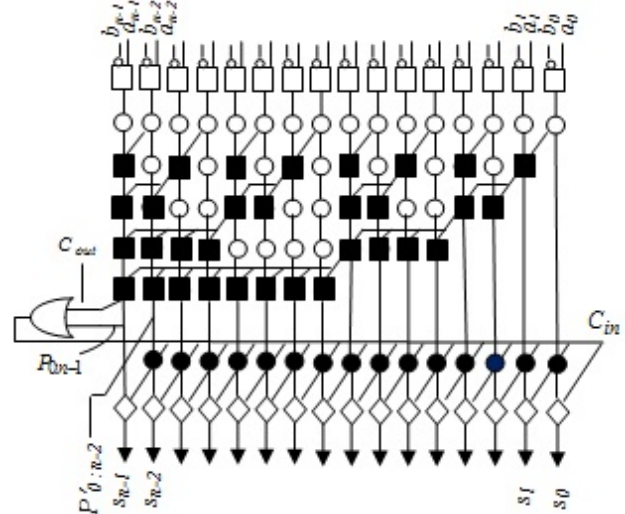


Fig. 4 Parallel prefix adder structure with End-around-carry.

The possibility of adding one bit with the delay of two logic levels enables computation of  $M$  and  $g(X) = \alpha + 1$ . Since  $s_j = P_{j,j} \oplus c_j$ , then  $M$  is given in the additional stage of new parallel-prefix adder. The value of  $M$  from  $M = P_{n-1:n-1} \oplus G_{0:n-2}$  is computed by EX-ORing of  $P_{n-1:n-1}$  and  $G_{0:n-2}$  of the MCP<sub>1</sub>G unit. The full circuit to evaluate  $M$  for  $n = 16$  is shown in Fig. 5.

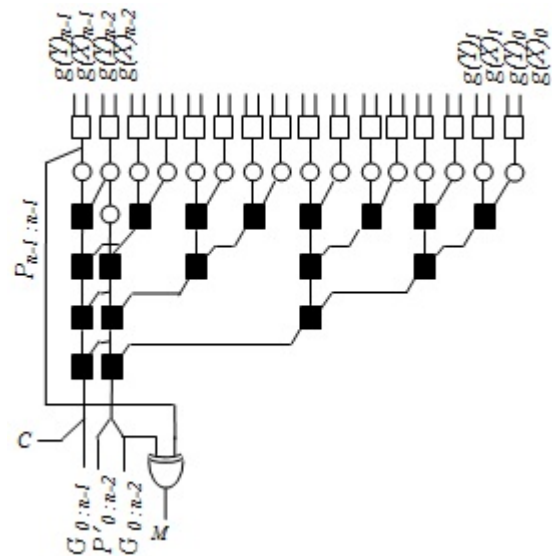


Fig. 5 MCP<sub>1</sub>G unit for  $n = 16$ .

The post-processing unit block diagram is shown in Fig. 6. It comprises a limited number of components, which can detect the overflow in process of addition of two numbers.

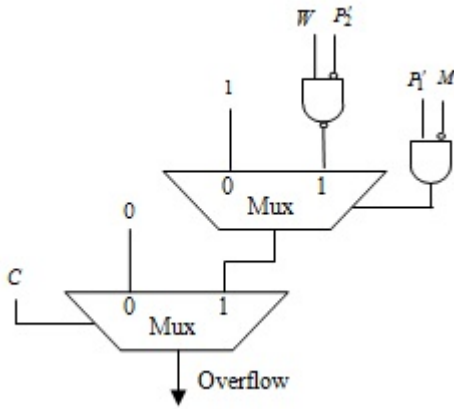


Fig. 6 post-processing unit.

The area and time (AT) characteristics of proposed circuit in order to overflow detection for RNS by moduli set  $\{2^n - 1, 2^n, 2^{n+1}\}$  are estimated using the standard unit-gate model used [18]. In this model, each gate of two-input such as AND, OR, NAND, NOR has area  $A = 1$  and delay  $T = 1$ . Also, for each 2-input gate XOR / XNOR there are  $A = T = 2$ .

The group number detection unit of shown in Fig. 1. comprises three main adders: one modulo  $(2^n)$  adder and two adder mod  $(2^n - 1)$ . For calculation of  $\beta$  modulo  $2^n$ , we used the parallel adder structure from [18] by  $A = 5n + (3/2)n \log_2 n$  and  $T = 2 \log_2 n + 4$ . As we said previously, for determination of values  $\omega$  and  $\alpha$ , applied the PPA with EAC (Fig. 4) which it uses  $(n - 1)$  black nodes,  $n$  input nodes (as square) and  $n / 2$  black square in each level where number of levels is equal be  $\log_2 n$ . Thus, AT parameters of any adder modulo  $(2^n - 1)$  are

$$\begin{aligned} A_{adder \text{ mod } 2^n - 1} &= 8n + 2n \log_2 n - 3 \\ T_{adder \text{ mod } 2^n - 1} &= 2 \log_2 n + 6. \end{aligned} \quad (24)$$

After the value determination of  $\alpha$  for group detection of each number, should be add  $\alpha$  with 1. Therefore,  $g(X)$  also obtains in the additional stage of PPA with EAC from Fig. 4. which it requires the hardware of  $2n$  and delay of 2 logic levels (see Fig. 7). Consequently, area and delay of GND unit are

$$\begin{aligned} A_1 &= 23n + \frac{11}{2} n \log_2 n - 6 \\ T_1 &= 4 \log_2 n + 14. \end{aligned} \quad (25)$$

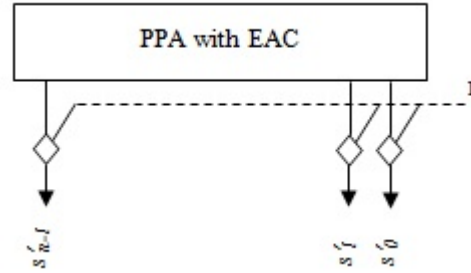


Fig.7 Final addition unit

The requirements for MCP'<sub>1</sub>G unit (Fig.5) are as follows:  $n$  input nodes,  $(n + 2)$  black square, and an additional gate. For determination of delay should be noticed the maximum number of black square that is required to working in parallel is  $\log_2 n$ . So, the MCP'<sub>1</sub>G area and delay can expressed as

$$\begin{aligned} A_2 &= 7n + 10 \\ T_2 &= 2 \log_2 n + 4. \end{aligned} \quad (26)$$

The post-processing unit contains a limited number of the gates and multiplexers. Notice that, a Mux<sub>2:1</sub> has  $A = 3$  and  $T = 2$ . So, the AT parameters of mentioned unit are

$$\begin{aligned} A_3 &= 8 \\ T_3 &= 5. \end{aligned} \quad (27)$$

Total delay of the circuit is determined by a path consisting one unit of group detection, MCP'<sub>1</sub>G unit and post-processing unit. The total area and delay of the designed overflow detection circuit are

$$\begin{aligned} A_{tot} &= 3A_1 + A_2 + A_3 = 76n + \frac{33}{2} n \log_2 n \\ T_{tot} &= T_1 + T_2 + T_3 = 6 \log_2 n + 23. \end{aligned} \quad (28)$$

## 4. Comparison

One of the fastest and most efficient RNS comparator for the moduli set  $\{2^n - 1, 2^n, 2^{n+1}\}$  are introduced in references [17] and [19] respectively. In a generic approach, after a residue to binary convert, comparison operation can be done by using  $n$  or  $(n + 1)$  bits comparator which has a delay of residue to binary converter plus delay of a  $(n + 1)$  bit Binary Comparator (BC). In Table 3 proposed technique is compared with other methods.

Table 3: Comparison Area and Delay of proposed method with other methods using unit-gate model

<i>Design</i>	<i>Area</i>	<i>Delay</i>
[17]	$115n + 186$	$4n + \log_2 n + 36$
[19]	$96n + n \log_2 n + 16$	$8n + \log_2 n + 12$
[20] - CI	$56n + 22 + A_{BC}$	$16n + 4 + \tau_{BC}$
[20] - CII	$96n + 24 + A_{BC}$	$4n + 4 + \tau_{BC}$
[20] - CIII	$80n + 18 + A_{BC}$	$4n + 4 + \tau_{BC}$
Proposed method	$76n + (33/2)n \log_2 n$	$6 \log_2 n + 23$

The most effective overflow detection circuit based on reverse converters can be built on the base on Converter I from [20]. In Converter I and also Converters II and III from [20], the minimum delay is  $O(n)$  whereas, delay of proposed method is factor of  $O(\log_2 n)$ .

As seen from Table 3, the proposed approach for overflow detection in moduli set  $\{2^n-1, 2^n, 2^n+1\}$  is faster than previous works. However, the hardware cost of the presented method is more. It is essential to remark that, although the proposed design consumes more hardware but it demonstrates significant improvement in terms of delay, especially for large  $n$ . Furthermore, our proposed method detects overflow without applying a complete comparator or reverse converter.

## 5. Conclusions

Detecting overflow is one of the most important and complex operations in residue number system. In this paper, a novel and different method has been presented for detecting overflow in moduli set  $\{2^n-1, 2^n, 2^n+1\}$ . Our proposed technique is based on group of numbers which leads to the correct result without doing a complete comparison or need to use the residue to binary converter. The presented approach has significant reduction in delay, compared to other methods.

## References

[1] T. Tomczak, "Fast Sign Detection for RNS  $\{2^n-1, 2^n, 2^n+1\}$ ", IEEE Transactions on Circuits and Systems I: Regular Papers, Vol. 55, Iss. 6, 2008, pp. 1502-1511.  
 [2] N. S. Szabo and R. I. Tanaka, Residue Arithmetic and Its Application to Computer Technology, New York: McGraw-Hill, 1967.  
 [3] W. A. Chren, Jr. "A new residue number system division algorithm", Comput. Math. Appl., Vol. 19, No. 7, 1990, pp. 13-29.

[4] H. Bronnimann, I. Z. Emiris, V. Y. Pan, and S. Pion, "Computing exact geometric predicates using modular arithmetic with single precision", in Proc. 13<sup>th</sup> Annu. Symp. Comput. Geom., ACM press, 1997.  
 [5] R. C. Debnath and D. A. Pucknell, "On Multiplicative Overflow detection in Residue Number System", Electronics Letters, Vol. 14, No. 5, 1978  
 [6] R. Conway and J. Nelson, "Improved RNS FIR Filter Architectures", IEEE Trans. On Circuits and Systems-II: Express Briefs, Vol. 51, No.1, 2004.  
 [7] P. G. Fernandez and et al., "A RNS-Based Matrix-Vector-Multiply FCT Architecture for DCT Computation", Proc. 43<sup>th</sup> IEEE Midwest Symposium on circuits and Systems 2000, pp. 350-353.  
 [8] L. Yang and L. Hanzo, "Redundant Residue number System Based ERROR Correction Codes", IEEE VTS 54<sup>th</sup> on Vehicular Technology Conference, 2001, Vol. 3, pp. 1472-1467.  
 [9] J. Ramirez, et al., "Fast RNS FPL-Based Communication", Proc. 12<sup>th</sup> Int'l Conf. Field Programmable Logic, 2002, pp. 472-481.  
 [10] R. Rivest, A. Shamir, and L. Adleman, "A Method for obtaining Digital Signatures and Public Key Cryptosystems", Comm. ACM, Vol. 21, No. 2, 1948, pp. 120-126.  
 [11] J. Bajard, and L. Imbert, "A Full RNS Implementation of RSA", IEEE Transactions on computers, Vol. 53, No. 6, 2004, pp. 769-774.  
 [12] B. Parhami, "Computer arithmetic: algorithms and hardware designs", New York : Oxford University Press, 2000.  
 [13] M. Askarzadeh, M. Hosseinzadeh and K. Navi, "A New approach to overflow detection in moduli set  $\{2^n-3, 2^n-1, 2^n+1, 2^n+3\}$ ", Second International Conference on Computer and Electrical Engineering, 2009, pp. 439-442.  
 [14] M. shang, H. JianHao, Z. Lin and L. Xiang, "An efficient RNS parity checker for moduli set  $\{2^n-1, 2^n+1, 2^{2n}+1\}$  and its applications", Springer Journal of Science in China Series F: Information Sciences", Vol. 51, No. 10, 2008, pp. 1563-1571.  
 [15] A. Omondi and B. Premkumar, "Residue Number Systems: Theory and Implementation", Imperial College Press, 2007.

- [16] M. Rouhifar, M. Hosseinzadeh and M. Teshnehlab, "A new approach to Overflow detection in moduli set  $\{2^n-1, 2^n, 2^n+1\}$ ", International Journal of Computational Intelligence and Information Security, Vol. 2, No.3, 2011, pp. 35-43.
- [17] E. Gholami, R. Farshidi, M. Hosseinzadeh and K. Navi, "High speed residue number system comparison for the moduli set  $\{2^n-1, 2^n, 2^n+1\}$ ", Journal of communication and computer, Vol. 6, No. 3, 2009, pp. 40-46.
- [18] R. Zimmerman, "Efficient VLSI implementation of modulo  $(2^n \pm 1)$  addition and multiplication", in Proc. 14<sup>th</sup> IEEE Symp. Comput. Arithm. , 1999, pp. 158-167.
- [19] BI. Shao-quiang and W. J. Groos, "Efficient residue comparison algorithm for general Moduli sets", IEEE International Circuits and Systems, 2005, pp. 1601-1604.
- [20] Y. Wang, X. Song, M. Aboulhamid and H. Shen, "Adder based residue to binary converters for  $\{2^n-1, 2^n, 2^n+1\}$ ", 2002, pp. 1772-1779.

**Mehrin Rouhifar** received her B.Sc. in Computer Software Engineering from Islamic Azad University, Shabestar branch, Iran in 2008. Recently, she is received the M.Sc. degree in Computer System Architecture from Islamic Azad University, Tabriz branch, Iran in 2011. Her main research interests include Computer Arithmetic, Residue Number System, VLSI Design and Network reliability.

**Mehdi Hosseinzadeh** was born in Dezful, a city in the southwestern of Iran, in 1981. Received B.Sc. in Computer Hardware Engineering from Islamic Azad University, Dezful branch, Iran in 2003. He also received the M.Sc. and Ph.D. degrees in Computer System Architecture from the Science and Research Branch, Islamic Azad University, Tehran, Iran in 2005 and 2008, respectively. He is currently Assistant Professor in Department of Computer Engineering of Science and Research Branch of Islamic Azad University, Tehran, Iran. His research interests are Computer Arithmetic with emphasis on Residue Number System, Cryptography, Network Security and E-Commerce.

**Saeid Bahanfar** received the B.Sc. degree in Computer Software Engineering from Payam Noor University (PNU), Tabriz branch, Iran in 2008. Currently, he is a M.Sc. student of Computer System Architecture in Islamic Azad University, Tabriz branch, Iran. His research interests include Residue Number System and VLSI Design.

**Mohammad Teshnehlab** is professor at Department of Control Engineering, Faculty of Electrical Engineering, K. N. Toosi University, Tehran, Iran. His current research interests include Fuzzy, Neural Network, Soft Computing, Evolutionary Filtering and Simultaneous Localization and Mapping.



# A Novel Feature Selection method for Fault Detection and Diagnosis of Control Valves

Binoy B. Nair<sup>1</sup>, Vamsi Preetam M. T.<sup>2</sup>, Vandana R. Panicker<sup>3</sup>, Grishma Kumar V.<sup>4</sup> and Tharanya A.<sup>5</sup>

<sup>1</sup> Assistant Professor (Sr.),  
Department of Electronics and Communication Engineering  
Amrita Vishwa Vidyapeetham,  
Coimbatore, Tamilnadu, India

<sup>2</sup> Department of Electronics and Communication Engineering  
Amrita Vishwa Vidyapeetham,  
Coimbatore, Tamilnadu, India

<sup>3</sup> Department of Electronics and Communication Engineering  
Amrita Vishwa Vidyapeetham,  
Coimbatore, Tamilnadu, India

<sup>4</sup> Department of Electronics and Communication Engineering  
Amrita Vishwa Vidyapeetham,  
Coimbatore, Tamilnadu, India

<sup>5</sup> Department of Electronics and Communication Engineering  
Amrita Vishwa Vidyapeetham,  
Coimbatore, Tamilnadu, India

## Abstract

In this paper, a novel method for feature selection and its application to fault detection and Isolation (FDI) of control valves is presented. The proposed system uses an artificial bee colony (ABC) optimized minimum redundancy maximum relevance (mRMR) based feature selection method to identify the important features from the measured control valve parameters. The selected features are then given to a naïve Bayes classifier to detect nineteen different types of faults. The performance of the proposed feature selection system is compared to that of six other feature selection techniques and the proposed system is found to be superior.

**Keywords:** *Feature Selection, Control Valves, Fault Detection and Diagnosis, Artificial bee colony, Feature selection, naïve Bayes.*

## 1. Introduction

Control valves are extensively used in industry to control various parameters such as flow, temperature, pressure, liquid level etc. For this reason it is of vital importance

that its condition is monitored continuously and deviations in its function be noted to prevent and control hazardous consequences that may follow. Timely fault detection and diagnosis in control valves can be used to develop maintenance strategies and consequently, the plant's overall downtime and hence, the resulting maintenance costs can be brought under control.

Fast Fourier Transforms (FFTs) are one of the oldest methods for FDI and have been widely used in fault detection. FFTs have been used for fault detection in gas turbine engines [1], rotor bars [2], [3] and induction motors [4]. Statistical techniques like multivariate statistical projection method (MSPM) [5] and dynamic PCA have been used [6] with varying degrees of success. Signal processing techniques like the Kalman filter [7], [8] have also been employed. Application of data mining techniques like support vector machines (SVMs) [9], [10], [11], [12], artificial neural networks (ANNs) have also been widely used [13], [14], [15], [16], [17], [18], [19], [20], [21], [22].

However, it is seen that the features required for the purpose of classification of faults are usually selected based on expert knowledge rather than automatically. A suboptimal feature set can compromise the accuracy of the classification system, leading to poor performance of the system. In the present study, a feature selection mechanism which can identify the importance of features from a large feature set is proposed. The performance of the proposed system is compared to that of six other feature selection techniques and the proposed system was found to outperform all the other techniques considered, by producing highest classification accuracy for the smallest number of faults. The dataset used for validating the proposed system is the Development and Application of Methods for Actuator Diagnosis in Industrial Control Systems (DAMADICS) standard benchmark dataset.

The rest of the paper is organized as follows: section 2 presents an overview of the DAMADICS benchmark, section 3 presents the design of the proposed system and the results are presented in section 4.

## 2. DAMADICS

This section presents the overview of the DAMADICS benchmark dataset used in the present study.

DAMADICS benchmark was developed for real time training of an actuator system [23],[24]. This benchmark has become a standard for analyzing wide range of FDI methods in terms of standard performance. The DAMADICS benchmark data was designed for comparing various FDI methods by real time testing on industrial actuators in the Lublin sugar factory in Poland. The benchmark is based on the complete working of electro – pneumatic valve actuator used in almost all industrial applications. The testing was performed by inducing abrupt (sudden) and incipient (gradually developing) faults to the actuators and recording the data.

The structure of benchmark actuator system [23], [24] is given in Fig. 1. For designing the benchmark data, five available measurements and one control value signal have been considered (measurements being made at every second). They are: process control external signal CV, values of liquid pressure on the valve inlet P1' and outlet P2', stem displacement X', liquid flow rate F' and liquid temperature T'. The apostrophe denotes signals that are measured. The set of main variables used in benchmark, as given in Fig. 1 is as follows: CV (process control external signal), CVI (internal current acting on E/P unit), E/P (electro-pneumatic transducer), F (main pipeline flow rate), Fv (control valve flow rate), Fv3 (actuator by-pass pipeline flow rate), FT (flow rate transmitter), P (positioned), P1,P2

(pressures on valve: inlet and outlet), Ps E/P (transducer output pressure), PSP (positioner supply pressure unit), PT (pressure transmitter), Pz (positioner air supply pressure), S (pneumatic servo-motor), T1 (liquid temperature), TT (temperature transmitter), V (control valve), V1, V2 and V3 (cut-off valves), X (valve plug displacement), ZC (internal controller), ZT (stem position transmitter).

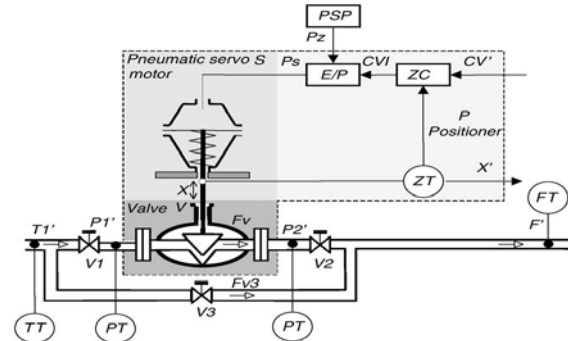


Fig.1 Structure of benchmark actuator system [23],[24]

There are 19 actuator faults which have been considered in the benchmark study [40]. These faults are:

Valve clogging ( $f_1$ ), Valve plug or Valve seat sedimentation ( $f_2$ ), Valve plug or Valve seat erosion ( $f_3$ ), increase of valve friction ( $f_4$ ), external leakage ( $f_5$ ), internal leakage ( $f_6$ ), medium evaporation or critical flow ( $f_7$ ), twisted servomotor stem ( $f_8$ ), servomotor housing or terminal tightness ( $f_9$ ), servomotor diaphragm perforation ( $f_{10}$ ), servomotor spring fault ( $f_{11}$ ), electro pneumatic transducer fault ( $f_{12}$ ), stem displacement sensor fault ( $f_{13}$ ), pressure sensor fault ( $f_{14}$ ), positioner spring fault ( $f_{15}$ ), positioner supply pressure drop ( $f_{16}$ ), unexpected pressure change across valve ( $f_{17}$ ), fully or partly opened bypass valve ( $f_{18}$ ) and flow rate sensor fault ( $f_{19}$ ).

The faults are grouped based on the severity of the fault as abrupt (large, medium and small) and incipient [23],[24]. The dataset consists of the data for the following simulated fault groups: abrupt-large for  $f_1, f_2, f_7, f_8, f_{10}, f_{11}, f_{12}, f_{13}, f_{14}, f_{15}, f_{16}, f_{17}$  and  $f_{18}$ ; abrupt-medium for  $f_1, f_7, f_8, f_{10}, f_{12}, f_{13}, f_{14}, f_{16}$ , and  $f_{18}$ ; abrupt-small for  $f_1, f_7, f_8, f_{10}, f_{12}, f_{13}, f_{14}, f_{16}$ , and  $f_{18}$  and  $f_{19}$ ; incipient for  $f_2, f_3, f_4, f_5, f_6, f_9, f_{11}$  and  $f_{13}$ . It must be noted that same fault can manifest itself with different levels of severity under different circumstances. Hence some faults, for example,  $f_1$  have been simulated at different fault severity levels, resulting in distinct measured data.

## 3. Design of the Proposed System

The system proposed in the present paper performs FDI in three steps:

**Step 1:** Extract statistical parameters (average, median, minimum, maximum, standard deviation, kurtosis, skew and Variance) using moving windows, from each of the six measured parameters.

**Step 2:** Select important features for fault classification.

**Step 3:** Use the selected features for identification of the fault and its type using Naïve Bayes classifier.

As can be seen, extracting eight parameters from six initial features creates a feature set with  $(6*8) = 48$  features and when taken together with the initial feature set, the total number of features in the feature set becomes 54. Also, the number of measurements made is large (a total of 65535 measurements for each of the initial features) as well. This makes the task of feature selection quite challenging.

### 3.1 Feature Selection

The proposed feature selection system is derived from ABC [25],[26] and mRMR [27] algorithms. This method was developed using principles of the ABC and mutual information (MI) [28].

The ABC algorithm is an optimization algorithm that uses the behavior of the bees while searching for food [25]. A bee colony is an organized team work system where each bee contributes significant information to the system. There are three types of worker bees which involve in collecting nectar viz. employed bees, onlooker bees and scout bees. The ABC algorithm considers the position of food source as the possible solution of the optimization problem and the food source corresponds to the quality (fitness) of the associated solution [26]. The number of the employed bees or the onlooker bees is equal to the number of solutions in the population. The initial population of  $N$  solutions is randomly generated. Each solution is a  $D$ -dimensional vector where  $D$  is the number of parameters to be optimised. They are relevance and redundancy in this case. The population of solutions is subject to repeated search processes by the employed bees, onlooker bees and scout bees. A solution is randomly chosen and compared with the current solution. The objective function used here will be the mRMR function. The fitness function of each solution is given by

$$fit_i = \frac{1}{1 + f(i)} \quad (1)$$

where  $f(i)$  is the objective function of the  $i^{th}$  solution. If the fitness function of the new chosen solution is greater than the existing one, then the new solution is memorized and the old one is discarded. The employed bees share the information i.e, fitness value of the solutions in their memory with the onlooker bees.

The probability of each solution based on its fitness, is calculated by

$$p_i = \frac{fit_i}{\sum_{k=1}^N fit_k} \quad (2)$$

where  $fit_i$  is the fitness value of the solution  $i$  and  $N$  is the number of solutions in the population. Candidate solutions are produced using the formula

$$v_{ij} = x_{ij} + \varphi_{ij} (x_{ij} - x_{kj}) \quad (3)$$

where  $k \in \{1, 2, \dots, N\}$  and  $j \in \{1, 2, \dots, D\}$  and  $\varphi_{ij}$  is a random number ranging between -1 and 1. This ensures that values generated are different from those already existing. And also the newly generated solutions lie within the defined boundary. The parameter exceeding its limit is set to its limit value.

The performance of each candidate solution is compared with that of the existing solution. If the new solution has equal or better fitness value than the old solution, the old one is discarded with the new one occupying its position. Else, the old one is retained. In other words, a greedy selection mechanism is used for the selection process.

If an optimal solution cannot be obtained from a population within the predefined number of cycles i.e. limit then, that population is abandoned and replaced with a new population. ABC algorithm is used here to optimize the redundancy and relevance parameters of mRMR function. The mRMR method proposed in [27] uses the principle of mutual Information. The mutual information between two variables  $A$  and  $B$  can be defined as

$$I(A; B) = \log_2 \left( \frac{P(A|B)}{P(A)} \right) \quad (4)$$

Maximum Relevance orders features based on the mutual information between individual features  $x_i$  and target class  $h$  such that the feature with the highest mutual information is the most relevant feature. The relationship is expressed as follows:

$$\max V_p V_1 = \frac{1}{|G|} \sum_{h \in G} I(h, V) \quad (5)$$

Max Relevance often shows a high inter-dependence among the features. When two features are highly dependent on one another, the class-discriminative power of these two features would not change much if either one of them were to be removed and if not removed they become redundant as they convey the same characteristics. The minimal redundancy condition can be added to select mutually exclusive features of the dataset. The following relationship helps establish the minimum redundancy measure.

$$\min W_p W_1 = \frac{1}{|G|} \sum_{i,j \in G} I(i, j) \quad (6)$$

The criterion combining the above two parameters is called “minimal-redundancy-maximal-relevance”. It was seen that the two measures could be used together to form two combinations for the purpose of improving the feature selection process [26]. The two combinations considered were:

$$\max(V_1 - W_1) \quad (7)$$

$$\max(V_1/W_1) \quad (8)$$

Here Eq. (7) forms MID: Mutual Information Difference criterion and Eq. (8) forms MIQ: Mutual Information Quotient criterion. It was observed in [30] that MID gave a better performance when compared to MIQ. This was found to be the case in the present study, as well.

Redundancy is often a matter of concern when dealing with large datasets. It was noticed that redundancy caused a negative effect on the accuracy of the classifying system. But it cannot be presumed that the relevance factor only facilitated the increase in accuracy. The conditions are seen to be purely situational. That is, depending on the dataset under study, either of the two, relevance or redundancy may drastically affect the percentage of accuracy.

The following expression defines the proposed optimization criterion

$$\max(a \cdot V_1 - b \cdot W_1) \quad (9)$$

where a and b are constants describing the weightage to be given to relevance and redundancy for selecting the optimal feature set,  $V_1$  is relevance and  $W_1$  is redundancy.

The value of the constants a and b are arrived at from the ABC algorithm. This algorithm was noticed to be applicable only for discrete datasets. And so, when continuous datasets are to be analyzed, discretization has to be done. Discretization comes with the disadvantage of loss of data which will further reduce the accuracy. This can be alternated by scaling the data and employing logarithmic functions. Thus, the relationship is modified for continuous datasets as follows:

$$\max(a \times \log(V_1) - b \times \log(W_1)) \quad (10)$$

The output of the proposed feature selection mechanism is the set of features in the decreasing order of importance.

### 3.2 Naïve Bayes Classification

Naïve Bayes classifier is a simple technique for supervised learning based on probability theory and is highly suitable for datasets containing large number of attributes [30]. Also small amounts of noise in the data do not affect the system. It works on Bayes theorem and the relation is given as follows

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (11)$$

where X is a tuple belonging to class C and H is some hypothesis under consideration. If two or more features are highly correlated, then the weightage for that feature is made high by the system and the result of classification is biased towards values with higher weightage thus pulling down the accuracy values.

## 4. Results

From the DAMADICS benchmark data six measurements have been considered, viz. process control external signal (CV), pressure on the valve inlet (P1), pressure on the valve outlet(P2), stem displacement (X), liquid flow rate (F), liquid temperature (T). As the first step, statistical parameters were extracted from each of the six initial features listed above. These parameters are average, median, minimum, maximum, standard deviation, kurtosis, skew and variance. Hence, a total of 54 features including the original six features are obtained.

The features were extracted for different moving averages (MA) and the corresponding accuracies using a naïve Bayes classifier were obtained (see Table 1). Use of MA is equivalent to passing the measured signal through a low pass filter. Increase in the number of points used for calculating the MA implies reduction in the cutoff frequency of the filtered signal. This is used to reduce the noise (which usually manifests itself as high frequency signals) in the measured signal and hence improve the fault detection accuracy of the system. However, the number of points used for calculating the MA cannot be arbitrarily high since, for larger moving averages (or equivalently, lower cut off frequencies), the filtered signal will begin to lose not only the high frequency noise but also the lower frequencies that may be useful. Also, processing time increases with higher moving points which results in slower classification.

Table 1 shows the result of the effort at identifying the optimum number of points for computing MA. An exhaustive search method to maximize the accuracy was carried out. It was observed that 100 point MA is optimal for abrupt medium, abrupt small and incipient faults.

Table 1: Accuracy (%) for faults from statistical features.

Moving Average (points)	Abrupt Small (%)	Abrupt medium (%)	Abrupt large (%)	Incipient (%)
40	72.1	65.0	78.5	32.9
60	70.4	75.9	80.8	39.5
70	73.5	82.8	94.0	49.4
80	79.0	79.0	81.4	37.8
90	72.6	88.9	82.8	65.5
100	<b>89.4</b>	<b>94.3</b>	<b>96.3</b>	<b>69.9</b>
110	81.1	89.4	82.3	60.5
120	87.6	89.0	94.5	49.7

The 54 features extracted using the 100 point MA are subjected to feature selection using following feature selection methods: Relief-F, MI,  $\chi^2$ , MID, MIQ, information gain (info gain), gain ratio and the proposed feature selection algorithm. Various feature selection algorithms applied to 100 point MA window large abrupt fault are shown in Table 2.

Table 2: Accuracy in % vs. no.of features for abrupt large faults (coefficients for proposed method,  $a=0.9106$ ;  $b=0.0131$ )

no. of features	Proposed method	MIQ	MID	Info gain	Gain ratio	$\chi^2$	Relief -F
3	78.1	80.7	85.0	87.5	30.8	87.5	24.8
6	80.8	85.9	85.7	93.3	84.5	93.3	87.3
9	86.8	87.2	88.0	94.0	86.4	94.0	89.6
12	89.1	88.3	88.1	94.7	91.8	94.2	89.9
15	89.2	89.0	89.0	94.5	94.4	94.5	90.3
18	95.1	89.8	83.9	95.1	95.0	95.1	92.3
21	95.4	83.9	84.1	95.2	95.2	95.5	96.1
24	95.9	84.5	84.1	95.8	95.4	95.6	96.6
27	95.9	84.5	84.6	95.7	95.8	95.7	96.4
30	<b>96.8</b>	84.7	84.6	96.3	95.9	96.3	96.4
33	96.3	84.7	84.7	96.3	96.1	96.3	96.3
36	96.3	95.9	84.7	96.3	96.3	96.3	96.3
39	96.3	96.0	95.9	96.3	96.3	96.3	96.3
42	96.3	96.2	96.2	96.3	96.3	96.3	96.3
45	96.3	96.2	96.2	96.3	96.3	96.3	96.3
48	96.3	96.3	96.2	96.3	96.3	96.3	96.3
51	96.3	96.3	96.3	96.3	96.3	96.3	96.3
54	96.3	96.3	96.3	96.3	96.3	96.3	96.3

For Abrupt medium and small faults also the proposed system showed relatively better accuracy of 95.0% for 36 features and 89.5% for 39 features respectively. The accuracies of the various feature selection algorithms for

abrupt medium faults and abrupt small faults are shown in Table 3 and Table 4 respectively.

Table 3: Accuracy in % vs. no.of features for abrupt medium faults (coefficients for proposed method,  $a=0.3462$ ;  $b=0.9386$ )

no. of features	Proposed method	MIQ	MID	Info gain	Gain ratio	$\chi^2$	Relief -F
3	83.9	80.3	83.9	75.2	20.0	75.2	82.4
6	84.2	80.6	84.2	84.0	77.7	84.0	83.2
9	70.1	87.3	87.2	91.2	92.3	91.2	84.6
12	72.4	87.7	87.7	92.1	92.6	92.1	86.2
15	73.6	89.2	89.2	92.1	92.2	92.1	92.8
18	73.9	76.2	76.2	93.2	92.9	94.0	93.3
21	75.7	76.4	76.2	94.0	94.0	94.0	93.4
24	76.7	77.1	77.1	94.0	94.1	94.2	94.6
27	78.4	77.7	77.3	94.3	94.0	94.3	94.6
30	85.1	81.2	78.5	94.3	94.3	94.3	94.5
33	86.2	86.2	85.2	94.3	94.3	94.3	94.5
36	<b>95.0</b>	95.0	86.9	94.3	94.3	94.3	94.5
39	95.0	95.0	95.0	94.3	94.3	94.3	94.5
42	95.0	95.0	95.0	94.3	94.3	94.3	94.5
45	95.0	95.0	95.0	94.3	94.3	94.3	94.3
48	95.0	95.0	95.0	94.3	94.3	94.3	94.3
51	94.3	94.3	94.3	94.3	94.3	94.3	94.3
54	94.3	94.3	94.3	94.3	94.3	94.3	94.3

Table 4: Accuracy vs. no.of features for abrupt small faults (coefficients for proposed method,  $a=0.0021$ ;  $b=0.9386$ )

no. of features	Proposed method	MIQ	MID	Info gain	Gain ratio	$\chi^2$	Relief -F
3	53.9	73.2	78.2	74.5	29.2	74.5	76.1
6	64.6	73.9	78.3	82.7	73.2	82.7	77.7
9	76.1	81.1	81.2	81.7	78.7	81.7	77.7
12	77.2	81.9	81.9	85.7	83.5	85.7	80.0
15	88.0	83.0	83.0	86.1	85.6	86.1	81.8
18	89.3	74.3	74.3	89.0	89.3	89.0	82.2
21	89.4	74.8	74.5	89.3	89.1	89.3	82.3
24	89.4	76.2	76.1	89.3	89.1	89.3	83.4
27	89.1	76.6	76.3	89.4	89.1	89.4	83.4
30	89.1	78.9	77.1	89.4	89.4	89.4	85.3
33	89.1	84.8	84.3	89.4	89.4	89.4	85.3
36	89.1	89.1	85.0	89.5	89.4	89.5	85.9
39	<b>89.5</b>	89.2	89.1	89.4	89.4	89.4	89.4
42	89.5	89.4	89.4	89.4	89.4	89.4	89.4
45	89.4	89.4	89.4	89.4	89.4	89.4	89.4
48	89.4	89.4	89.4	89.4	89.4	89.4	89.4
51	89.4	89.4	89.4	89.4	89.4	89.4	89.4
54	89.4	89.4	89.4	89.4	89.4	89.4	89.4



Results of feature selection for incipient fault are shown in Table 5. It can be observed that the proposed method shows accuracy of 70.7% for 36 features. Information gain shows better accuracy of 71.6% for 9 features. Also other methods like chi square, Relief-F and gain ratio give slightly better results when compared to the proposed method.

Table 5: Accuracy in % vs. no.of features for abrupt small faults (coefficients of the proposed method:  $a=0.9106$ ;  $b=0.0131$ )

no. of features	Proposed method	MIQ	MID	Info gain	Gain ratio	chi <sup>2</sup>	Relief-F
3	52.9	52.9	52.9	64.8	22.2	64.8	54.8
6	56.8	56.8	56.8	69.5	57.3	68.6	66.4
9	69.8	69.8	64.1	<b>71.6</b>	65.0	71.3	69.3
12	67.3	67.3	67.3	71.1	66.0	71.1	67.9
15	70.4	70.3	70.4	71.0	69.2	71.0	68.2
18	70.4	70.3	70.4	69.8	71.0	69.8	68.1
21	70.4	70.3	70.4	69.8	71.3	69.8	67.9
24	70.4	70.3	70.4	69.8	71.0	69.8	68.0
27	70.4	70.3	70.4	69.8	69.8	69.8	69.9
30	70.4	70.3	70.4	69.8	69.8	69.8	69.9
33	70.4	70.3	70.4	69.9	69.9	69.9	69.9
36	70.7	70.7	70.4	69.9	69.9	69.9	69.9
39	70.7	70.7	70.2	69.9	69.9	69.9	69.9
42	70.7	70.7	70.7	69.9	69.9	69.9	69.9
45	70.7	70.7	70.7	69.9	69.9	69.9	69.9
48	70.0	70.0	70.7	69.9	69.9	69.9	69.9
51	69.9	69.9	70.7	69.9	69.9	69.9	69.9
54	69.9	69.9	69.9	69.9	69.9	69.9	69.9

It can be seen from the above results that the proposed feature selection system is well capable of identifying the best features in a dataset and that the FDI system presented in this paper can be successfully used for identifying faults in actuators with a very high degree of accuracy.

## References

- [1] X. Dai, Z. Gao, T. Breikin and H. Wang, "Disturbance Attenuation in Fault Detection of Gas Turbine Engines: A Discrete Robust Observer Design", IEEE Transactions on Systems, Man, And Cybernetics, 2009, Vol. 39, No.2, pp. 243-239.
- [2] B. Ayhan, M. Y. Chow, H. J. Trussell and M.H. Song, "A Case Study on the Comparison of Non-parametric Spectrum Methods for Broken Rotor Bar Fault Detection", in Proceedings of the 29<sup>th</sup> Annual Conference of the Industrial Electronics Society, 2003, Vol. 3, pp. 2835-2840.
- [3] X. Wang and D. Zhang, "Optimization Method of Fault Feature Extraction of Broken Rotor Bar in Squirrel Cage Induction Motors", in Proceedings of the IEEE International Conference on Information and Automation, 2010, pp. 1622-1625.
- [4] G. King, M. Tarbouchi and D. McGaughey, "Rotor Fault Detection in Induction Motors Using the Fast Orthogonal Search Algorithm", in Proceedings of the International Symposium on Industrial Electronics, 2010, pp. 2621-2625.
- [5] J. Liang and N. Wang, "Faults Detection and Isolation Based on PCA: An Industrial Reheating Furnace Case Study", in Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, 2003, Vol. 2, pp. 1193-1198.
- [6] J. Mina and C. Verde, "Fault Detection Using Dynamic Principal Component Analysis by Average Estimation", in Proceedings of the 2nd International Conference on Electrical and Electronics Engineering (ICEEE) and XI Conference on Electrical Engineering (CIE), 2005, pp. 374-377.
- [7] N. Tudoroiu and M. Zaheeruddin, "Fault Detection And Diagnosis Of Valve Actuators In HVAC Systems", in Proceedings of the IEEE Conference on Control Applications, 2005, pp. 1281-1286.
- [8] N. Tudoroiu and M. Zaheeruddin, "Fault Detection and Diagnosis of Valve Actuators in Discharge Air Temperature (DAT) Systems, using Interactive Unscented Kalman Filter Estimation", in Proceedings of the IEEE International Symposium on Industrial Electronics, 2006, pp. 2665-2670.
- [9] K. Choi, S. M. Namburu, M. S. Azam, J. Luo, K. R. Pattipati, and A. Patterson-Hine, "Fault Diagnosis in HVAC chillers", IEEE Instrumentation & Measurement Magazine, 2005, pp. 24-32.
- [10] L. Hu, K. Cao, H. Xu and B. Li, "Fault Diagnosis of Hydraulic Actuator based on Least Squares Support Vector Machines", in Proceedings of the IEEE International Conference on Automation and Logistics, 2007, pp. 985-989.
- [11] J. Gao, W. Shi, J. Tan and F. Zhong, "Support vector machines based approach for fault diagnosis of valves in reciprocating pumps", in Proceedings of the IEEE Canadian Conference on Electrical and Computer Engineering, 2002, pp. 1622-1627.
- [12] F. He and W. Shi, "WPT-SVMs Based Approach for Fault Detection of Valves in Reciprocating Pumps", in Proceedings of the American Control Conference, 2002, pp. 4566-4570.
- [13] J. Middleton, P. Urwin and M. Al-Akaidi, "Fault Detection And Diagnosis In Gas Control Systems", Intelligent Measuring Systems for Control Applications, IEE Colloquium, 1995, pp. 8/1 - 8/3.
- [14] D. Linaric and V. Koroman, "Fault Diagnosis of a Hydraulic Actuator using Neural Network", in International Conference on Industrial Electronics (ICIT), 2003, pp. 106-111.

- [15] M. Karpenko, N. Sepehri and D. Scuse, "Neural Network Detection and Identification of Actuator Faults in a Pneumatic Process Control Valve", in Proceedings of IEEE International Symposium on Computational Intelligence in Robotics and Automation, 2001, pp. 166-171.
- [16] L. Hongmei, W. Shaoping and O. Pingchao, "Fault Diagnosis Based on Improved Elman Neural Network for a Hydraulic Servo System", in Proceedings of the Conference on Robotics, Automation and Mechatronics, 2006, pp. 1-6.
- [17] X. Z. Gao and S. J. Ovaska, "Motor Fault Detection Using Elman Neural Network with Genetic Algorithm-aided Training", in Proceedings of the International Conference on Systems, Man, and Cybernetics, 2000, Vol. 4, pp. 2386-2392.
- [18] B. Koppen-Seliger and P. M. Frank, "Fault Detection and Isolation in Technical Processes with Neural Networks", in Proceedings of the 34<sup>th</sup> Conference on Decision & Control, 1995, Vol. 3, pp. 2414-2419.
- [19] R. Sreedhar, B. Fernhndez and G. Y. Masada, "A Neural Network Based Adaptive Fault Detection Scheme", in Proceedings of the American control conference, 1995, Vol.5, pp. 3259-3263.
- [20] J. Tian, M. Gao, L. Cao and K. Li, "Fault Detection of Oil Pump Based on Fuzzy Neural Network", in Proceedings of the Third International Conference on Natural Computation (ICNC), 2007, pp. 636-640.
- [21] M. Y. Chow and P. V. Goode, "The Advantages and Challenges of Machine Fault Detection Using Artificial Neural Network and Fuzzy Logic Technologies", in Proceedings of the 36th Midwest Symposium on Circuits and Systems, 1993, Vol. 1, pp. 708-711.
- [22] A. Ramezanifar, A. Afshar and S. K. Y. Nikravesh, "Intelligent Fault Isolation of Control Valves in a Power Plant", in Proceedings of the First International Conference on Innovative Computing, Information and Control (ICICIC), 2006, pp.1-4.
- [23] M. Bartys, R. Patton, M. Syfert, S. de las Heras and J. Queveda, "Introduction to the DAMADICS actuator FDI benchmark study", Control Engineering Practice, 2006, Vol.14, pp. 577-596.
- [24] M. Bartys and M. Syfert, "Using DAMADICS Actuator Benchmark Library", 'DABLib' Simulink Library Help File, ver.1.22, 2002, pp.1-32.
- [25] M. Subotic, M. Tuba and N. Stanarevic, "Parallelization of the artificial bee colony (ABC) algorithm", In proceedings of the 11th WSEAS International Conference on Neural Networks, 2010, pp. 191-196.
- [26] S. M. Saab, N. K. T. El-Omari and H. H. Owaied, "Developing Optimization Algorithm using Artificial Bee Colony System", UBiCC Journal, 2009, Vol.4, No 5, pp. 391-396.
- [27] H. Peng, F. Long and C. Ding , "Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance and Min-Redundancy", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, Vol. 27, No.8, pp. 1226-1236.
- [28] R. Bose, "Information Theory, Coding and Cryptography", Second Ed., Tata McGraw-Hill, New Delhi, 2008.
- [29] C. Ding and H. Peng, "Minimum Redundancy Feature Selection From Microarray Gene Expression Data", Journal for Bioinformatics and Computational Biology, 2005, Vol.3, No.2, pp. 185-205.
- [30] C. A. Ratanamahatana and D. Gunopulos, "Scaling up the Naive Bayesian Classifier: Using Decision Trees for Feature Selection", in proceedings of Workshop on Data Cleaning and Preprocessing (DCAP 2002), IEEE International Conference on Data Mining (ICDM 2002) , 2002, pp. 475-487.

# A Survey on Data Mining and Pattern Recognition Techniques for Soil Data Mining

Dr. D. Ashok Kumar<sup>\*1</sup>, N. Kannathasan<sup>#2</sup>

<sup>\*1</sup>Government Arts College, Tiruchirapalli- 620 022, Tamil Nadu, India

<sup>#2</sup>Mahatma Gandhi Government Arts College, Mahe, P.O. New Mahe-673 311, U. T of Puducherry, India

**Abstract**— Data mining has emerged as one of the major research domain in the recent decades in order to extract implicit and useful knowledge. This knowledge can be comprehended by humans easily. Initially, this knowledge extraction was computed and evaluated manually using statistical techniques. Subsequently, semi-automated data mining techniques emerged because of the advancement in the technology. Such advancement was also in the form of storage which increases the demands of analysis. In such case, semi-automated techniques have become inefficient. Therefore, automated data mining techniques were introduced to synthesis knowledge efficiently. A survey of the available literature on data mining and pattern recognition for soil data mining is presented in this paper. Data mining in Agricultural soil datasets is a relatively novel research field. Efficient techniques can be developed and tailored for solving complex soil datasets using data mining.

**Keywords**— *Data Mining, Pattern Recognition, Soil Data Mining*

## I. INTRODUCTION

This Data mining software applications includes various methodologies that have been developed by both commercial and research centers. These techniques have been used for industrial, commercial and scientific purposes. For example, data mining has been used to analyze large datasets and establish useful classification and patterns in the datasets. Agricultural and biological research studies have used various techniques of data analysis including, natural trees, statistical machine learning and other analysis methods [16]. This paper outlines research which may establish if new data mining techniques will improve the effectiveness and accuracy of the Classification of large soil datasets. In particular, this research work aims to compare the performance of the data mining algorithms with soil limitations and soil conditions in respect of the following characteristics: Acidity, Alkalinity and sodicity, Salinity, Low cation exchange capacity, Phosphorus fixation, Cracking and swelling properties, Depth, Soil density and Nutrient content. The use of standard statistical analysis techniques is both time consuming and expensive. If alternative techniques can be found to improve this process, an improvement in the classification of soils may result.

In many developing countries, hunger is forcing people to cultivate land that is unsuitable for agriculture and which can only be converted to agricultural use through enormous efforts

and costs, such as those involved in the construction of terraces. Each country is known for its core competence. India's is agriculture. Yet, it only accounts for 17 per cent of the total Gross Domestic Product. With the pressure of urbanization, it is going to be a challenge to produce food for more people with less land and water.

Agriculture or farming forms the backbone of any country economy, since a large population lives in rural areas and is directly or indirectly dependent on agriculture for a living. Income from farming forms the main source for the farming community. The essential requirements for crop harvesting are water resources and capital to buy seeds, fertilizers, pesticides, labor etc. Most farmers raise the required capital by compromising on other necessary expenditures, and when it is still insufficient they resort to credit from sources like banks and private financial institutions. In such a situation, the repayment is dependent on the success of the crop. If the crop fails even once due to several factors, like bad weather pattern; soil type; improper, excessive, and untimely application of both fertilizers and pesticides; adulterated seeds and pesticides etc. then he is pushed into an acute crisis causing severe stress [58]. In addition, the plant growth depends on multiple factors such as soil type, crop type, and weather. Due to lack of plant growth information and expert advice, most of the farmers fail to get a good yield.

Most knowledge of soil in nature comes from soil survey efforts. Soil survey, or soil mapping, is the process of determining the soil types or other properties of the soil cover over a landscape, and mapping them for others to understand and use. Primary data for the soil survey are acquired by field sampling and supported by remote sensing.

The test dataset using for this research work collected from World Soil Information – ISRIC (International Soil Reference and Information Centre). Version 3.1 of the ISRIC-WISE database (WISE3-World Inventory of Soil Emission Potentials) was compiled from a wide range of soil profile data collected by many soil professionals world wide. All profiles have been harmonized with respect to the original Legend (1974) and Revised Legend (1988) of FAO-Unesco. Thereby the primary soil data and any secondary data derived from them can be linked using GIS to the spatial units of the soil map of the world as well as more recent Soil and Terrain (SOTER) databases through the soil legend code.

WISE3 is a relational database, compiled using MS-ACCESS. It can handle data on: (a) soil classification; (b) soil horizon data; (c) source of data; and methods used for determining analytical data. Profile data in WISE3 originate from over 260 different sources, both analogue and digital. Some 40% of the profiles were extracted from auxiliary datasets, including various Soil and Terrain (SOTER) databases and the FAO Soil Database (FAO-SDB), which, in turn, hold data collated from a wide range of sources.

WISE3 holds selected attribute data for 10,253 soil profiles, with some 47,800 horizons, from 149 countries. Individual profiles have been sampled, described, and analyzed according to methods and standards in use in the originating countries. There is no uniform set of properties for which all profiles have analytical data, generally because only selected measurements were planned during the original surveys. Methods used for laboratory determinations of specific soil properties vary between laboratories and over time. Some times, results for the same property cannot be compared directly. WISE3 will inevitably include gaps, being a compilation of legacy soil data derived from traditional soil survey. These can be of a taxonomic, geographic, and soil analytical nature. As a result, the amount of data available for modeling is some times much less than expected. Adroit use of the data, however, will permit a wide range of agricultural and environmental applications at a global and continental scale (1:500000 and broader) [44].

The analysis of these datasets with various data mining techniques may yield outcomes useful to researchers in future.

## II. MATERIALS AND METHODS

The rapid growth of interest in data mining is due to the (i) falling cost of large storage devices and increasing ease of collecting data over networks, (ii) development of robust and efficient machine learning algorithms to process this data, and (iii) falling cost of computational power, enabling use of computationally intensive methods for data analysis [37].

Data Mining (DM) represents a set of specific methods and algorithms aimed solely at extracting patterns from raw data [18]. The DM process has developed due to the immense volume of data that must be handled easier in areas such as: business, medical industry, astronomy, genetics or banking field. Also, the success and the extraordinary development of hardware technologies led to the big capacity of storage on hard-disks, fact that challenged the appearance of many problems in manipulating immense volumes of data. Of course the most important aspect here is the fast growth of the Internet.

The core of the DM process lies in applying methods and algorithms in order to discover and extract patterns from stored data but before this step data must be pre-processed. It is well known that simple use of DM algorithms does not produce good results. Thus, the overall process of finding useful knowledge in raw data involves the sequential adhibition of the following steps: developing an understanding of the application domain, creating a target dataset based on

an intelligent way of selecting data by focusing on a subset of variables or data samples, data cleaning and pre-processing, data reduction and projection, choosing the data mining task, choosing the data mining algorithm, the data mining step, interpreting mined patterns with possible return to any of the previous steps and consolidating discovered knowledge.

The DM contains many study areas such as machine-learning, pattern recognition in data, databases, statistics, artificial intelligence, data acquisition for expert systems and data visualization. The most important goal here is to extract patterns from data and to bring useful knowledge into an understandable form to the human observer. It is recommended that obtained information to be facile to interpret for the easiness of use. The entire process aims to obtain high-level data from low level data.

Data mining involves fitting models to or determining patterns from observed data. The fitted models play the role of inferred knowledge. Typically, a data mining algorithm constitutes some combination of the following three components.

- The model: The function of the model (e.g., classification, clustering) and its representational form (e.g. linear discriminants, neural networks). A model contains parameters that are to be determined from the data.
- The preference criterion: A basis for preference of one model or set of parameters over another, depending on the given data.
- The search algorithm: The specification of an algorithm for finding particular models and parameters, given the data, model(s), and a preference criterion.

A particular data mining algorithm is usually an instantiation of the model/preference/search components. The more common model functions in current data mining practice include:

1. Classification [41], [38], [42], [6], [39]: classifies a data item into one of several predefined categorical classes.
2. Regression [19], [12], [64], [45]: maps a data item to a real valued prediction variable.
3. Clustering [61], [50], [47], [52], [29], [31], [62], and [21]: maps a data item into one of several clusters, where clusters are natural groupings of data items based on similarity metrics or probability density models.
4. Rule generation [60], [35], [40], [43], [23], [55], [53], [67]: extracts classification rules from the data.
5. Discovering association rules [2], [63], [5], and [34]: describes association relationship among different attributes.
6. Summarization [32], [65], [25], [20]: provides a compact description for a subset of data.

7. Dependency modeling [22], [7]: describes significant dependencies among variables.
8. Sequence analysis [10], [33]: models sequential patterns, like time-series analysis. The goal is to model the states of the process generating the sequence or to extract and report deviation and trends over time.

Though, there are lots of techniques available in the data mining, few methodologies such as Artificial Neural Networks, K nearest neighbor, K means approach, are popular currently depends on the nature of the data.

**Artificial Neural Network:** Artificial Neural Networks (ANN) is systems inspired by the research on human brain (Hammerstrom, 1993). Artificial Neural Networks (ANN) networks in which each node represents a neuron and each link represents the way two neurons interact. Each neuron performs very simple tasks, while the network representing of the work of all its neurons is able to perform the more complex task. A neural network is an interconnected set of input/output units where each connection has a weight associated with it. The network learns by fine tuning the weights so as able to predict the call label of input samples during testing phase. Artificial neural network is a new techniques used in flood forecast. The advantage of ANN approach in modeling the rain fall and run off relationship over the conventional techniques flood forecast. Neural network has several advantages over conventional method in computing. Any problem having more time for getting solution, ANN is highly suitable states that the neural network method successfully predicts the pest attack incidences for one week in advance.

Pedotransfer functions (PTFs) provide an alternative by estimating soil parameters from more readily available soil data. The two common methods used to develop PTFs are multiple-linear regression method and ANN. Multiple linear regression and neural network model (feed-forward back propagation network) were employed to develop a pedotransfer function for predicting soil parameters using easily measurable characteristics of clay, sand, silt, SP, Bd and organic carbon[51].

Artificial Neural Networks have been successful in the classification of other soil properties, such as dry land salinity (Spencer *et al.* 2004). Due to their ability to solve complex or noisy problems, Artificial Neural Networks are considered to be a suitable tool for a difficult problem such as the estimation of organic carbon in soil.

**Support Vector Machines:** Support Vector Machines (SVM) is binary classifiers (Borges, 1998; Cortes and Vapnik, 1995). SVM is able to classify data samples in two disjoint classes. The basic idea behind is classifying the sample data into linearly separable. Support Vector Machines (SVMs) are a set of related supervised learning methods used for classification and regression. In simple words given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that

predicts whether a new example falls into one category or the other.

SVM is used to assess the spatiotemporal characteristics of the soil moisture products [4].

**Decision trees:** The decision tree is one of the popular classification algorithms in current use in Data Mining and Machine Learning. Decision tree is a new field of machine learning which is involving the algorithmic acquisition of structured knowledge in forms such as concepts, decision trees and discrimination nets or production rules. Application of data mining techniques on drought related data for drought risk management shows the success on Advanced Geospatial Decision Support System (GDSS). Leisa J Armstrong states that data mining approach is one of the approaches used for crop decision making.

Research has been conducted in Australia to estimate a range of soil properties, including organic carbon (Henderson *et al.* 2001). The nation-wide database had 11,483 soil points available to predict organic carbon in the soil. An enhanced decision trees tool (Cubist), catering for continuous outputs was used for this study. A correlation of up to 0.64 was obtained between the predicted and actual organic carbon levels.

**K nearest neighbor:** K nearest neighbor techniques is one of the classification techniques in data mining. It does not have any learning phase because it uses the training set every time a classification performed. Nearest Neighbor search (NN) also known as proximity search, similarity search or closest point search is an optimization problem for finding closest points in metric spaces.

K nearest neighbor is applied for simulating daily precipitation and other weather variables (Rajagopalan and Lall, 1999).

**Bayesian networks:** A Bayesian network is a graphical model that encodes probabilistic relationships among variables of interest. When used in conjunction with statistical techniques, the graphical model has several advantages for data analysis. One, because the model encodes dependencies among all variables, it readily handles situations where some data entries are missing. Two, a Bayesian network can be used to learn causal relationships and hence can be used to gain understanding about a problem domain and to predict the consequences of intervention. Three, because the model has both a causal and probabilistic semantics, it is an ideal representation for combining prior knowledge (which often comes in causal form) and data. Four, Bayesian statistical methods in conjunction with Bayesian networks offer an efficient and principled approach for avoiding the over fitting of data Development of a data mining application for agriculture based on Bayesian networks were studied by Huang *et al.* (2008). According to him, Bayesian network is a powerful tool for dealing uncertainties and widely used in agriculture datasets. He developed the model for agriculture application based on the Bayesian network learning method. The results indicate that Bayesian Networks are a feasible and efficient.



Bayesian approach improves hydrogeological site characterization even when using low-resolution resistivity surveys [52].

**K means approach:** K means method is one of the most used clustering techniques in the data mining. The idea behind the K means algorithms is very simple that certain partition of the data in K clusters, the centers of the cluster can be computed as the mean of the all sample belonging to a cluster. The center of the cluster can be considered as the representative of the cluster. The center is quite close to all samples in the cluster.

K Means approach was used to classify the soil and plants (*Camps-Valls et al., 2003*).

**Fuzzy logic:** Fuzzy logic is a form of multi valued logic derived from Fuzzy set theory to deal with reasoning that is approximate rather than accurate. In contrast with "crisp logic", where binary sets have binary logic, fuzzy logic variables may have a truth value that ranges between 0 and 1 and is not constrained to the two truth values of classic propositional logic [46]. Furthermore, when linguistic variables are used, these degrees may be managed by specific functions. Fuzzy logic emerged as a consequence of the 1965 proposal of Fuzzy set theory by Lotfi zadeh [1] [66]. Though fuzzy logic has been applied to many fields, from control theory to artificial intelligence, it still remains controversial among most statisticians, who prefer Bayesian logic, and some control engineers, who prefer traditional two-valued logic.

Fuzzy logic is used to the prediction of soil erosion in a large watershed (B.Mitra et al., ScienceDirect, Nov.1998).

**Genetic Algorithm:** The Genetic Algorithm (GA) is a search heuristic that mimics the process of natural evolution. This heuristic is routinely used to generate useful solutions to optimization and search problems. Genetic algorithms belong to the larger class of Evolutionary Algorithm (EA), which generates solutions to optimization problems using techniques inspired by natural evolution, such as inheritance, mutation, selection and crossover.

Soil liquefaction is a type of ground failure related to earthquakes. It takes place when the effective stress within soil reaches zero as a result of an increase in pore water pressure during earthquake vibration (Youd, 1992). Soil liquefaction can cause major damage to buildings, roads, bridges, dams and lifeline systems, like the earthquakes.

Genetic Algorithm approach is used for assessing the liquefaction potential of sandy soils (G. Sen et al. Nat. Hazards Earth Syst. Sci., 2010).

**Ant Colony Optimization:** The Ant Colony Optimization (ACO) algorithm is probabilistic technique for solving computational problems which can be reduced to finding good paths through graphs. This algorithm is a member of ant colony algorithms family, in swarm intelligence methods, and it constitutes some Meta heuristic optimizations. Initially proposed by Marco Dorigo in 1992 in his Ph.D. thesis [13] [17], the first algorithm was aiming to search for an optimal path in a graph, based on the behavior of ants seeking a path

between their colony and a source of food. The original idea has since diversified to solve a wider class of numerical problems, and as a result, several problems have emerged, drawing on various aspects of the behavior of ants.

Ant Colony Optimization is applied for estimating unsaturated soil hydraulic parameters (K.C.Abbaspour *et al., ELSEVIER, 2001*).

**Particle Swarm Optimization:** Particle Swarm Optimization (PSO) is a method for performing numerical optimization without explicit knowledge of the gradient of the problem to be optimized. PSO is originally attributed to Kennedy, Eberhart, and Shri [28] [54] and was first intended for simulating social behavior. The algorithm was simplified and it was observed to be performing optimization. The book by Kennedy and Eberhart [27] describes many philosophical aspects of PSO and swarm intelligence. An extensive survey of PSO applications is made by Poli [48] [49].

Particle Swarm Optimization is used for analysis of Soil erosion characteristics (Li Yunkai et al, Springer, Sep.2009).

**Simulated Annealing:** Simulated Annealing (SA) is a generic probabilistic Meta heuristic for the global optimization problem of applied mathematics, namely locating a good approximation to the global optimum of a given function in a large search space. It is often used when the search space is discrete (e.g., all tours that visit a given set of cities). For certain problems, simulated annealing may be more effective than exhaustive enumeration provided that the goal is merely to find an acceptably good solution in a fixed amount of time, rather than the best possible solution. The method was independently described by Scott Kirkpatrick, C. Daniel Gelatt and Mario P. Vecchi in 1983 [30] and by Vlado Cerny in 1985 [9]. The method is an adaptation of the Metropolis Hastings algorithm, a Monte Carlo method to generate sample states of a thermodynamic system, invented by N. Metropolis et al. in 1953 [36].

Simulated Annealing is used for analyzing Soil Properties (R.M. Lark et al., ScienceDirect, March, 2003).

### III. RESULTS AND DISCUSSION

The purpose of the study is to examine the most effective techniques to extract new knowledge and information from existing soil profile data contained within ISRIC-WISE soil data set. Several data mining techniques are in agriculture and allied area. Few of techniques are discussed here. K means method is used to forecast the pollution in the atmosphere (*Jorquera et al., 2001*). Different possible changes of weather are analyzed using SVM (*Tripathi et al., 2006*). K means approach is used for classifying soil in combination with GPS readings (*Verheyen et al., 2001*). Wine Fermentation process monitored using data mining techniques. Taste sensors are used to obtain data from the fermentation process to be classified using ANNs (*Riul et al., 2004*).

A brief survey of the related work in the area of soil mining is that the data involved here are high dimensional data and

dimensionality reduction was addressed in classical methods such as Principal Component Analysis (PCA) [24]. There is a growing literature demonstrating the predictive capacity of the soil landscape paradigm using digital data and empirical numerical modeling techniques as specified by Christopher et al., [11]. The Eigen decomposition of empirical covariance matrix is performed and the data points are linearly projected. When the information relevant for classification is present in eigenvectors associated with small eigenvalues are removed, then this could lead to degradation in classification accuracy. Examples of spatial prediction have been provided, across a range of physiographical range of environment and spatial extents, for a number of soil properties by Gessler et al., [21] Tenenbaum et al., [59] introduced the concept of Isomap, a global dimensionality reduction algorithm. The CDDR (classification constrained dimensionality reduction) algorithm [15] was only demonstrated for two classes and the performance was analyzed for simulated data. Bui et al., [8] demonstrated the potential for the discovery of knowledge embedded in survey of landscape model using rule induction techniques based on decision trees. It has the ability to mimic soil map using samples taken from it, and by implication it also captures the embedded knowledge. Related to agriculture, many countries are still facing a multitude of problems to maximize productivity [26]. Another concept of CDDR plots the classification error probability and its confidence interval using K nearest neighbour classifier [14]. Normally there is a decrease in error probability as dimension increases, and the optimal value is reached when dimension value varies between 12 - 14, which has been proved using entropic graph algorithm. However the food production has improved significantly during last two decades by providing it with good seeds, fertilizers, and pesticides and modern farming equipment [57]. The agriculture sector has seen a tremendous improvement.

#### IV. CONCLUSIONS

In this research survey, data mining and pattern recognition techniques for soil data mining studied. The survey aims to come out of the techniques being used in the agricultural soil science and its allied area.

The recommendations arising from this research survey are: A comparison of different data mining techniques could produce an efficient algorithm for soil classification for multiple classes. The benefits of a greater understanding of soils could improve productivity in farming, maintain biodiversity, reduce reliance on fertilizers and create a better integrated soil management system for both the private and public sectors.

#### ACKNOWLEDGMENT

The authors would like to thank the editor and the anonymous reviewers for their valuable comments and suggestions.

#### REFERENCES

- [1] "Fuzzy Logic". Stanford Encyclopedia of Philosophy. Stanford University. 2006-07-23. Retrieved 2008-09-29.
- [2] Agrawal R., Imielinski T., and Swami A., Mining association Rules between sets of items in large databases, in Proceedings of 1993 ACM SIGMOD International Conference on Management of Data, (Washington D.C.), pp. 207-216, May 1993.
- [3] Alahakoon D., Halgamuge S.K., and Srinivasan B, Dynamic self organizing maps with controlled growth for knowledge discovery, IEEE Transactions on Neural Networks, vol. 11, pp. 601-614, 2000.
- [4] Anish C. Turlapaty, Valentine Anantharaj, Nicolas H. Younan, Spatio-temporal consistency analysis of AMSR-E soil moisture data using wavelet-based feature extraction and one-class SVM, In the Proceedings of the Annual Conference Baltimore, Maryland, March 9-13, 2009.
- [5] Au W. H. and Chan K. C. C., An effective algorithm for discovering fuzzy rules in relational databases, in Proceedings of IEEE International Conference on Fuzzy Systems FUZZ IEEE 98, (Alaska), pp. 1314-1319, May 1998.
- [6] Banerjee M, Mitra S, and Pal S.K, Rough fuzzy MLP: Knowledge encoding and classification, IEEE Transactions on Neural Networks, vol. 9, pp. 1203-1216, 1998.
- [7] Bosc P., Pivert O., and Ughetto L., Database mining for the discovery of extended functional dependencies, in Proceedings of NAFIPS 99, (New York, USA), pp. 580-584, June 1999.
- [8] Bui E. N., Loughhead A. and Comer R., Extracting Soil Landscape Rules from Previous Soil Surveys. Australian Journal of Soil Science, 37:495508, 1999.
- [9] Cerny V., A thermo dynamical approach to the traveling salesman problem: an efficient simulation algorithm. Journal of Optimization Theory and Applications, 45:41-51, 1985.
- [10] Chiang D. A., Chow L. R., and Wang Y. F., Mining time series data by a fuzzy linguistic summary system," Fuzzy Sets and Systems, vol. 112, pp. 419-432, 2000.
- [11] Christopher J. Moran and Elisabeth Bui N., Spatial Data Mining for Enhanced Soil Map Modeling. In the Proceedings of the International Journal of Geographical Information Science, 2002.
- [12] Ciesielski V and Palstra G, Using a hybrid neural/expert system for database mining in market survey data, in Proc. Second International Conference on Knowledge Discovery and Data Mining (KDD-96), (Portland, OR), p. 38, AAAI Press, Aug. 2-4, 1996.
- [13] Colomi A., Dorigo et M., Maniezzo V., Distributed Optimization by Ant Colonies, actes de la première conference euro penne sur la vie artificielle, Paris, France, Elsevier Publishing, 134-142, 1991.
- [14] Costa A. and Hero A. O. Geodesic Entropic Graphs for Dimension and Entropy Estimation in Manifold Learning. In the Proceedings of IEEE Transaction Signal Processing, volume 52, pages 2210-2221, 2004.
- [15] Costa J. A. and Hero A. O., III. Classification Constrained Dimensionality Reduction. In IEEE International Conference on Acoustic Speech, and Signal Processing, volume 5, pages 1077-1080, March 2005.
- [16] Cunningham S. J and Holmes G. Developing innovative applications in agriculture using data mining, In the Proceedings of the Southeast Asia regional Computer Confederation Conference, 1999.
- [17] Dorigo M., Optimization, Learning and Natural Algorithms, PhD thesis, Politecnico di Milano, Italie, 1992.
- [18] Fayadd, U., Piatetsky-Shapiro, G., and Smyth, P, Data Mining to Knowledge Discovery in Databases, AAAI Press / the MIT Press, Massachusetts Institute of Technology. ISBN 0-262-56097-6 Fayap, 1996.
- [19] Fayyad U.M, Piatetsky-Shapiro G, Smyth P., and Uthurusamy R., eds., Advances in Knowledge Discovery and Data Mining. Menlo Park, CA: AAAI/MIT Press, 1996.
- [20] George R. and Srikanth R., Data summarization using genetic algorithms and fuzzy logic, in Genetic Algorithms and Soft Computing (F. Herrera and J. L. Verdegay, eds.), pp. 599-611, Heidelberg: Springer-Verlag, 1996.
- [21] Gessler P. E., Moore D., McKenzie N. J. and Ryan P... Soil Landscape Modeling and Spatial Prediction of Soil Attributes. In the Proceedings

- of the International Journal of Geographical Information Systems, volume 9, pages 421-432, 1995.
- [22] Hale J. and Shenoj S., Analyzing FD inference in relational databases, *Data and Knowledge Engineering*, vol. 18, pp. 167-183, 1996.
- [23] Hu X. and Cercone N., Mining knowledge rules from databases: A rough set approach, in *Proceedings of the 12<sup>th</sup> International Conference on Data Engineering*, (Washington), pp. 96-105, IEEE Computer Society, Feb. 1996.
- [24] Jain A. K. and Dubes R. C., *Algorithm for Clustering Data*. Prentice Hall, 1998.
- [25] Kacprzyk J. and Zadrozny S., Data mining via linguistic summaries of data: an interactive approach, in *Proceedings of IIZUKA 98*, (Fukuoka, Japan), pp. 668-671, October 1998.
- [26] Katyal J. C., Paroda R. S., Reddy M. N., Aupam Varma and N. Hanumanta Rao. *Agricultural Scientists Perception on Indian Agriculture: Scene Scenario and Vision*. National Academy of Agricultural Science, 2000.
- [27] Kennedy J. Eberhart R.C. *Swarm Intelligence*. Morgan Kaufmann. ISBN 1-55860-595-9., 2001
- [28] Kennedy, J., Eberhart, R., "Particle Swarm Optimization". *Proceedings of IEEE International Conference on Neural Networks. IV*. pp. 1942-1948, 1995.
- [29] Kiem H. and Phuc D., Using rough genetic and Kohonen's Neural network for conceptual cluster discovery in data mining, in *Proceedings of RSFDGrC'99*, (Yamaguchi, Japan), pp. 448-452, November 1999.
- [30] Kirkpatrick S., Gelatt C.D, Vecchi M.P. Optimization by Simulated Annealing. *Science New Series* 220 (4598):671-680. Doi:10.1126/science.220.4598.671. ISSN 00368075. , 1983-05-13
- [31] Kohonen, Kaski S., Lagus K., Salojarvi J., Honkela J., Paatero V., and Saarela A., Self organization of a massive document collection, *IEEE Transactions on Neural Networks*, vol. 11, pp. 574-585, 2000.
- [32] Lee D. H. and Kim M. H, Database summarization using fuzzy ISA hierarchies, *IEEE Transactions on Systems Man and Cybernetics. Part B-Cybernetics*, vol. 27, pp. 68-78, 1997.
- [33] Lee R. S. T. and Liu J. N. K., Tropical cyclone identification and tracking system using integrated neural oscillatory leastic graph matching and hybrid RBF network track mining techniques, *IEEE Transactions on Neural Networks*, vol. 11, pp. 680-689, 2000.
- [34] Lopes C., Pacheco M., Vellasco M., and Passos E., Rule evolver: An evolutionary approach for data mining, in *Proceedings of RSFDGrC'99*, (Yamaguchi, Japan), pp. 458-462, November 1999.
- [35] Lu H.J., Setiono R., and Liu H., Effective data mining using neural networks, *IEEE Transactions on Knowledge and Data Engineering*, vol. 8, pp. 957-961, 1996.
- [36] Metropolis N., Rosenbluth A.W., Rosenbluth M.N., Teller A.H. and Teller E... Equations of State Calculations by Fast Computing Machines. *Journal of Chemical Physics*, 21(6):1087-1092, 1953.
- [37] Mitchell T.M, Machine learning and data mining, *Communications of the ACM*, vol. 42, no. 11, 1999.
- [38] Mitra S and Pal S.K, Fuzzy self organization, inferencing and rule generation, *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 26, pp. 608-620, 1996.
- [39] Mitra S, Mitra P, and Pal S.K, Evolutionary modular Design of rough knowledge-based network using fuzzy attributes Neuro computing, vol. 36, pp. 45-66, 2001.
- [40] Mitra S. and Hayashi Y., Neuro-fuzzy rule generation: Survey in soft computing framework, *IEEE Transactions on Neural Networks*, vol. 11, pp. 748-768, 2000.
- [41] Mitra S. and Pal S.K, Fuzzy multi-layer perceptron, Inferencing and rule generation, *IEEE Transactions on Neural Networks*, vol. 6, pp. 51-63, 1995.
- [42] Mitra S., De R.K, and Pal S.K, Knowledge-based fuzzy MLP for classification and rule generation, *IEEE Transactions on Neural Networks*, vol. 8, pp. 1338-1350, 1997.
- [43] Mollestad T. and Skowron A., A rough set framework for data mining of propositional default rules," *Lecture Notes in Computer Science*, vol. 1079, pp. 448-457, 1996.
- [44] Niels H. Batjes, *ISRIC-WISE Harmonized Global Soil Profile Dataset (Ver. 3.1) - A Report -2008/2*
- [45] Noda E, Freitas A.A, and Lopes H.S, Discovering Interesting prediction rules with a genetic algorithm, in *Proceedings of IEEE Congress on Evolutionary Computation CEC 99*, (Washington DC), pp. 1322-1329, July 1999.
- [46] Novak, V., Perfilieva, I. and Mockor, J. *Mathematical principles of fuzzy logic* Dodrecht: Kluwer Academic. ISBN 0-7923-8595-0, 1999.
- [47] Pedrycz W, Conditional fuzzy c-means, *Pattern Recognition Letters*, vol. 17, pp. 625-632, 1996.
- [48] Poli, R. An analysis of publications on Particle swarm optimization applications. Technical Report CSM-469 (Department of Computer Science, University of Essex, UK), 2007
- [49] Poli, R. Analysis of the publications on the Applications of particle swarm optimization". *Journal of Artificial Evolution and Applications: 1-10*. Doi:10.1155/2008/685175., 2008
- [50] Russell S and Lodwick W, Fuzzy clustering in data mining for telco database marketing campaigns, in *Proceedings of NAFIPS 99*, (New York), pp. 720-726, June 1999.
- [51] Sarmadian F., Taghizadeh R., Mehrjardi and. Akbarzadeh A, Optimization of Pedotransfer Functions Using an Artificial Neural Network, *Australian Journal of Basic and Applied Sciences*, 3(1): 323-329, ISSN 1991-8178., 2009,
- [52] Shalvi D and De Claris N, Unsupervised neural network approach to medical data mining techniques, in *Proceedings of IEEE International Joint Conference on Neural Networks*, (Alaska), pp. 171-176, May 1998.
- [53] Shan N. and Ziarko W., Data-based acquisition and incremental modification of classification rules, *Computational Intelligence*, vol. 11, pp. 357-370, 1995.
- [54] Shi, Y. Eberhart, R.C., A modified particle swarm optimizer". *Proceedings of IEEE International Conference on Evolutionary Computation*. pp. 69-73., 1998
- [55] Skowron A., Extracting laws from decision tables - a rough set approach, *Computational Intelligence*, vol. 11, pp. 371-388, 1995.
- [56] Souheil Ezzedine, Yoram Rubin, and Jinsong Chen, Bayesian method for hydro geological site characterization using borehole and geophysical survey data: Theory and application to the Lawrence Livermore National Laboratory Superfund site, *Water Resources Research*, vol. 35, No. 9, Pages 2671-2683, September, 1999.
- [57] Subba Rao. *Indian Agriculture past Laurels and Future Challenges*, *Indian Agriculture: Current Status, Prospects and Challenges*. Convention of Indian Agricultural Universities Association, 27:58-77, December 2002.
- [58] Sudarshan Reddy S, Vedantha S, Venkateshwar Rao B, Sundar Ram Reddy and Venkat Reddy. *Gathering Agrarian Crisis Farmers Suicides in Warangal district. Citizens Report*, 1998.
- [59] Tenenbaum J. B., De Silva and Langford C., A Global Geometric Framework for Dimensionality Reduction. 290(5500):2319-2323, 2000.
- [60] Tickle A.B, Andrews R., Golea M., and Diederich J., The truth will come to light: Directions and challenges in extracting the knowledge embedded within trained artificial neural networks, *IEEE Transactions on Neural Networks*, vol. 9, pp. 1057-1068, 1998.
- [61] Turksen I.B, Fuzzy data mining and expert system Development, in *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics*, (San Diego, CA), pp. 2057-2061, October 1998.
- [62] Vesanto J. and Alhoniemi E., Clustering of the self organizing map, *IEEE Transactions on Neural Networks*, vol. 11, pp. 586-600, 2000.
- [63] Wei Q. and Chen G., Mining generalized association rules with fuzzy taxonomic structures, in *Proceedings of NAFIPS 99*, (New York), pp. 477-481, June 1999.
- [64] Xu K, Wang Z, and Leung K.S, Using a new type of non Linear integral for multi-regression: an application of evolutionary Algorithms in data mining, in *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics*, (San Diego, CA), pp. 2326-2331, October 1998.
- [65] Yager R. R., On linguistic summaries of data, in *Knowledge Discovery in Databases (W. Frawley and G. Piatetsky-Shapiro, eds.)*, pp. 347-363, Menlo Park, CA: AAAI/MIT Press, 1991.
- [66] Zadeh L.A... "Fuzzy sets", *Information and Control* 8 (3): 338-353, 1965

- [67] Zhang Y.Q., Fraser M.D., Gagliano R.A., and Kandel A., Granular neural networks for numerical-linguistic data fusion and knowledge discovery, IEEE Transactions on Neural Networks, vol.11, pp. 658-667,2000.

#### AUTHORS BIOGRAPHY



D. Ashok kumar did his Master degree in Mathematics and Computer Applications in 1995 and completed Ph.D., on Intelligent Partitional Clustering Algorithm's in 2008, from Gandhigram Rural Institute–Deemed University, Gandhigram, Tamil Nadu, INDIA. He is currently working as Senior Grade Assistant Professor and Head in the Department of Computer Science, Government Arts College, Tiruchirapalli–620 022, Tamil Nadu, INDIA. His research interest includes Pattern Recognition and Data Mining by various soft computing approaches viz., Neural Networks, Genetic Algorithms, Fuzzy Logic, Rough set, etc.,



N. Kannathasan is a Senior Grade Assistant Professor of Computer Science at the Mahatma Gandhi Government Arts College, Mahe, U.T. of Puducherry, INDIA. Prior to joining MGGAC, Mahe, he served at the Bharathidasan Government College for Women, Puducherry, Velammal College of Management and Computer Studies, Chennai, SRM Arts and Science College, Chennai, and AVC College, Mayiladuthurai. He received his M.Phil. Computer Science from the Bharathidasan University, Tiruchirappalli and M.Sc. Computer Science from Ayya Nadar Janaki Ammal College, Sivakasi.

# Markov Model for Reliable Packet Delivery in Wireless Sensor Networks

Vijay Kumar(Member, IEEE)<sup>1</sup>, R. B. Patel (Member, IEEE)<sup>2</sup>, Manpreet Singh (Member, IEEE)<sup>3</sup> and Rohit Vaid<sup>4</sup>

<sup>1</sup> Department of Computer Engineering, M. M. University,  
Mullana (Ambala) 133207, India

<sup>2</sup> Department of Computer Engineering, M. M. University,  
Mullana (Ambala) 133207, India

<sup>3</sup> Department of Computer Engineering, M. M. University,  
Mullana (Ambala) 133207, India

<sup>4</sup> Department of Computer Engineering, M. M. University,  
Mullana (Ambala) 133207, India

## Abstract

This paper presents a model for reliable packet delivery in Wireless Sensor Networks based on Discrete Parameter Markov Chain with absorbing state. We have demonstrated the comparison between cooperative and non cooperative automatic repeat request (ARQ) techniques with the suitable examples in terms of reliability and delay in packet transmission.

**Keywords:** Reliability, Absorbing State, Wireless Sensor Network, Markov chain.

## 1. Introduction

Wireless sensor networks (WSNs) [1][2] are the topic of intense academic and industrial studies. Research is mainly focused on energy saving schemes to increase the lifetime of these networks [4][5]. There is an exciting new wave in sensor applications-wireless sensor networking-which enables sensors and actuators to be deployed independent of costs and physical constraints of wiring. For a wireless sensor network to deliver real world benefits, it must support the following requirements in deployment: scalability, reliability, responsiveness, power efficiency and mobility.

The complex inter-relationships between these characteristics are a balance; if they are not managed properly, the network can suffer from overhead that negates its applicability. In order to ensure that the network supports the application's requirements, it is important to understand how each of these characteristics affects the reliability.

### 1.1. Scalability and Reliability

Network reliability and scalability are closely coupled and typically they act against each other. In other words, it is very difficult to build a reliable ad hoc network as the number of nodes increases [7]. This is due to network overhead that comes with increased size of network. In ad hoc network, there is no predefined topology or shape. Therefore, any node wishing to communicate with other nodes should generate more control packets than data packets. Moreover, as network size increases, there is more risk that communication links get broken, which will end up with creating more control packets. In summary, more overhead is unavoidable in a larger scale wireless sensor network to keep the communication path intact.

### 1.2. Reliability and power efficiency

Power efficiency also plays a very important role in this complex equation. To design a low power wireless sensor network, the duty cycle of each node needs to be reduced. The drawback is that as the node stays longer in sleep mode [3] to save the power, there is less probability that the node can communicate with its neighbors and may also lower the reliability due to lack of exchange of control packets and delays in the packet delivery.

### 1.3. Reliability and responsiveness

Ability of the network to adapt quickly the changes in the topology is known as responsiveness. For better responsiveness, there should be more issue and exchange of control packets in ad hoc network, which will naturally result in less reliability.



### 1.4. Mobility and reliability

A wireless sensor network that includes a number of mobile nodes should have high responsiveness to deal with the mobility. The mobility effect on responsiveness will compound the reliability challenge.

Many applications for wireless sensor networks require immediate and guaranteed action; for example medical emergency alarm, fire alarm detection, instruction detection [6]. In these situations packets has to be transported in a reliable way and in time through the sensor network. Thus, besides the energy consumption, delay and data reliability becomes very relevant for the proper functioning of the network.

Direct communication between any node and sink could be subject only to just a small delay, if the distance between the source and the destination is short, but it suffers an important energy wasting when the distance increases. Therefore often multihop short range communications through other sensor nodes, acting as intermediate relay, are preferred in order to reduce the energy consumption in the network. In such a scenario it is necessary to define efficient technique that can ensure reliable communication with very tight delay constraint. In this work we focus attention on the control of data and reliability in multihop scenario.

A simple implementation of ARQ is represented by the Stop and Wait technique that consists in waiting the acknowledgement of each transmitted packet before transmitting the next one, and retransmit the same packet in case it is lost or wrongly, received by destination [8].

We extend here this analysis by introducing the investigation of the delay required by the reliable data delivery task. To this aim we investigate the delay required by a cooperative ARQ mechanism to correctly deliver a packet through a multihop linear path from a source node to the sink. In particular we analyze the delay and the coverage range of the nodes in the path, therefore the relation between delay and the number of cooperative relays included in the forwarding process.

## 2. System Model

Fig. 1 shows the network structure with linear multihop path consist of source node (node  $n = 1$ ), destination (node  $n = N$ ) and  $(N-2)*t$  intermediate relay nodes deployed at equal distance where  $t$  is the number of parallel path of intermediate relay nodes between source and destination. Each path is composed by  $Z = N - 1$  links. Suppose that all the nodes have circular radio coverage with the same transmission range  $R_t$ . When a sensor transmits a packet, it is received by all the sensors in a listen state inside the coverage area of the sender.

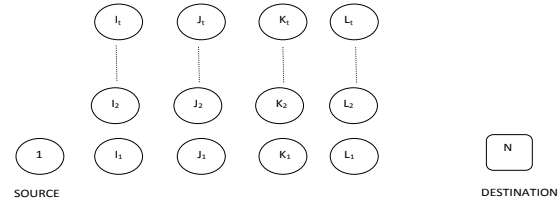


Fig.1 Network structure with Linear Multi-hop path

When a packet is transmitted, it can be forwarded towards the destination by only those nodes which are closer to the destination, then the transmitter.

### 2.1 Discrete Parameter Markov Chain with Absorbing State

Packet transfer from source to destination via intermediate forwarders can be treated as a state diagram of discrete parameter Markov chain with absorbing state. An absorbing state is a state from which there is zero probability of exiting. An absorbing Markov system is a Markov system that contains at least one absorbing state, and is such that it is possible to get from each non absorbing state to some absorbing state in one or more time steps. Consider  $p$  be the probability of successful transmission of a packet to an intermediate relay node inside the coverage range. Therefore  $1-p$  will be the probability of unsuccessful transmission of packet.

For each; node  $n$ , the probability to correctly deliver a packet to a node that is  $R_t$  links distant is equal to  $p$ . So the probability that the packet is not correctly received by this node  $(1 - p)$ , while it is correctly received from the immediately previous node with a probability  $p$ ; so with a probability  $(1 - p) p$  the packet will be forwarded by the previous node. If also this node has not correctly received the packet send by node  $n$ , event that occur with a probability  $(1 - p)^2$ , with a probability  $(1 - p)^2 p$  the packet will be forwarded by the node previous to previous. If none of the node in the coverage area of the transmitter receives a correct packet it is necessary to ask the retransmission of the packet by the source node. It is possible to describe the process concerning one data packet forwarding from the source node  $n = 1$  to the destination  $n = N$  with a discrete time Markov chain with absorbing state. Packet transmitted by a node will be further forwarded by a node in the coverage range of the transmitter which is furthest node from the source and has correctly received the packet.

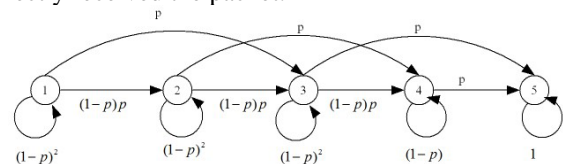


Fig 2 Packet transmission in Cooperative ARQ as a discrete parameter Markov Chain with absorbing state

Consider a single multihop linear path consisting five sensors with four links as shown in fig. 2. Assume transmission range of each sensor node is  $R_t=2$  unit. State transition probability matrix for a successful transmission of a packet under cooperative automatic repeat request will be as under:

$$P_{Success} = \begin{bmatrix} (1-p)^2 & p(1-p) & p & 0 & 0 \\ 0 & (1-p)^2 & p(1-p) & p & 0 \\ 0 & 0 & (1-p)^2 & p(1-p) & p \\ 0 & 0 & 0 & (1-p) & p \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Similarly we can find the probability matrix for link error by replacing  $(1-p)$  with  $q$ . In fig. 2 states 1 through 4 are transient state while state 5 is an absorbing state.

In general, we consider a Markov chain with  $n$  states,  $s_1, s_2, \dots, s_n$ .  $S_n$  will be the absorbing state, and the remaining state will be transient. The transition probability matrix of such a chain may be partitioned so that

$$P = \begin{bmatrix} Q & C \\ 0 & 1 \end{bmatrix}$$

Where  $Q$  is an  $(n-1)$  by  $(n-1)$  substochastic matrix, describing the probabilities of transition only among the transient states.  $C$  is a column vector and  $0$  is a row vector of  $(n-1)$  zeros. Now the  $k$ -step transition probability matrix  $P^k$  has the form

$$P^k = \begin{bmatrix} Q^k & C' \\ 0 & 1 \end{bmatrix}$$

Where  $C'$  is a column vector whose elements will be of no further use and hence need not be computed. The  $(i, j)$  entry of matrix  $Q^k$  denotes the probability of arriving in (transient) state  $s_j$  after exactly  $k$  steps starting from

(transient) state  $s_i$ . It can be shown that  $\sum_{k=0}^{\infty} Q^k$  converges

as  $t$  approaches infinity. This imply that the inverse matrix  $(I-Q)^{-1}$ , called the fundamental matrix,  $M$ , exists and is given by

$$M = (I - Q)^{-1} = I + Q + Q^2 + \dots = \sum_{k=0}^{\infty} Q^k$$

The fundamental matrix is used for calculating the expected no. of steps to absorption. The number of times, starting in state  $i$ , and expected to visit state  $j$  before absorption is the  $ij^{th}$  entry of  $M$ . The total no. of steps expected before absorption equals the total no. of visits expected to make to all the non absorption states. This is the sum of all the entries in the  $i^{th}$  row of  $M$ .

Suppose  $p=0.8$ , then  $Q$  will be as under

$$Q_{Success} = \begin{bmatrix} .04 & .16 & .8 & 0 \\ 0 & .04 & .16 & .8 \\ 0 & 0 & .04 & .16 \\ 0 & 0 & 0 & .2 \end{bmatrix}$$

Therefore fundamental matrix  $M = (I - Q)^{-1}$

$$M = \begin{bmatrix} 25/24 & 25/144 & 775/864 & 305/864 \\ 0 & 25/24 & 25/144 & 155/144 \\ 0 & 0 & 25/24 & 5/24 \\ 0 & 0 & 0 & 5/4 \end{bmatrix}$$

Thus the states 1, 2, 3 and 4 are respectively executed 25/24, 25/144, 775/864, 305/864 times on the average. If  $t_1, t_2, t_3$  and  $t_4$  respectively is the time for one time execution of the states 1,2,3 and 4 then total time required to transmit a packet from source node 1 to destination node 5 is equal to :

$$T = 25/24 t_1 + 25/144 t_2 + 775/864 t_3 + 305/864 t_4 \text{ unit times.}$$

If  $t_1=t_2=t_3=t_4=t$  then  $T=2.4645$  unit times.

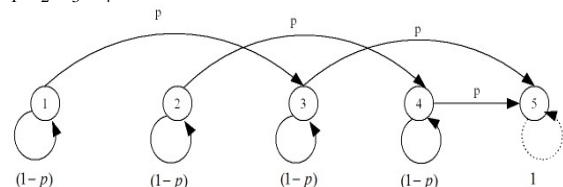


Fig 3 Packet transmission in Non-Cooperative ARQ as a discrete parameter Markov Chain with absorbing state

In non-Cooperative ARQ, a packet transmitted by source node is received by a node at distance  $R_t$  towards the destination from source and is forwarded by the node if packet received correctly otherwise transmitter is forced for retransmission. Other intermediate nodes between the transmitter and the node at distance  $R_t$  remains in sleep mode as they will never be involved in packet forwarding process. State transition probability matrix for successful transmission of the packet for non-cooperative ARQ will be as under:

$$P_{Success} = \begin{bmatrix} 1-p & 0 & p & 0 & 0 \\ 0 & 1-p & 0 & p & 0 \\ 0 & 0 & 1-p & 0 & p \\ 0 & 0 & 0 & (1-p) & p \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Suppose  $p=0.8$ , then  $Q$  will be as under

$$Q_{Success} = \begin{bmatrix} .2 & 0 & .8 & 0 \\ 0 & .2 & 0 & .8 \\ 0 & 0 & .2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Therefore fundamental matrix

$$M = (I - Q)^{-1}$$

$$= \begin{bmatrix} 5/4 & 0 & 5/4 & 0 \\ 0 & 5/4 & 0 & 1 \\ 0 & 0 & 5/4 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Thus the states 1, 2, 3 and 4 are respectively executed 5/4, 0, 5/4, 0 times on the average if source node is considered as node 1. If  $t_1, t_2, t_3$  and  $t_4$  respectively is the time for one time execution of the states 1, 2, 3 and 4 then total time required to transmit a packet from source node 1 to destination node 5 is equal to :

$$T = 5/4 t_1 + 5/4 t_3 \text{ units times.}$$

If  $t_1=t_2=t_3=t_4=t$  then  $T=2.5$  unit times.

### CONCLUSION

In this work we have presented Markov model to analyze the performance of cooperative and non cooperative ARQ in terms of delay and power efficiency. It has been observed that packet delivery is more reliable and timely in case of cooperative ARQ, where as non cooperative ARQ is better in terms of power efficiency of sensor nodes as most of the sensors do not participate in packet forwarding process.

### REFERENCES

- [1] F. Akyildiz, W. Su, Y. S. Subramaniam, and E. Cayirci, "A Survey on sensor networks", IEEE Communication Magazine, Volume: 40 Issue:8, pp: 102-114, August 2002.
- [2] Tubaishat M & Madria S, "Sensor networks: an overview" IEEE Potentials Volume 22, Issue 2, April- May 2003 pp: 20-23, 2003.
- [3] C. F. Chiasserini and M. Garetto, "An analytical model for wireless sensor networks with sleeping nodes", IEEE Trans. Mobile Computing, vol. 5, no. 12, pp: 1706-1718, 2006.
- [4] Pal, Y., Awasthi, L.K., Singh, A.J., "Maximize the Lifetime of Object Tracking Sensor Network with Node-to-Node Activation Scheme", in Proceeding of Advance Computing Conference, pp: 1200 – 1205, 2009.
- [5] Yan-liang Jin, Hao-jie Lin, Zhu-ming Zhang, Zhen Zhang, Xu-yuan Zhang, "Estimating the Reliability and Lifetime of Wireless Sensor Network", in Proceeding of Wireless Communications, Networking and Mobile Computing (WiCOM 2008), pp: 2181 – 2186, 2008.
- [6] L. Bernardo, R. Oliveria, R. Tiago, P. Pinto, "A Fire Monitoring Application for Scattered Wireless Sensor Network", in the proceeding of WinSys 2007, on July 28-31, 2007.
- [7] Wenyu Cai, Xinyu Jin, Yu Zhang, Kangsheng Chen, Jun Tang, "Research on Reliability Model of Large-Scale Wireless Sensor Networks", in Proceeding of Wireless Communications, Networking and Mobile Computing (WiCOM 2006) pp: 1-4, 2006.
- [8] AboElFotouh, H.M.F., Iyengar, S.S., Chakrabarty, K, "Computing reliability and message delay for Cooperative wireless distributed sensor networks subject to random

failures", in IEEE Transactions on Reliability, pp:145 – 155, 2005.

### AUTHOR BIOGRAPHY



**Prof. Vijay Kumar** born in Kanpur, India, on 30<sup>th</sup> June 1972. He received his B.E & M.E. degrees from Kumaon University Nainital (U.P) and Thapar University Patiala (Punjab) respectively. He has supervised 8 M. Tech and 1 M. Phil candidates. His research interests are in Wireless Sensor Networks, Reliability Theory and Artificial Neural Networks, etc. He has about 16 years experience in teaching. He is also a member of IEEE.



**Dr. R. B. Patel** received PhD from IIT Roorkee in Computer Science & Engineering, PDF from Highest Institute of Education, Science & Technology (HIEST), Athens, Greece, MS (Software Systems) from BITS Pilani and B. E. in Computer Engineering from M. M. M. Engineering College, Gorakhpur, UP. Dr. Patel is in teaching and Research & Development since 1991. He has supervised 30 M. Tech, 7 M. Phil and 2 PhD Thesis. He is currently supervising 3 M. Tech, and 8 PhD students. He has published more than 120 research papers in International/National Journals and Refereed International Conferences. He had been awarded for Best Research paper many times in India and abroad. He has written numbers books for engineering courses (These are "Fundamentals of Computing and Programming in C", "Theory of Automata and Formal Languages", "Expert Data Structures with C," "Expert Data Structures with C++," "Art and Craft of C" and "Go Through C". His research interests are in Mobile & Distributed Computing, Mobile Agent Security and Fault Tolerance, development infrastructure for mobile & Peer-To-Peer computing, Device and Computation Management, Cluster Computing, Sensor Networks, etc.



**Dr. Manpreet Singh** is working as Professor. & Head of computer science and Engineering department at MMEC, M. M. University Mullana, Ambala, India. He obtained his Ph.D. (Computer Science) from Kurukshetra University. He has number of publications in International journals/Conferences to his credit. His current research interest includes Grid Computing, Wireless communications, MANETs etc.



**Rohit Vaid** received his M. Tech. degree from Maharishi Markandeshwar University Mullana, Ambala (Haryana) respectively. His research interests are in Mobile & Distributed Computing, Mobile Agent Security and Fault Tolerance, Cluster Computing, Wireless Sensor Networks, etc.

# Comparative Study of VoIP over WiMax and WiFi

M. Atif Qureshi<sup>\*1</sup>, Arjumand Younus<sup>\*2</sup>, Muhammad Saeed<sup>#3</sup>, Farhan Ahmed Siddiqui<sup>#4</sup>, Nasir Touheed<sup>\*5</sup>,  
and M. Shahid Qureshi<sup>\*6</sup>

**\* Faculty of Computer Science, Institute of Business Administration  
Karachi, Pakistan**

**# Department of Computer Science, Karachi University  
Karachi, Pakistan**

## Abstract

VoIP is a technology in great demand these days. Its interactive nature makes it very appealing for users and today it is one of the most dominant technologies for communication. With the growth over wireless networks the option to have voice communication over wireless has been considered - the choices are VoIP over WiFi or VoIP over WiMax. This paper studies and compares the two options and summarizes the results.

**Keywords:** Packet loss, jitter, throughput, congestion window size, QoS.

## 1. Introduction

Recently wireless technology has grown immensely in popularity and usage becoming a medium of choice for networks. The wireless communication revolution is bringing fundamental changes to data networking, telecommunication, and is making integrated networks a reality. By freeing the user from the cord, personal communications networks, wireless LAN's, mobile radio networks and cellular systems, harbor the promise of fully distributed mobile computing and communications, anytime, anywhere.

A similar trend is seen in the world of voice communication and now transmission of voice over wireless communication links is very common as is obvious from the huge adoption of mobile telephony around the world [1]. One example of a rapidly growing voice application is VoIP as can be evidenced from high success rates of applications like Skype [2]. Voice over Internet Protocol (VoIP) technology facilitates packet based IP networks to carry digitized voice, it uses Internet Protocol for transmission of voice as packets over IP networks [12] thereby dramatically improving bandwidth

efficiency and facilitates creation of new services. VoIP has enabled service providers to offer telephony services along with traditional data services using the same IP infrastructure and this in turn leads to improvement of business models.

However one fundamental question that arises is: "Can we get good VoIP quality over wireless networks while at the same time maintaining its traditional role for data services?"

We have addressed this question in this study by doing measurement analysis of VoIP over both WiFi and WiMax networks. The approach adopted is based on simulation using the well-known networking research simulation tool ns2 [3]. We performed two experiments: one for the case of IEEE 802.11 and the other for the case of IEEE 802.16.

VoIP packets are sent in conjunction with TCP packets and the performance of network is analyzed through various characteristics such as jitter, packet losses, throughput and delay.

This paper is organized as follows: Section 2 discusses the issues that arise when using VoIP over wireless networks. Section 3 explains the simulation scenario. Section 4 presents measurement results and graphs along with explanation along with an explanation of the results.

## 2. Voice going Wireless

Voice is the method of choice for real time communications [4]. Voice is so important to human communications that we have constructed entire networks centered around voice, namely, the public switched telephone network (PSTN) [5] and the analog/digital

cellular networks [6]. Computer networks were originally developed with data transmission in mind, but the needs of Internet users today are diverse; no longer is the need for transmitting only data traffic over the Internet but there is also need to make VoIP calls, play online games and watch streaming media. Indeed, voice over the Internet Protocol (VoIP) is growing rapidly and is expected to do so for the near future. A new and powerful development for data communications is the emergence of wireless local area networks (WLANs) in the embodiment of the 802.11 a, b, g standards [7, 8], collectively referred to as Wi-Fi [8]. Because of the proliferation and expected expansion of Wi-Fi networks, considerable attention is now being turned to voice over Wi-Fi, with some companies already offering proprietary networks, handsets, and solutions. However deployment of VoIP over WiFi poses some serious problems and concerns. This is the main reason why the shift is now towards WiMax.

In this paper we take up a comparative study based on measurement analysis of “simulated packet traces.” The results are compared to see which option is more viable: VoIP over WiFi or VoIP over WiMax.

## 2.1 VoIP Issues on IEEE 802.11

Wireless Local Area Networks (WLANs) are increasingly making their way into residential, commercial, industrial and public areas. As VoIP applications flourish [2] voice will be a significant driver for widespread adoption and integration of WLAN. As such voice capacity of a WLAN, which is defined as the maximum number of voice connections that can be supported with satisfied quality, has been investigated in the literature [9, 10]. The capacity of G.711 VoIP using constant bit rate (CBR) model and a 10 ms packetization interval is 6 calls. The two main problems encountered when VoIP is used over WiFi are:

- The system capacity for voice can be quite low for WLAN.
- VoIP traffic and traditional data traffic such as Web traffic, emails etc. can mingle with each other thereby bringing down VoIP performance.

These problems exist mainly due to the following reasons:

- a) There is large per-packet overhead imposed by WiFi for each VoIP packet – for both protocol headers and WiFi contention.
- b) Design of 802.11 protocols is such that it allows clients to access the channel in a distributed manner which causes a contention for the network which is particularly evident in the case of VoIP due to the real-time nature of the traffic.

Hence in the case of VoIP over WLAN the perceived throughput and real throughput have a large difference. Even though it does seem as an attractive alternative to cellular wireless telephony it has several drawbacks as we shall further investigate in section 4 of this paper.

## 2.2 VoIP on IEEE 802.16

IEEE 802.16 [11] is the “de facto” standard for broadband wireless communication. It is considered as the missing link for the “last mile” connection in Wireless Metropolitan Area Networks (WMAN). It represents a serious alternative to the wired network, such as DSL and cable modem. Besides Quality of Service (QoS) support, the IEEE 802.16 standard is currently offering a nominal data rate up to 100 Mega Bit Per Second (Mbps), and a covering area around 50 kilometers. Thus, a deployment of multimedia services such as Voice over IP (VoIP), Video on Demand (VoD) and video conferencing is now possible, which will open new markets and business opportunities for vendors and service providers.

Concerning QoS support, the 802.16 standard proposes to classify, at the MAC layer, the applications according to their QoS service requirement (real time applications with stringent delay requirement, best effort applications with minimum guaranteed bandwidth) as well as their packet arrival pattern (fixed / variable data packets at periodic / aperiodic intervals). For this aim, the initial standard proposes four classes of traffic, and the 802.16e [11] amendment adds another class:

- Unsolicited grant service (UGS): supports Constant Bit Rate (CBR) services, such as T1/E1 emulation and VoIP without silence suppression.
- Real-time polling service (rtPS): supports real-time services with variable size data on a periodic basis, such as MPEG and VoIP with silence suppression.
- Extended rtPS : recently introduced by the 802.16e standard, it combines UGS and rtPS. That is, it guaranties periodic unsolicited grants, but the grantsize can be changed by request. It was speciallyintroduced to support VoIP traffics [11].



- Non Real-Time Polling service (nrtPS): supports non real-time services that require variable size data bursts on regular basis, such as File Transport Protocol (FTP) service.
- Best effort (BE): for applications that do not require QoS such as Hyper Text Transfer Protocol (HTTP).

Due to the above-mentioned QoS implementations on IEEE 802.16 VoIP performs better on WiMax as we shall see in the next section.

### 3. Experimental Setup

To investigate performance of VoIP with TCP on IEEE 802.11 and IEEE 802.16 simulations were undertaken using TCP flows along with CBR flows (defined on top of UDP flows). UDP was used for the VoIP data flow and the UDP packet properties were those of the G.711 codec [13].

Figure 1 shows the simulation setup in ns2. In this network both VoIP and TCP/IP data traffic will be used to test the network performance for VoIP.

The setup is composed of two wired nodes, three mobile nodes and a base station serving as the access point for the WiFi network in case of Experiment 1 and for the WiMax network in case of Experiment 2. In both the experiments the deployment of the network was kept the same but the TCP and VoIP flows were varied each time.

Also the number of flows was varied: the simulation part was done with ns2 whereas for analysis purposes the Linux utilities xgraph and gnuplot were used.

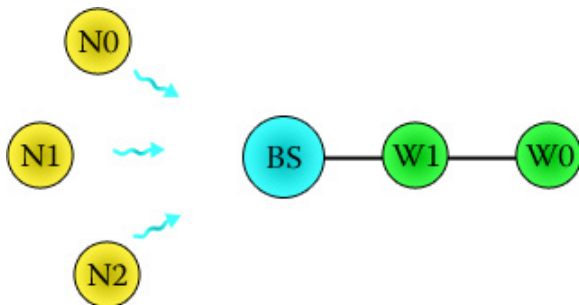


Fig. 1 Setup for Experiment

#### 3.1 Experimental Details for Flows of TCP and VoIP

VoIP is basically CBR UDP: typical data rates and packet sizes can be obtained for voice codecs by doing a search for VoIP - typical data rates EXCLUSIVE of header overhead are from 5.3 to 64 kbps, depending on the implementation and application. Packet sizes are usually kept short to minimize latency.

Hence in the ns2 simulation the VoIP packets have been modeled through CBR UDP with a data rate of 80 bytes and a delay of 20 milliseconds which is typical specification for G.711 codec [14].

In the case of the 802.11 scenario two TCP flows are set up: one from node N0 to wired node W0 (it is run from 5 seconds to end of simulation) and the other from wired node W1 to node N2 (it is run from 15 seconds to end of simulation). The VoIP packets are sent from node N0 to wired node W0 and from N2 to wired node W1. There are 16 VoIP flows instantiated simultaneously between N0 and W0 and their start time is 40 seconds, two of them are stopped at 100 second while remaining two at 120 seconds. Between N2 and W1 there are 4 simultaneous VoIP sessions with start times 100 seconds and ending times of the 4 are 120 seconds for first two, 140 seconds for third and 150 seconds for the last one.

In the case of 802.16 scenario the same example as the one provided by NS2 Simulator for IEEE 802.16 network [15] has been used and the topology for it has been shown in Figure 1. In the case of the 802.16 scenario three TCP flows are set up: one from node N0, node N2 and node N3 to wired node W1. Their start times are 0.1, 0.2 and 0.3 seconds and they stop when simulation ends; the VoIP packets are sent from node N0 to wired node W1. There are 8 VoIP flows instantiated simultaneously between N0 and W1 and their start time is 40 seconds out of which two are stopped at 60 seconds and remaining are allowed to run till the end of the simulation.

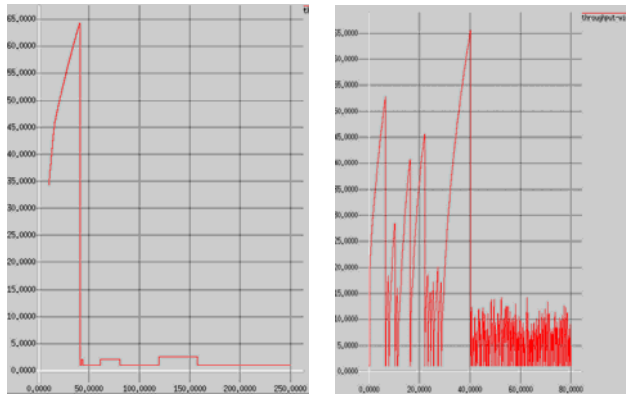
### 4. Experimental Results

This section presents the results for the two experiments. We plotted graphs for throughput, jitter and packet losses in both cases.

#### 4.1 The 802.11 and 802.16 Results Compared

The throughput graph for both cases is shown in Figure 2. From the above graphs it is clear that VoIP over WiFi makes TCP capacity inefficient and as soon as VoIP flows are started the TCP congestion window drops and does not rise again until and unless the VoIP packet sending process drops. So this makes it clear that throughput of VoIP and TCP both are affected by deployment of VoIP over 802.11.

The first graph i.e. Figure 2a shows the scenario for the 802.11 networks and throughput that TCP achieves in presence of VoIP packets being transmitted.

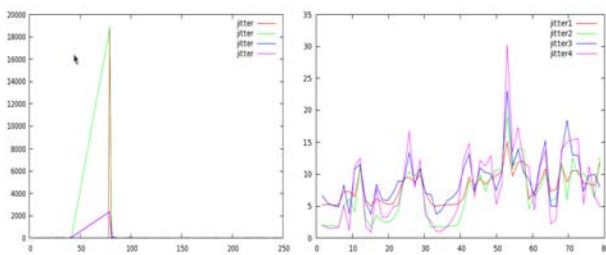


(a) IEEE 802.11 (b) IEEE 802.16

**Fig. 2 Throughput Window for TCP flows in 802.11 and 802.16 Networks**

The graph shows that in presence of VoIP flows on WiFi the TCP capacity is marginally reduced and congestion window is affected badly. On the other hand in the case of WiMax networks the TCP does achieve an acceptable throughput hence demonstrating that WiMax is better suited for real-time services like VoIP.

The next graph in Figure 3 shows the jitter experienced by the TCP packets when VoIP flows and TCP flows exist simultaneously on a wireless link.



(a) IEEE 802.11 (b) IEEE 802.16

**Fig. 3 Jitter for TCP flows in 802.11 and 802.16 Networks**

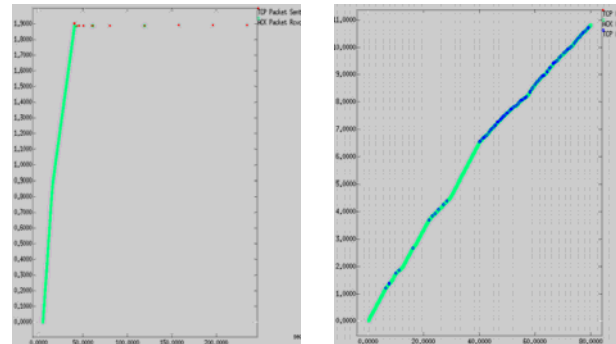
Again this graph shows that VoIP existence on WLAN kills the TCP capacity of the network with very high jitter at times when VoIP packets co-exist and no jitter when there are no VoIP packets. However in the WiMax scenario VoIP packets do not make the network unsuitable for TCP thus proving the claim of WiMax community that “it is the ideal standard for both voice and data.”

Figure 4 shows the graph for packet losses experienced by the TCP packets when VoIP flows and TCP flows exist simultaneously on a wireless link.

The red dots indicate sent packets, green dots are for received packets and blue dots mark the dropped packets over the network.

In the case of IEEE 802.11 networks we have almost no losses when VoIP is not being sent but as soon as we begin to send VoIP packets the congestion window gets

halved due to packet drops at queue. Hence there is almost no further sending and receiving of packets and the network is unutilized by TCP since VoIP completely occupies it. Unlike that IEEE 802.16 networks although do show a packet loss but it is tolerable



(a) IEEE 802.11 (b) IEEE 802.16

**Fig. 4 Packet Losses for TCP flows in 802.11 and 802.16 Networks**

#### 4.2 Other Characteristics

Although the graph has not been shown for the delay, it was however noted by analyzing the packet traces that as soon as CBR traffic was introduced into the network it took quite a long time for the TCP packets to arrive at the destination whilst at the same time VoIP quality suffered.

Moreover fairness was almost non-existent when number of flows was increased; the options of 4, 8 and 16 flows were tried for each case. In case of 16 flows the link containing VoIP traffic behaved as if it is down and due to that bandwidth of network was not equally shared.

#### 4.3 Explanation of the Results

The results obtained and analyzed above were much expected due to the very nature of the two technologies of WiFi and WiMax.

There are three great problems inherent to the WLANs that can harm VoIP performance are:

- The inefficiency of the 802.11 MAC protocol.
- The signal instability caused by electromagnetic phenomena
- The competition for bandwidth usage between voice traffic and data traffic.

These problems not only make the performance of VoIP suffer over 802.11 but also render the network useless for data by choking TCP. On the other hand WiMax’s better performance is attributed to its better QoS services.

WiMax is quite well suited to the promising VoIP applications.

#### 4. Conclusions

All our findings complement the characteristics of both the networks and help in further establishing the fact that WiMax is better suited to VoIP than WiFi.

#### References

- [1] D. P. Hole and F. A. Tobagi, "Capacity of an IEEE 802.11b wireless LAN supporting VoIP," in Proc. IEEE ICC, Jun. 2004, vol. 1, pp. 196–201. Goode B., September 2002. Voice Over Internet Protocol (VoIP). Invited Paper. Proceedings of the IEEE, Vol. 90, no. 9.
- [2] Skype: <http://www.skype.com>  
Network Simulator 2 <http://www.isi.edu/nsnam/ns/>
- [3] B. Teitelbaum, "Leading-edge voice communications for the MITC," Sept. 12, 2003 at <http://people.internet2.edu/~ben/>.
- [4] Forman, G. 2003. An extensive empirical study of feature selection metrics for text classification. J. Mach. Learn. Res. 3 (Mar. 2003), 1289-1305.
- [5] J. C. Bellamy, Digital Telephony, John Wiley & Sons, 2000.
- [6] T.S. Rappaport, Wireless Communications: Principles and Practice, Prentice Hall, second edition, 2002.
- [7] ISO/IEC and IEEE Draft International Standards, "Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," ISO/IEC 8802-11, IEEE P802.11/D10, Jan. 1999.
- [8] [http://wi-fiplanet.webopedia.com/TERM/w/Wi\\_Fi.html](http://wi-fiplanet.webopedia.com/TERM/w/Wi_Fi.html)
- [9] F. Anjum, M. Elaoud, D. Famolari, A. Ghosh, R. Vaidyanathan, A. Dutta, P. Agrawal, T. Kodama, and Y. Katsube. Voice performance in WLAN networks-an experimental study. Global Telecommunications Conference, 2003. GLOBECOM'03. IEEE, 6, 2003.
- [10] S. Garg and M. Kappes. An experimental study of throughput for udp and voip traffic in ieee 802.11b networks. Wireless Communications and Networking, 2003. WCNC 2003. 2003 IEEE, 3:1748–1753 vol.3, 16-20 March 2003.
- [11] IEEE standard for local and metropolitan area networks, Part 16: Air Interface for fixed broadband wireless access systems, IEEE Standard 802.16, October 2004.
- [12] Goode B., September 2002. Voice Over Internet Protocol (VoIP). Invited Paper. Proceedings of the IEEE, Vol. 90, no. 9.
- [13] Voice over IP – Per Call Bandwidth Consumption  
[http://www.cisco.com/en/US/tech/tk652/tk698/technologies\\_tech\\_note09186a0080094ae2.shtml](http://www.cisco.com/en/US/tech/tk652/tk698/technologies_tech_note09186a0080094ae2.shtml)
- [14] [http://www.cisco.com/en/US/tech/tk652/tk698/technologies\\_tech\\_note09186a0080094ae2.shtml](http://www.cisco.com/en/US/tech/tk652/tk698/technologies_tech_note09186a0080094ae2.shtml)
- [15] <http://cnlab.kaist.ac.kr>

# IBook: Interactive and Semantic Multimedia Content Generation for eLearning

Arjumand Younus\*<sup>1</sup>, M. Atif Qureshi\*<sup>2</sup>, Muhammad Saeed\*<sup>3</sup>, Syed Asim Ali\*<sup>4</sup>, Nasir Touheed\*<sup>5</sup>,  
and M. Shahid Qureshi\*<sup>6</sup>

\* Faculty of Computer Science, Institute of Business Administration  
Karachi, Pakistan

# Department of Computer Science, Karachi University  
Karachi, Pakistan

## Abstract

Over the years the World Wide Web has seen a major transformation with dynamic content and interactivity being delivered through Web 2.0 and provision of meaning to Web content through the Semantic Web. Web 2.0 has given rise to special methods of eLearning; we believe that interactive multimedia and semantic technologies applied together can further enable effective reuse of such applications thereby taking eLearning a step further. As proof of this idea we present IBook which is an eLearning application that uses concepts from both the fields of Web 2.0 and Semantic Web. It presents multimedia in a form that enhances the user's learning experience through the use of Web 2.0 and Semantic Web.

**Keywords:** *Web 2.0, Semantic Web, Multimedia, eLearning.*

## 1. Introduction

With a proliferation of Web 2.0 services and applications there has been a major paradigm shift in the way we envision the World Wide Web [3, 4]. We have witnessed an evolution of the Web from the first generation to the third generation [1, 2] and at present we live somewhere between the age of second generation and third generation Web content. This age can be termed as a "transition stage" between Web 2.0 [3, 4] and the envisioned Semantic Web [5] and in this transition phase there has been a realization of new concepts such as e-Science, e-Education, e-Learning, e-Commerce, e-Government etc.

The realization of these new technologies has given birth to new forms of multimedia in the World Wide Web and this is in particular the case with eLearning [6] with many adaptive hypermedia learning applications being developed. This paper also presents one such application

IBook and what sets it apart from other similar works are its additional features of interactive multimedia content facilitating effective learning with semantic technologies i.e. XML[7].

IBook is an application that takes an innovative approach for eLearning which lies in both domains: Web 2.0 and the Semantic Web. Linear text was challenged by the world of the Internet which led to the creation of hypertext but even that suffered some drawbacks which led to the concept of hypermedia [9]. The students of today have done away with books and look to the Internet to support their learning. A widespread argument now exists among teachers, educators and psychologists that advanced comprehension is acquired through interacting with the content [8] and this is the fundamental motivation behind IBook. We feel that semantically connected data in multiple dimensions can bring a remarkable change in the learning curve and experience and this is where IBook plays its role.

As is clear from the name IBook is an interactive, multimedia based book which provides the reader with additional forms of presentations for enhanced delivery of the book's contents. Moreover the book not only follows its classical front view but also possesses great details to explain it further by adding relevant video content as well as voice over feature to retain readers' attention to the most. Hence IBook is an advanced multimedia platform for eLearning. With IBook the educator can add flexibility and easy adaptation to new and changing user requirements through support for a reusable metadata structure.

The remainder of this paper is organized as follows: section II explains the necessary background with respect to the generations of Web content, section III explains in detail the IBook features and functionalities with illustrations. Section IV presents the architecture and implementation details of the IBook framework with an overview of how semantic technologies are incorporated into it. Section V concludes the paper with a discussion of possible future works.

## 2. Background

As mentioned in section I IBook is an application from the areas of Web 2.0 and Semantic Web and this section provides a brief overview of each of these areas.

Some researchers characterize the Web evolution in terms of generations with the first generation containing static HTML content [1, 2] which was and is still being replaced by dynamic, on-the-fly Web content giving rise to the second generation of Web technologies and applications [4]. Second generation Web technology mainly focused on addressing needs of humans but in contrast third generation Web technology is more focused at making content that is machine-processable.

### 2.1 Web 2.0

The term Web 2.0 stands not for a system but a design philosophy applied to the first generation Web content and with application of this design philosophy emerged a whole new range of applications which facilitated interactive information sharing, interoperability, user-centered design and collaboration on the World Wide Web. Web 2.0 applications include a whole new array of applications some examples being social networking sites like Facebook and MySpace, video sharing sites like YouTube, wikis, blogs, mashups and folksonomies [14].

The fundamental idea behind Web 2.0 is the use of the “Web as a platform” with software applications moving from the desktop to this easily accessible platform enabling rich interaction and user participation – the two things we also bring into IBook.

### 2.2 Semantic Web

The Semantic Web is the third generation Web platform which is more focused towards meaning of information and services on the Web making it possible for machines to process the content in order to enhance the user experience. It is more of a vision of the early pioneers of the World Wide Web and this vision sees the Web as a universal medium for data, information and knowledge exchange [10]. It proposes markup of content on the Web

with the help of formal ontologies for structuring of the underlying data for the purpose of comprehensiveness and machine understanding. The Semantic Web is an extension of the current Web in which information is given well-defined meaning, enabling computers and people to work in co-operation.

### 2.3 Integration of Web 2.0 with the Semantic Web

Earlier when O'Reilly Media and MediaLive hosted the first Web 2.0 conference in 2004 and the term “Web 2.0” was used the inventor of the World Wide Web Sir Tim-Berners Lee discarded it as being a “buzzword” or “piece of jargon” but recently some researchers have presented a different viewpoint [11, 12, 13]. Researchers are now talking about a merger of the two ideas of Web 2.0 and the Semantic Web and are now upholding the belief that the two fields are “complementary rather than competing with goals being in harmony and each bringing its own strength into the picture” [11]. This is also the line of reasoning we follow in this paper and advocate the idea of integration of Web 2.0 technologies with the Semantic Web ideas for effective methods of eLearning.

## 3. IBook Features and Functionalities

We now present a high-level view of IBook describing its features and functionalities in detail.

This section mainly contains a description of the features through that IBook can support and the implementation details are explained in the next section.

The programming platform used was Adobe Flash with Action Script 3.0. Figure 1 presents the front end view of IBook:

Broadly viewed we can define IBook features in terms of following characteristics of the Web 2.0 and the Semantic Web:

- Non-linear Textual Nature
- User Interactivity
- Multimedia Support
- Content Personalization and Reuse

### 3.1 Non-linear Textual Nature

The innovations in Web technology have refined the traditional use of the book adding to it the dimension of random access rather than linear, sequential access. This is what is also supported in IBook. As mentioned in the introduction IBook presents the content of the book in its classical format with the added feature of having hyperlinks to each chapter. This eases the process of



navigating into the book for content and gives the user an extra level of interactivity which closely mimics the real-world book as shown in the table of contents view shown in Figure 1. When the user clicks a particular chapter for viewing he is presented with the view shown in Figure 2. Here the reader is not only able to read the chapter's contents but can also listen to it with voice over feature: as soon as chapter opens the text of the chapter is played with the portion that is being played highlighted in yellow. The voice over facility is what makes IBook particularly unique and sets it apart from other works in the eLearning domain: this is the first such work which gives user an extra level of multimedia interactivity with voice over capability thereby being able to grasp his attention towards the content of the book. The reader is also given the capability to stop or pause the audio at any point thereby adding interactivity to the reading/listening process. Navigation features are also included within each chapter of IBook while the reader browses through the book.



Figure 1 IBook Front View

These quizzes can be user-defined and how this is achieved is explained in detail in the next section.

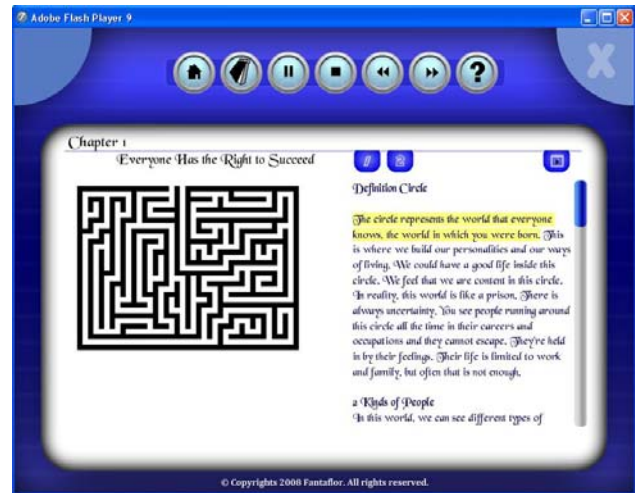


Figure 2 IBook Chapter View



Figure 3 Quiz 1 Linkage Activity

### 3.2 User Interactivity

IBook also contains additional features to enable user interactivity in order to facilitate deep comprehension of the subject; these features incorporate the facility to check if the user understands the chapter's contents through chapter quizzes and enhancement of that understanding through video summary of the chapter. The first quiz as shown in Figure 3 is termed "Linkage Activity" and asks user to match the correct terms with each other by drawing a line between them and the second quiz as shown in Figure 4 is called "Drag and Drop Activity" and it asks the user to fill in the appropriate boxes with the connected terms by dragging and dropping. These quizzes are included in each chapter and test the user's knowledge of the chapter's content hence enhancing the user interactive experience and providing more efficient concept delivery.

### 3.3 Multimedia Support

Another useful feature provided in IBook is the concept of video summaries corresponding to each chapter; an illustration is presented in Figure 5. As the figure shows when the user clicks on the video summary button a popup video appears which summarizes the contents of the current chapter in video form. Like the interactive quizzes the video summary can be user-defined and supports rich user-defined semantics as explained in the next section. Hence we can see that IBook provides a complete multimedia platform with audio, video, images and interactive content to enhance the user's learning experience which sharpens the learning curve greatly.

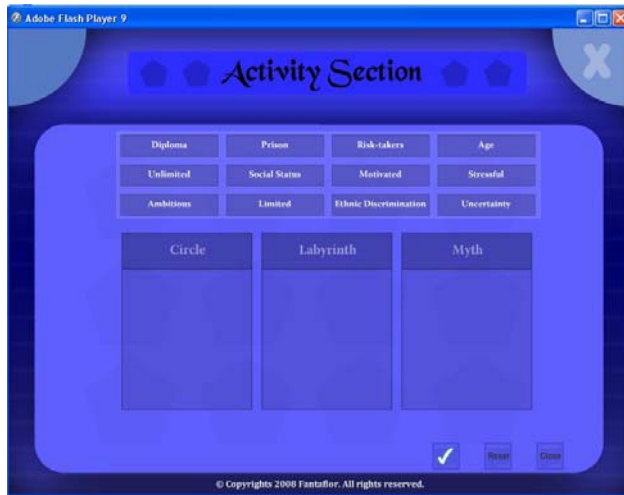


Figure 4 Quiz 2 Drag and Drop Activity



Figure 5 Chapter Video Summary

### 3.4 Content Personalization and Reuse

One of the key aspects of the IBook system is the ability to support personalization of the book's content. As mentioned in the Introduction Web 2.0 enables rich dynamicity into Web content and this is what we have taken advantage of for IBook: the content and the underlying structure are stored separately from the details of presentation of that content and this enables effective adaptation of the content to specific needs of users. In addition to personal adaptation the dynamic nature of Web 2.0 enables effective reuse of the underlying structure and this reuse has been enabled largely with the help of Semantic Metadata standards

## 4. IBook Framework and Architecture

This section explains the architectural framework of IBook. As explained in earlier sections IBook uses technologies from the domain of Web 2.0 for interactivity and Semantic Web for reusable, user-defined content support through metadata standards.

### 4.1 High-Level Architectural View

Figure 6 depicts the high-level architectural view of IBook. The semantic module is basically composed of three parts with the first part describing the book's text and audio content, the second and third part contain metadata and user-defined semantics for video summary and the quiz activity.

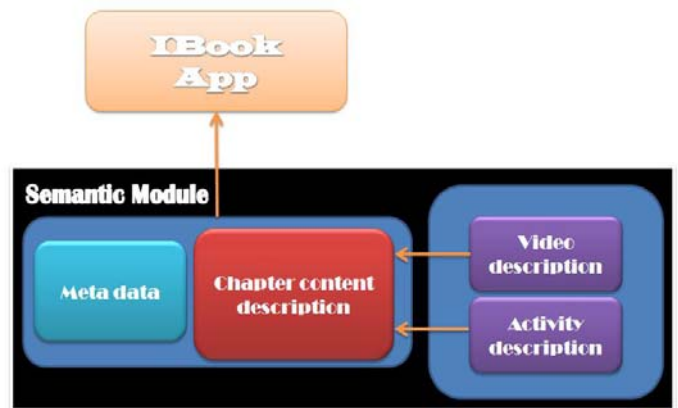


Figure 6 IBook High-Level Architecture

The module for video summary and quiz activity is kept as a stand-alone for the purpose of greater flexibility. This also has the added benefit of using these modules independently i.e. creation of video content as a standalone video presentation or creation of quiz activities as standalone quizzes. We now explain each of these modules in detail.

### 4.2 Semantic Metadata Module

The module for managing the metadata uses the semantic technology to facilitate user-defined content creation and reuse of that content. XML is also a widely used standard for Web 2.0 development designed for markup in documents of arbitrary structure. For each part of the IBook we use XML specification which makes the content machine-processable and adds descriptive information about the IBook resources for purpose of finding, managing and using them more effectively. The XML portion for each chapter is shown below:

```
<book>
  <toc>
    <Title>Book Title</Title>
    <Heading>TOC</Heading>
    <data>
      <![CDATA[Chapter Contents Here]]>
    </data>
    <image>TOC Image</image>
    <audio>TOC Audio</audio>
    <pos>Audio Positions</pos>
    <pos>Audio Positions</pos>
    <pos>Audio Positions</pos>
  </toc>
  <content>
    <Title>Chapter 1 Title</Title>
    <Heading>Chap Heading</Heading>
    <data>
      <![CDATA[Chapter Contents Here]]>
    </data>
    <image>Chapter Image</image>
    <topicAudio>Audio</topicAudio>
    <audio>Chapter Audio</audio>
    <videoTitle></videoTitle>
    <pos>Audio Positions</pos>
    <pos>Audio Positions</pos>
    <pos>Audio Positions</pos>
  </content>
  <content>
    .
    .
    .
  </content>
</book>
```

**Table 1 XML Specification for IBook**

As is clear from this XML structure the <toc></toc> nested part specifies the contents for the table of contents of the book: including the image for table of content, heading and audio for voice over. Similarly the tags for content i.e. <content></contents> allow the user to give specification for chapter's audio, images and video summary title. The <pos> tags specify the cue positions for the audio that is played with the highlight feature; these positions mark the positions of sentences in the audio and decide when the next sentence is to be highlighted. As this structure shows the user can fill the IBook with his changing data and can also specify many other features of his choice.

In a similar manner the quiz activities along with its questions and answers and the video summary can be specified through XML tags based on each chapter's

contents. The video summary is basically an animation made up of a sequence of images with effects in between; also there is the additional capability of using text in between the moving images for description. The video module is basically a stand-alone multimedia presentation generation system which combines text and graphics in real-time enabled through XML metadata specification.

This semantic module provided by IBook enhances the user's experience with multimedia interactivity and hence replaces the traditional book concept.

## 4.2 The Web 2.0 and Semantic Web Experience

The Semantic Web takes the Web experience further by making transforming computers from a passive entity to an active entity in the process and offers a generic infrastructure for interchange, integration and creative reuse of structured data: these features of the Semantic Web can help it overcome the problems and limitations of the current Web 2.0. As demonstrated by IBook adding semantics to Web 2.0 provides more reuse possibilities and creates richer links between content items: audio, video and interactive quizzes in the IBook case. We further advocate the case presented in [11]: 1) using layers of Web 2.0 to lead towards the Semantic Web dream and 2) using semantic technologies for providing a robust and extensible basis for emerging Web 2.0 applications.

## 4. Conclusions

The collaboration between Web 2.0 technologies and Semantic Web technologies can give birth to exciting, interactive applications. We experimented with the idea and developed one such eLearning application IBook which introduces how enhanced learning can be delivered through modern Web technology. The future works for this application include extension of the semantic technologies to make it closer to be realized as a Semantic Web application and one example of this can be incorporation of semantic search technology within the book framework. Another significant future extension we plan for IBook is the automation and ease of use for the content generation process for IBook which at present is manual.

All in all we believe applications like IBook are a significant move towards the envisioned "Semantic Web" which is likely to become a reality by efficient and effective use of Web 2.0 technologies.

## References

- [1] J. van Ossenbruggen, J. Geurts, F. Cornelissen, L. Hardman and L. Rutledge, "Towards Second and Third Generation Web-

Based Multimedia,” In Proceedings of the Tenth International World Wide Web Conference (WWW10), May 2001, Hong Kong, ACM Press, 479-488.

[2] S. Decker et al., “Knowledge Representation on the Web,” In F. Baader, editor, International Workshop on Description Logics (DL’00), 2000.

[3] P. Anderson. “What is Web 2.0? Ideas, technologies and implications for education.” Technical report, JISC, 2007.

[4] T. O’Reilly. “What is Web 2.0: Design Patterns and Business Models for the next generation of software,” 2005 <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>

[5] T. Berners-Lee, J. Hendler and O. Lassila. “The Semantic Web,” Scientific American, May 2001.

[6] L. Stojanovic, S. Staab and R. Studer, “Elearning based on the Semantic Web,” World Conference on the WWW and Internet (WebNet), Orlando, Florida, USA, 2001.

[7] V. Geroimenko and C. Chen, “Visualizing the Semantic Web: XML-based Internet and Information Visualization”, London: Springer-Verlag, 2003.

[8] A. El Saddik, S. Fischer and R. Steinmetz, “Reusable Multimedia Content in Web-Based Learning Systems”, IEEE Multimedia, July-Sept 2001, 30-38.

[9] T. Nelson, “A File Structure for the Complex, the Changing, and the Indeterminate.” In Proceedings of ACM National Conference, 1965, pp. 84–100.

[10] Herman, Ivan (March 7, 2008). "Semantic Web Activity Statement". W3C. Retrieved March 13, 2008.

[11] A. Ankolekar, M. Krötzsch, T. Tran, and D. Vrandečić. The two cultures: Mashing up web 2.0 and the semantic web. *Web Semant.*, 6(1):70–75, 2008.

[12] U. Bojars, J.G. Breslin, A. Finn, S. Decker : "Using the Semantic Web for linking and reusing data across Web 2.0 communities", *J. Web Sem.* 6: 21-28, 2008.

[13] J. Hendler, J. Goldbeck, “Metcalf’s law, web 2.0 and the Semantic Web.” *J. Web Sem.* 6(1):14-20, 2008.

[14] Sadasd Graham Cormode and Balachander Krishnamurthy. “Key differences between Web 1.0 and Web 2.0.” *First Monday*, 13(6), June 2008.

# Applying RFID Technology to construct an Elegant Hospital Environment

A.Anny Leema<sup>1</sup>, Dr.Hemalatha.M<sup>2</sup>

<sup>1</sup> Assistant Professor, Computer Applications Department, B.S.Abdur Rahman University  
PhD Scholar, Karpagam University, Coimbatore, India

<sup>2</sup> Head of MSc. Software Systems,  
Karpagam University, Coimbatore, India

## Abstract

Radio frequency identification (RFID) technology has seen increasing adoption rates in applications that range from supply chain management, asset tracking, Medical/Health Care applications, People tracking, Manufacturing, Retail, Warehouses, and Livestock Timing. Of these, Medical/Health care applications are of more importance because minute errors in it can cost heavy financial and personal losses. The success of these applications depends heavily on the quality of the data stream generated by the RFID readers. Efficient and accurate data cleaning is an essential task for the successful deployment of RFID systems. Hence this paper gives the brief introduction of RFID terminologies and cleaning methods to provide accurate RFID data to applications. It also outlines a patient management system which helps hospitals to build a better, more collaborative environment between different departments, such as the wards, medication, examination, and payment. Indeed used in hospital to bring down the health care costs, optimizing business processes, streamline patient identification processes and improve patient safety.

*Keywords: RFID technology, cleaning methods, smart hospital, health care systems*

## 1. Introduction

RFID technology uses radio-frequency waves to automatically identify people or objects. There are several methods of

identification, but the most common is to store a serial number that identifies a person or object, and perhaps other information, on a microchip that is attached to an antenna (the chip and the antenna together are called an RFID transponder or an RFID tag). The antenna enables the chip to transmit the identification information to a reader [2]. The reader converts the radio waves reflected back from the RFID tag into digital information that can then be passed on to computers that can make use of it. RFID is automatic and fast and will replace the barcode system in the near future. The big difference between RFID and barcodes is line-of-sight technology. That is, a scanner has to see the barcode to read it, which means people usually have to orient the barcode toward a scanner for it to be read, RFID by contrast, doesn't require line of sight. RFID tags can be read as long as they are within range of a reader[5].

RFID technology is widely used in diverse application such as supply chain automation, Asset tracking, Medical/Health Care applications, People tracking, Manufacturing, Retail, Warehouses, and Livestock Timing. Of these, Medical/Health care applications are of more importance because minute errors in it can cost heavy financial and personal losses. For hospitals and healthcare systems, increasing the operational efficiency is the primary target. It is a tough task to keep up the effectiveness and monitor each and every patient [7]. However, utilization of RFID (Radio Frequency Identification) technology in addition to reducing the health care costs facilitates automating and streamlining patient identification processes in



hospitals and use of mobile devices like PDA, smart phones, design of health care management systems etc., [4]. The emerging RFID technology is rapidly becoming the standard for tracking inventory, identifying patients, and managing personnel in hospitals [7]. In hospitals patient safety is critically important; lives are at stake, and zero defects should be the established standard. At the same time, hospitals are pressured to reduce costs. Therefore, when developing strategic objectives, technologies that reduce operating expenses while providing increased patient safety must be thoroughly tested and evaluated. Radio frequency identification (RFID) is one technology that holds great promise.

Recent years, in almost every country in the world, substantial financial resources have been allocated to the health care sector. Technological development and modern medicine practices are amongst the outstanding factors triggering this shift. Achieving a high operational efficiency in the health care sector is an essential goal for organizational performance evaluation. Efficiency uses to be considered as the primary indicator of hospital performance [1].

The goal of this paper is to show how RFID contributes to build an elegant hospital by optimizing business processes, reducing errors and improving patient safety. This section starts by a short introduction to the RFID technology and define some of its main concepts and standards. The second section describes some interesting hospital use cases that could benefit from RFID and the third section outlines the cleaning methods and finally developed a health care system. We also summarized the open problems that still have to be solved before RFID is fully adopted by the healthcare community.

## 2. Building an elegant Hospital

### 2.1 Existing Problems in Hospital

Healthcare providers (i.e., hospitals) traditionally use a paper-based 'flow chart' to capture patient information during registration time, which is updated by the on duty nurse and handed over to the incoming staff at the end of each shift. Although, the nurses spent large amount of time on

updating the paperwork at the bedside of the patient, it is not always accurate, because this is handwritten.

In thousands of hospitals across the world, blood transfusion is an everyday business, but fraught with risks. This is because contaminated blood may be transfused to a healthy patient or receiving wrong type of blood. Data from US hospitals shows an alarming number of cases of medical negligence or mistakes, many of which are related to blood transfusion.

Many health professionals are concerned about the growing number of patients who are misidentified before, during or after medical treatment. Indeed, patient identification error may lead to improper dosage of medication to patient, as well as having invasive procedure done. Other related patient identification errors could lead to inaccurate lab work and results reported for the wrong person, having effects such as misdiagnoses and serious medication errors [4].

### 2.2 Potential Benefits of RFID technology

The RFID solution to the above said problem is to embed a tag into the blood bag label itself. The parametric who transfuses the blood can scan the bag before transferring. He typically enters the patient ID number and the patient also has a wrist band RFID tag which identifies him uniquely. In case the wrong blood bag is scanned, the reader can throw up a warning given below and the patient is saved from wrong treatment.[3]

<b>WARNING BLOOD MISMATCH!!!</b>	
The Blood bag is for patient ANNY Patient ID A1000	The patient on the bed is ANNIE Patient ID is A0100

The RFID patient tracking kit consists of RFID wristbands, a PDA handheld reader, a desktop HF reader and necessary software. Because of automated data capture, the RFID patient tracking kit brings improved

efficiency. The waterproof, non-allergic wristband can be reprogrammed to enable patient information to be stored and transferred to and from RFID readers, information systems, and medical devices in hospital. The Handheld RFID reader is used to receive the patient's real-time information just beside the beds, whereas desktop reader is used to read/write wristband's information beside computer to save time. Hospitals can use this RFID patient tracking kit to boost efficiency and accuracy while reducing costly and dangerous errors, and giving patient more privacy. [6]

Patients are monitored in many hospitals whether proper care is given or not. These systems tend to reduce the data-entry workload of nurses, and also let them spend more time caring for patients and automate the process of billing. Additionally, hospitals are tracking high-value assets, including gurneys, wheel chairs, oxygen pumps and defibrillators. These systems reduce the time employees spend looking for assets, improve asset utilization and enhance the hospitals' ability to performed scheduled maintenance.

Patient bracelets embedded with RFID technology securely tracks patient movement from admission to discharge. The Orthopedic Hospital of Oklahoma (OHO) uses RFID technology and thin client computing solution from Sun Microsystems Inc. to significantly enhance the overall hospital experience for its patients.

An Active Wave RFID system can be used to track patients, doctors and expensive equipment in hospitals in real time. RFID tags can be attached to the ID bracelets of all patients, or just patients requiring special attention, so their location can be tracked continuously. It also Restrict access to drugs, pediatrics, and other high-threat areas to authorized staff.

Moreover, RFID readers are placed at strategic places within the hospital:

- RFID gates are inclined at entrances and exits of the hospital.

- At least one RFID reader can be placed for each operating theatre.

- RFID sensors are placed in strategic galleries and important offices. In the best case, every office should contain an RFID reader: either placed next to the door or under the desks.

- The staff members (doctors, nurses, caregivers and other employees) each have a handheld (PDA, mobile phone, etc.) equipped with an RFID reader and possibly with a wireless (e.g. WiFi) connection to the web.

### 3 Cleaning Methods

Data cleaning is essential for the correct interpretation and analysis of RFID data. To increase patients' safety the major challenge is to clean the data so that it is "fit for use". Efficient and accurate data cleaning is an essential task for the successful deployment of RFID systems. The standard data-cleaning mechanism in most systems is a *temporal "smoothing filter"*. The goal is to reduce or eliminate dropped readings by giving each tag more opportunities to be read within the smoothing window. SMURF which was proposed by UC Berkeley dynamically adjusts the size of a window to pre-treat RFID data. However, SMURF does not work well in determining the size of slide window for frequently moving tags.

In EPS (Extensible Sensor Stream Processing) the static size of the window is the limitation of the approach because large window induces false positives and small window cannot fill false negatives. A new cleaning approach based on the virtual spatial granularity, named bSpace overcome the weakness of the existing techniques. It uses a Bayesian estimation strategy to compute the times that the tag has been detected, in order to fill up false negatives for dynamic tags; and it uses the rules to solve false positives.

In order for RFID technology to become feasible, RFID middleware must be able to produce reliable streams describing the physical world. Cleaning of RFID data sets can be an expensive problem. Existing work on RFID cleaning mainly focused on improving the accuracy of a stand-alone technique and largely ignored costs. Cost conscious cleaning method is based on Dynamic Bayesian networks. The above said cleaning algorithms have its own benefits and drawbacks. SMURF is one of the

recognized data cleaning approaches. [8] However it does not have good performance when tag moves rapidly in and out of reader's communication range, reading frequency and velocity of tag movement. SMURF gives only the empirical value of  $\delta$  and does not tell how to calculate it [9]. To improve the algorithm performance the size of the sliding window is computed by adjusting the parameter  $\delta$ . The simulation shows the error rate is lower and not completely removed.

#### 4 Patient Management System

The important data (e.g., patient ID, name, age, location, drug allergies, blood group, drugs that the patient is on today) can be stored in the patient's back-end databases for processing. The databases containing patient data can also be linked through Internet into other hospitals databases [5]. The Patient Management System's administrator can issue unused tag (wristband) to every patient at registration time. Healthcare professionals (e.g., doctors, consultants) can edit/update password protected patient's medical record for increased patient and data security by clicking the Update Patient Button. This PMS can be implemented in departments (e.g., medicine, surgery, obstetrics and gynecology, pediatrics) in both public and private hospitals for fast and accurate patient identification without human intervention. Using HPMS, health care providers (e.g., hospitals) have a chance to track fast and accurate patient identification, improve patient's safety by capturing basic data (such as patient unique ID, name, blood group, drug allergies, drugs that the patient is on today), prevent/reduce medical errors, increases efficiency and productivity, and cost savings through wireless communication. The PMS also helps hospitals to build a better, more collaborative environment between different departments, such as the wards, medication, examination, and payment.

#### 5 Conclusions and Future Work

Health care is an important sector that can obtain great benefits from the use of the RFID technology. In this paper, we have analyzed the use of RFID in the health care sector and also described some interesting applications with promising perspectives. Although a number of great ideas and systems can be found in the literature, there is a number of issues that have

not been analyzed yet. [7] We summarize some points that should be addressed in the near future:

1. When talking about pasting radio frequency tags on drug packages, there are concerns that exposure to electromagnetic energy could affect product quality.
2. RFID-based systems can fail due to several reasons (e.g. RFID tags can be destroyed accidentally or, communications can be broken due to interferences). There is a need for real-time fault tolerant RFID systems able to deal with situations in which patients lives could be in danger.
3. RFID components interact wirelessly, thus, attackers have plenty of opportunities to eavesdrop communications and obtain private data of the patients. [6] These data can be used by the eavesdropper to blackmail patients, or by an insuring company to raise prices to their clients. Security and privacy in RFID technology is a very active research field that has the challenge to design scalable and cheap protocols to guarantee the privacy and security of RFID users.

#### References

- [1] P.F. Drucker, "The essential Drucker: selections from the management works of Peter F. Drucker", *New York: HarperBusiness*, 2001.
- [2]. Sudarshan S. Chawathe, Venkat Krishnamurthy, Sridhar Ramachandran, and Sanjay Sarma, "Managing RFID Data", In Proceedings of the 30th VLDB Conference, pp.1189-1195, 2004.
- [3]. Belal Chowdhury and Rajiv Khosla, "RFID-based Hospital Real-time Patient Management System," ICIS, pp.363-368, In proceedings of 6th IEEE/ACIS International Conference on Computer and Information Science (ICIS 2007), 2007.
- [4]. J. Fisher and T. Monahan, "Tracking the social dimensions of RFID systems in hospitals", In Proceedings of International Journal of Medical Informatics, Vol. 77, Issue 3, pp. 176-183, 2007.
- [5] S. Shepard, "RFID Radio Frequency Identification", *The McGraw-Hall Companies, Inc.* USA, 2005.
- [6] Agusti Solanas\* and Jordi Castellà-Roca "RFID technology for the health care sector" *CRISES Research Group, UNESCO Chair in Data Privacy, Dept. Computer Science and Mathematics, Rovira i Virgili University Tarragona, Catalonia, Spain*
- [7] O. Shoewu and O. Badejo, "Radio Frequency Identification Technology: Development, Application, and Security Issues", *The Pacific Journal of Science and Technology*, Vol.7, No.2, November 2006.
- [8] Ge Yu, " bspace: A Data Cleaning approach for RFID data streams based on virtual spatial

granularity , 2009 International conference on Hybrid intelligent systems.

[9] Lingyong meng, Fengqi Yu “RFID data cleaning based on Adaptive window” Vol :1 ,2010 IEEE

**First Author** A.Anny Leema completed MCA., MPhil., in Computer Science and working as an Assistant Professor in the dept of Computer Applications, B.S.Abdur Rahman University. She has ten years of experience in teaching. She has presented several papers in National conferences and two in International Conference. Her area of interest is Data Mining, Cybercrime and E-learning. She is currently doing her research in the area Data Warehousing under the guidance of Dr. Hemalatha., Assistant Professor and Head, dept of software systems in Karpagam University

**Second Author** Dr. Hemaltha completed MCA, MPhil., PhD in Computer Science and currently working as a Asst Professor and Head, dept of software systems in Karpagam University. She has ten years of experience in teaching. Published twenty seven papers in International journals and presented several papers in National and international conference. .Area of research is Data mining, Software Engineering, bioinformatics and Neural Network.

---

# Image Compression Using Wavelet Transform Based on the Lifting Scheme and its Implementation

A.Alice Blessie<sup>1</sup>, J. Nalini<sup>2</sup> and S.C.Ramesh<sup>3</sup>

<sup>1</sup> Applied Electronics, Anna University, PSN College of Engineering & Technology,  
Tirunelveli, Tamilnadu , India

<sup>2</sup> Electronics And Communication , Anna University , PSN College of Engineering & Technology,  
Tirunelveli, Tamilnadu, India

<sup>3</sup> Avionics, Anna University , PSN College of Engineering & Technology,  
Tirunelveli, Tamilnadu, India

## Abstract

This paper presents image compression using 9/7 wavelet transform based on the lifting scheme. This is simulated using ISE simulator and implemented in FPGA. The 9/7 wavelet transform performs well for the low frequency components. Implementation in FPGA is since because of its partial reconfiguration. The project mainly aims at retrieving the smooth images without any loss. This design may be used for both lossy and lossless compression.

**Keywords:** image compression, wavelet transform, implementation

## 1. Introduction

In many applications, such as image de-noising or compression, transforms are used to obtain a compact representation of the analyzed image. The wavelet transform relies on a set of functions that are translates and dilates of a single "mother" function, and provides sparse representation of a large class of real-world signals and images.

Image compression is minimizing the size in bytes of a graphics file without degrading the quality of the image to an unacceptable level. The reduction in file size allows more images to be stored in a given amount of disk or memory space. It also reduces the time required for images to be sent over the Internet or downloaded from Web pages.

There are several different ways in which image files can be compressed. For Internet use, the two most common compressed graphic image formats are the JPEG format

and the GIF format. The JPEG method is more often used for photographs, while the GIF method is commonly used for line art and other images in which geometric shapes are relatively simple.

Other techniques for image compression include the use of fractals and wavelets. These methods have not gained widespread acceptance for use on the Internet as of this writing. However, both methods offer promise because they offer higher compression ratios than the JPEG or GIF methods for some types of images. Another new method that may in time replace the GIF format is the PNG format.

A text file or program can be compressed without the introduction of errors, but only up to a certain extent. This is called lossless compression. Beyond this point, errors are introduced. In text and program files, it is crucial that compression be lossless because a single error can seriously damage the meaning of a text file, or cause a program not to run. In image compression, a small loss in quality is usually not noticeable. There is no "critical point" up to which compression works perfectly, but beyond which it becomes impossible. When there is some tolerance for loss, the compression factor can be greater than it can when there is no loss tolerance. For this reason, graphic images can be compressed more than text files or programs.

In JPEG also there are some limitations. In order to overcome those limitations ISO has come with new standard , which is based on new technology called the wavelet technology[1].



A field programmable gate array (FPGA) contains a matrix of reconfigurable gate array logic circuitry that, when configured, is connected in a way that creates a hardware implementation of a software application. Increasingly sophisticated tools are enabling embedded control system designers to more quickly create and more easily adapt FPGA-based applications. Unlike processors, FPGAs use dedicated hardware for processing logic and do not have an operating system. Because the processing paths are parallel, different operations do not have to compete for the same processing resources. That means speeds can be very fast, and multiple control loops can run on a single FPGA device at different rates. Also, the reconfigurability of FPGAs can provide designers with almost limitless flexibility. In manufacturing and automation contexts, FPGAs are well-suited for use in robotics and machine tool applications, as well as for fan, pump, compressor and conveyor control[2].

## 2. Proposed Methodology

The smooth variations in images are called the low frequency components where the sharp variations are the high frequency components. The low frequency components forms the base of an image where the high frequency components add upon them to refine the image. Hence the averages or the smooth variations demands more importance than details[3]. Hence performing 9/7 wavelet transform for smooth images gives better results. Lifting scheme is a technique for constructing second generation wavelet transform.

### 2.1 Discrete Wavelet Transform

“Discrete Wavelet Transform”, transforms discrete signal from time domain into time-frequency domain. The transformation product is set of coefficients organized in the way that enables not only spectrum analyses of the signal, but also spectral behavior of the signal in time. This is achieved by decomposing signal, breaking it into two components, each caring information about source signal. Filters from the filter bank used for decomposition come in pairs: low pass and high pass. The filtering is succeeded by down sampling (obtained filtering result is "re-sampled" so that every second coefficient is kept). Low pass filtered signal contains information about slow changing component of the signal, looking very similar to the original signal, only two times shorter in term of number of samples. High pass filtered signal contains information about fast changing component of the signal. In most cases high pass component is not so rich with data offering good property for compression. In some cases, such as audio or video signal, it is possible to discard

some of the samples of the high pass component without noticing any significant changes in signal. Filters from the filter bank are called "wavelets".

The other perspective to the same theory is based on the fact that some signals, such as audio or video signals often carry redundant information. For instance, looking at the digital picture reveals that neighboring pixels often differ very slightly. The idea is to find a mathematical relation that connects neighboring data samples (pixels) and reduces their number. Of course, inverse process is needed to reconstruct the original.

The wavelet transform (WT) has gained widespread acceptance in signal processing and image compression. Because of their inherent multi-resolution nature, wavelet-coding schemes are especially suitable for applications where scalability and tolerable degradation are important. Recently the JPEG committee has released its new image coding standard, JPEG-2000, which has been based upon DWT. Wavelet transform decomposes a signal into a set of basis functions. These basis functions are called wavelets.

Wavelets are obtained from a single prototype wavelet  $\psi(t)$  called mother wavelet by dilations and shifting:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \quad (1)$$

where  $a$  is the scaling parameter and  $b$  is the shifting parameter.

### 2.2 2-D for DWT

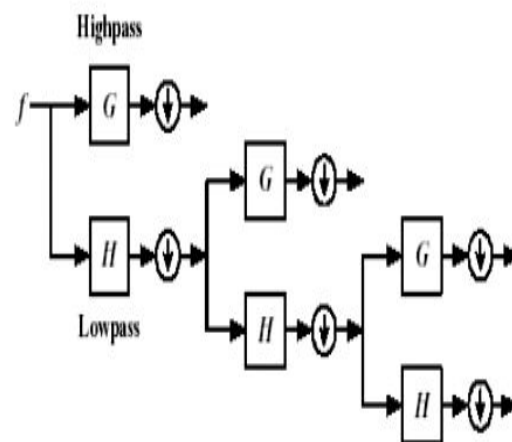


Fig 1. 2-D for discrete wavelet transform

### 2.3 The Lifting Scheme

The wavelet Lifting Scheme is a method for decomposing wavelet transforms into a set of stages. Lifting scheme algorithms have the advantage that they do not require temporary arrays in the calculations steps and have less computations. We use the lifting coefficients to represent the discrete wavelet transform kernel[4].

2.3.1 Three Steps in lifting scheme

a) Split step

It is also called lazy wavelet transform. It divides the input data into odd and even elements.

b) Predict step

This step predicts the odd elements from the even elements.

c) Update step

This replaces the even elements with an average.

3. Block Diagram

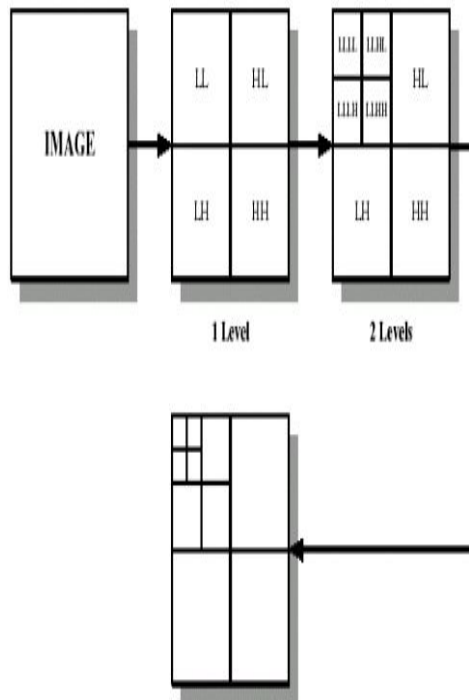


Fig. 2 Block diagram for DWT

3.1 Three stages of waveletting

The 512 by 512 pixel input image frame is processed with three stages of waveletting. In the first stage, 512 pixels of each row are used to compute 256 high pass coefficients (g) and 256 low pass coefficients (ff), figure 3. The coefficients are written back in place of the original row.

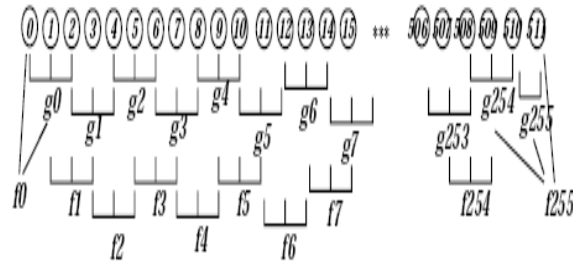


Fig. 3 Waveletting

Once all the 512 rows are processed, the filters are applied in the Y direction. This completes the first stage of waveletting. While conventional Mallot ordering scheme aggregates coefficients into the 4 quadrants, our ordering scheme interleaves the coefficients in the memory. The second stage of wave-letting only processes the low frequency coefficients from the first stage. This corresponds to the upper left hand quadrant in the Mallot scheme. Thus, second stage operates on row and columns of length 256, while the third stage operates on rows and columns of length 128. The aggregation of coefficients along the 3 stages under Mallot ordering is shown in figure4. The memory map with the interleaved ordering is shown in figure 5.

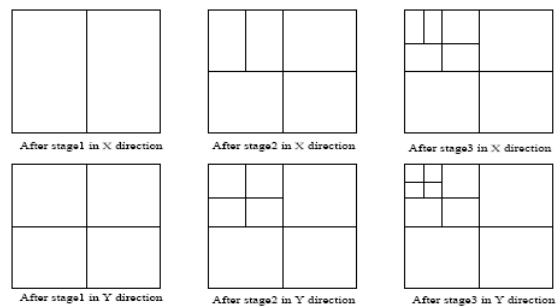


Fig. 4 Mallot Ordering

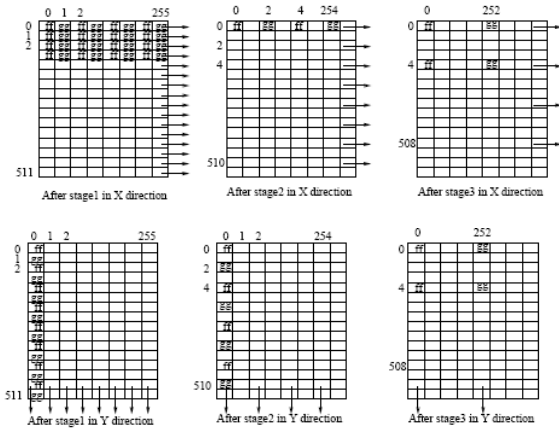


Fig. 5 Stages of waveletting



Fig. 7 The compressed image

#### 4. Results



Fig. 6 The original image

$$MSE = \frac{1}{512 \times 512} \sum_{x=1}^{512} \sum_{y=1}^{512} [p(x, y) - p'(x, y)]^2 \quad (2)$$

$$RMSE = \sqrt{MSE} \quad (3)$$

$$PSNR = 20 \log_{10}(255/RMSE) \quad (4)$$



Fig. 8 Lena image after 3-level of transform

#### 5. Conclusion

Real time signals are both time-limited (or space limited in the case of images) and band-limited. Time-limited signals can be efficiently represented by a basis of block functions (Dirac delta functions for infinitesimal small blocks). But block functions are not band-limited. Band limited signals on the other hand can be efficiently represented by a Fourier basis. But sines and cosines are not time-limited. Wavelets are localized in both time (space) and frequency (scale) domains. Hence it is easy to capture local features in a signal. Another achievement of a wavelet basis is that it supports multi resolution. In the windowed Fourier transform, the effect of the window is to localize the signal being analyzed. Because a single window is used for all frequencies, the resolution of the analysis is same at all

frequencies. To capture signal discontinuities (and spikes), one needs shorter windows, or shorter basis functions. At the same time, to analyze low frequency signal components, one needs longer basis functions. With a wavelet based decomposition, the window sizes vary. Thus it allows to analyze the signal at different resolution levels.

Computationally intensive problems often require a hardware intensive solution. Unlike a microprocessor with a single MAC unit, a hardware implementation achieves greater parallelism, and hence higher throughput. Reconfigurable hardware is best suited for rapid prototyping applications where the lead time for implementation can be critical. It is an ideal development environment, since bugs can be fixed and multiple design iterations can be done, without incurring any non recurring engineering costs. Reconfigurable hardware is also suited for applications with rapidly changing requirements. In effect, the same piece of silicon can be reused. With respect to limitations, achieving good timing/area performance on these FPGAs is much harder, when compared to an ASIC or a custom IC implementation. There are two reasons for this. The first pertains to the fixed size look-up tables. This leads to under utilization of the device. The second reason is that the pre-fabricated routing resources run out fast with higher device utilization.

### Acknowledgments

Thanks be to Holy God, who is a constant strengthener in all the moments of my life. I thank my parents and husband for their dedicative support towards my studies. Thanks to Dr.P.Suyambu, Chairman, PSN Group of Institutions, for His encouragement and motivation for paper publications. Thanks to Dr.C.Gopi, Principal, PSNCET, for His wonderful guidance in publishing our paper. Thanks to all those who have helped me throughout this project .

### References

- [1] Refael c. Gonzalez and Richard E. Woods "Digital Image Processing" Delhi, India:Pearson Education, 2003
- [2]Ayan sengupta, "Compressing still and moving images with wavelets", Multimedia Systems,Vol-2, No-3, 994
- [3]Robbins,"Advantages of FPGAs", Renee Control Engineering [Control Eng.]. Vol. 57, no. 2, pp. 60-62. Feb 2010.
- [4]A.Grzeszczak, M. K. Mandal, S. Panchanathan and T. Yeap, "VLSI Implementation of discrete wavelet transform" IEEE transactions on VLSI systems, vol 4, No 4,pp 421-433, Dec 1996
- [5] SIAM J. Math. Anal, "The lifting scheme: A construction of second generation wavelets," vol. 29, no. 2, pp. 511-546, 1997.

**A.Alice Blessie** has received her Bachelor of Engineering in Electronics and Communication Engineering from Anna University, Chennai in 2007 and now doing her master of engineering under the branch of Applied Electronics in Anna University, Tirunelveli. She is a student of PSN college of Engineering and Technology. She has participated in many National conferences and seminars. Her research interests include Digital Image processing, Neural networks.

**J.Nalini** has received her M.Sc (Electronics) from Bharathidasan University in the year 2005 and she has completed M.E (Communication systems) in Vinayaka Mission University in 2008.She is an Assistant Professor in the Department of Electronics And Communication Systems, PSN College of Engineering and Technology. Her research interests include Digital Image Processing and Wide Area Networks.

**S.C.Ramesh** is an associate professor of Department of Aeronautical Engineering. He has been served on programming committees and organizing committees of various international conferences and symposia. He is a member of IEEE.

# Incorporating Agent Technology for Enhancing the Effectiveness of E-learning System

N. Sivakumar<sup>1</sup>, K. Vivekanandan<sup>2</sup>, B. Arthi<sup>3</sup>, S.Sandhya<sup>4</sup>, Veenas Katta<sup>5</sup>

<sup>1</sup> Pondicherry Engineering college,  
Department of Computer Science and Engineering,  
Pondicherry, India

<sup>2</sup> Pondicherry Engineering college,  
Department of Computer Science and Engineering,  
Pondicherry, India

<sup>3</sup> Pondicherry Engineering college,  
Department of Computer Science and Engineering,  
Pondicherry, India

<sup>4</sup> Pondicherry Engineering college,  
Department of Computer Science and Engineering,  
Pondicherry, India

<sup>5</sup> Pondicherry Engineering college,  
Department of Computer Science and Engineering,  
Pondicherry, India

## Abstract

The advancement in internet and multimedia technologies with years of constant progress in developing software tools to support education have reshaped the way knowledge is delivered allowing E-learning to emerge as a solution to conventional learning methods. It has turned out that the learning process can significantly be improved if the learning content is specifically adapted to individual learners' preferences, learning progress and needs. The complexity of evaluating highly interactive e-learning environment has become an issue that is being addressed by educational developers. The main objective of our paper is to incorporate agent technology to enhance the effectiveness of e-learning system. Software agents have a great potential for supporting learning processes that target and deliver learning materials to learners. A possible way is to use software agents to extract and organize data in intelligent ways. This paper provides conceptualization of the agent based effective e-learning strategies. An agent based feedback oriented e-learning system accompanied by agent based testing for estimation of student's grade; dynamic generation of contents and expert query management system is also proposed. The use of agent

technology in these activities would considerably reduce the human intervention involved in managing e-learning processes.

Keywords: *E-learning, Agent technology, Multi-agent system.*

## 1. Introduction

In today's competitive world, professional training and learning is no longer limited to schools and colleges. A learning environment which focuses on the increasing individual and organizational performance would be more desirable. E-learning goes beyond the paradigm of traditional learning. E-learning refers to the use of Internet technology to deliver a broad array of solution that enhances knowledge and performance [1]. E-learning would not be effective without proper web usability and communication. The current system of e-learning is either domain specific or not completely personalized. This results in the inception of open intelligent e-learning infra-



structures that are more personalized, user friendly and effective means of e-learning.

Researches in the education field show that it is difficult to find a general strategy of teaching when human differences are taken into account. In traditional classroom students are able to interact with each other and their instructor is able to socially construct their knowledge. In technology based learning, this social aspect of learning is significantly reduced. The e-learning interaction is a one-on-one relationship between the student and the instructional content. This problem could be overcome by the usage of a recent technological advancement which is the development of agent based software. An agent based e-learning offers potential solution regarding the problems in conventional learning. An agent can be used in e-learning applications in different contexts. The various agent properties like autonomy, proactive and reactive behaviors, capability to co-operate and communicate with other agents makes it ideal for use in e-learning applications.

An agent in e-learning application is situated in the learning environment and performs the pedagogical tasks autonomously. Agent based intelligent system (ABIS) have proved their worth in multiple ways in education. ABIS goes far beyond conventional training records management and reporting. Learner's self-service, learning workflow, provisions of online-learning, collaborative learning and training resource management are some of the features of ABIS. They are basically used for content management and data persistence [2]. As enrichment over the ABIS, we propose to use agents for various other activities in the system like providing feedback to the educational analyst and e-learning administrator on the quality of the tutorial, offering self rating system for the e-learner, efficient dynamic contents viewing and maintaining updated query answering system. This would help to explore better the agent's property in an e-learning environment and reduce the overhead of human intervention providing an intelligent e-learning system for the end user.

## 2. Related Works

There are numerous researches happening in the field of software agents which has given rise to ideas in sophisticating E-learning. We present here some of the related works done by different research scholars in the areas of agent based e-learning, agent based architecture for distance learning, etc. This chapter helps us to identify the areas in which improvement can be enacted in the existing e-learning system.

In [1], a research note that provides a general introduction on e-learning has been discussed. This paper examines the links between knowledge management and content management and discusses in detail about the various tools necessary for knowledge management and content management. It also dealt in detail about the advantages of e-learning system and presents a consolidated six steps guide towards implementing e-learning. Agent based intelligent system have proved their worth in multiple ways. [2] introduced the application of an agent based intelligent system for enhancing e-learning. This paper reports on the conceptual structure evolved to define development process for pedagogical agents. An agent based e-learning environment where users interact collectively and intelligently with the environment is discussed in [3]. This paper proposes the employment of an agent based approach where agents are a natural metaphor of human acts and the learning systems are generally complex.

An agent-oriented software engineering methodology tropos is proposed for an e-learning system which incorporates various agents and gives a coarse grained analysis for the e-learning system [4]. The base agent model is enriched by the beliefs, goals and plans making the e-learning system more intelligent and flexible. [5] Proposed a multi-agent system for an e-learning system which consists of heterogeneous types of functional agents that executes few functionalities of the distance learning autonomously. Activities like perception, modeling, planning, coordination and task or plan execution are suggested in this paper. A theoretical consideration of a real multi-agent system along with performance comparison is proposed in [6]. This paper aims at full personalization of the e-learning process through an agent based e-learning system. In this paper agent-specific techniques are mainly used for estimation knowledge absorption, adjusting tasks to be suitable for an individual and optimization a whole performance of gaining knowledge to be optimal for each student.

[7] Illustrates advantages of customization of appropriate e-learning resources and fosters collaboration in e-learning environments. This paper proposes intelligent agents in this system would support retrieval of relevant learning materials, support instructional design and analyze data. Agents can be used to generate learning progress reports against predefined goals and can also document learning efficiency. [8] Investigates how e-learning applications are designed and how software systems improve their performance. It lists several educational perspectives that have been implemented and the nine distinctive stages of implementation. It also proposes better software simulation for social interactions and better performance of

applications by introducing a conjunction of static and dynamic profiling mechanisms.

The use of web mining techniques to build an agent that could recommend online learning activities is been discussed in [9]. Data mining techniques are used to extract hidden patterns from web logs. Association rules are used to train the recommender agents to build a model representing the web page access behavior or associations between on-line learning activities. The involvement of Resource Description Framework (RDF) assertions to describe the resources of e-learning system is discussed in [10]. The RDF assertions can be used to model the relationships between various components of the system and between the participants. RDF properties may be thought of as attributes of resources and correspond to traditional attribute-value pairs. The concrete RDF syntax uses XML constructs.

### 3. Agent Properties and their Capabilities

Agent technology appears to be a promising solution to challenges of modern environment. This appears as a high level of software abstraction and it is a part of artificial intelligence. An agent can be defined as "An encapsulated computer system that is situated in some environment and that is capable of flexible, autonomous action in that environment in order to meet its design objectives." Agent is a process which operates in the background and performs activities when specific events occur [6]. The various properties of agents make them more suitable to environments where human intervention creates a great overhead. Agents are capable of relieving human intervention significantly and help in proper functioning of the system. The various characteristics of agents are:

**Autonomy:** Autonomy corresponds to the independence of a party to act as it pleases. Autonomous agents have control both over their internal state and over their own behaviour.

**Heterogeneity:** Heterogeneity corresponds to the independence of the designer of a component to construct the component in any manner.

**Proactive:** A proactive agent is one that can act without any external prompts. It acts in anticipation of the future goals.

**Reactive:** the agent responds based on the input it received and according to the environment. It responds in timely fashion to the environmental change.

**Communication:** It can be defined as those interactions that preserve the autonomy of the parties concerned.

**Dynamism:** the agents are dynamic as their reaction is dynamic and varies according to the environment.

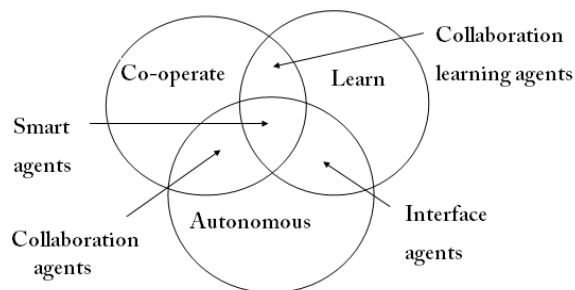


Fig 1. Representations of Agent Properties

These characteristics of agents when utilized efficiently at the correct place and time, would simplify the complications of many existing systems. Agents increase the robustness of the software by mining data to extract hidden patterns. The intelligent agent helps to obtain optimized result from data preserved in web page library [2]. Agents in e-learning system perform various tasks at various levels. Agent helps to model the user behavior, by passing interests and specifications of the user [6].

Agents hide the complexity of different tasks and monitor events and procedures. The agent's properties make them ideal for E-learning applications [7]. Agents apart from data mining, knowledge management, selecting tutorials for the user, they also help in collaboration of the system. User based agents significantly help reduce the administration duties of the course and focus on response to user's questions or prepare training materials. The agents have been used in many areas of e-learning system at present. Yet, there remains a myriad of contexts where agents can be incorporated to make e-learning more efficient and fundamentally change the way education is being delivered. In this following section of the paper, we discuss how agents can be incorporated for various activities in e-learning system and how they can be better utilized in a system.

### 4. Proposed Work

In this section we propose an e-learning system based on the concept of agent oriented software. The following agents can be utilized in an e-learning environment to make the e-learning system efficient.

#### 4.1 Personalization Agent

The perceiving capacity and the knowledge possessed vary from one person to another. In a static e-learning environment the tutorials or the resources do not vary and are not based on the capacity of the e-learner. For the user to understand the concepts clearly the learning resources should be interactive, responsive and engaging with

knowledge formation emphasized. The personalization agent used in an e-learning system would help the user to rank themselves. Based on their ranking, the agent selects learning materials and retrieves it based on cognitive style, personal preferences and prior knowledge. The agent uses a number of techniques and characteristics to filter retrieve and categorize documents according to user's predefined criteria. The personalization agent to a great extent helps the user to save time by personalizing the available resources and tutorials based on the user's self evaluation.

## 4.2 Evaluation Agent

The evaluation agent plays a crucial role in the system by evaluating the student's performance after the tutorial

session. It not only lets the user know where he/she stands but also offers direct and indirect feedback on the efficiency of the tutorial to the tutor. The problems to be generated dynamically for the user evaluation tests are stored in a questionnaire database. The agent determines the learner's level of understanding from the problem statement and the learner's answers. The user's score, difficulty level attempted, duration taken to answer the questions and the topics in which the test was taken are all stored in a database for further analysis in the future by the e-learning instructor.

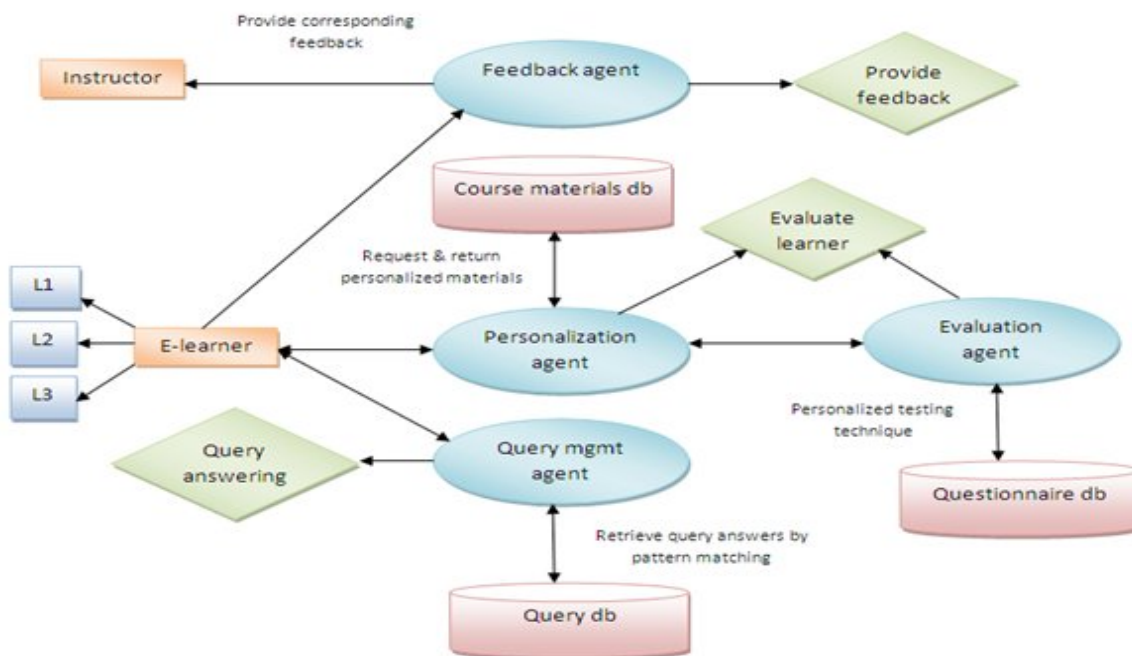


Fig 2. Agent Overview Diagram in Agent Based E-learning System

## 4.3 Query Management Agent

Query management system is very important for a learning environment. Queries and response to queries help the user understand topics clearly. The response to queries should be reliable, fast, clear and satisfactory to the e-learner. The major goal of this agent is to propose links and information that are considered relevant to the user's search. A discussion forum can be treated as a query management system. However in some cases a particular question/query raised by one of the learners may have been answered directly or some related answers may have already been present in some other context or in some other discussion

forum. The query management agent undertakes the responsibility of detecting and avoiding redundant questions posted. This is done by using pattern matching algorithm for texts and mining techniques. The query management agent helps in intelligent search to obtain optimized search results from the data already preserved. The agent deployed automatically searches for information relevant to a particular search query using domain characteristics. If no response to the query is found then the agent seeks an expert/instructor's advice.

#### 4.4 Feedback Agent

The ultimate goal of a system cannot be achieved without proper feedback. The effectiveness of any system depends greatly on the feedback timing and style. The feedback agent collects the feedback and rating of the tutorials from the user. A reliable feedback from the user would enable to improve the efficiency of the tutor and the quality of the resources used in learning practice. This information would help to determine the usefulness of a material for teaching specific topics and update materials to improve their ranking by interacting with the user.

#### 4.5 Agent Relationship

In multi-agent system (MAS), the interaction and communication between the agents plays a key role. The

agents interact with each other through message passing. The message passing involves processing of incoming messages, decoding, and takes corresponding actions. The interaction of the agents in the system is as shown in Fig.4. The e-learner ranks him based on his knowledge and the personalization agent provides learning materials to the learner based on the criteria. The user makes use of the tutorials and if any doubt arises, the user can report it. The query management system handles the question raised and responses to the query as early as possible in the most efficient way. The agent analyzes the user's performance and generates questionnaire accordingly.

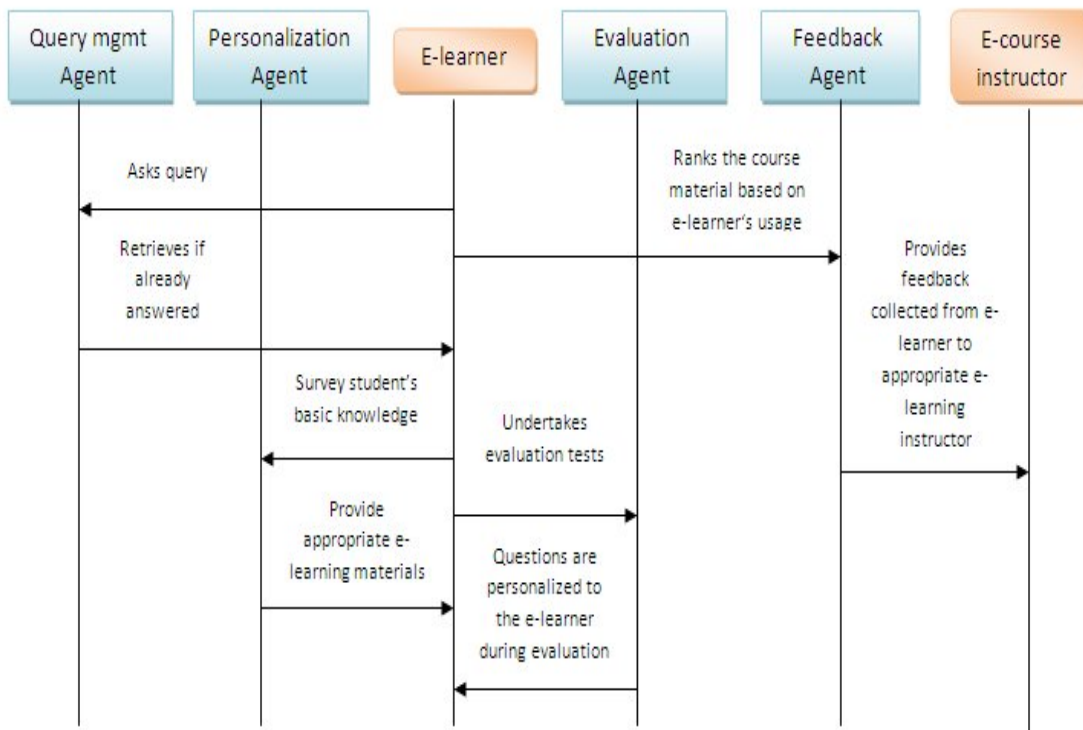


Fig 3. Agent Interaction Diagram.

### 5. Implementation

The agent based e-learning system was implemented using the Java Agent Development Environment (JADE). The agents communicate using the agent communication language (ACL) and Knowledge Query and Manipulation Language (KQML). The e-learning system was developed

for OOPS training. The tutorials concentrate on OOPS concepts and C basics. Let us consider the example of a user learning OOPS. If the user rates him to be well versed with the C concepts then the basics of it need not be dealt much and the agent provides resources on OOPS correspondingly. If the user had rated him to be not much familiar with the C concept then the agent retrieves materials on the basics of C before training the user in OOPS. For evaluating process the questions will be stored

in the questionnaire database. We classify the problems stored in the knowledge base into four difficult levels: easy medium, difficult, very difficult. The evaluation agent determines which difficulty level problem should be generated to the user. When the learner reaches a certain score of about 70% or more then the agent increases the difficulty level for the remaining problems. If the score is less than 40% the agent retrieves easy questions from the knowledge base and also rates the user's understanding on the concepts to be low. If the user consumes longer duration to answer questions on certain topics, there is a possibility that the user is either referring other resources or the users understanding on the particular concept is relatively low. So, more of application oriented questions are retrieved by the agent in the above case, in order to test

the learner's capability. If the learner is able to answer the application oriented questions on a certain topic then the theoretical questions on that topic can be skipped by the agent. Consider an OOPS learning session, where there are four tutorials and four instructors. The agent monitors and collects details on the average number of hits per e-learner for a particular tutorial and the total number of user for that particular learning material as in Table 1. Based on this data, the agent ranks the tutorials and provides feedback to the corresponding e-learning instructor. This is pictorially depicted in a graph through Fig 4. The agent provides this feedback to the instructor of the particular course after considering the average number of hits for the resource and its usage.

Table 1. Survey for feedback agent

<i>Learning material</i>	<i>E-learner instructor</i>	<i>Average no. of hits / E-learner</i>	<i>No. of E-learners used</i>	<i>Rating of E-learning material</i>
OOPS-1	Instructor1	2	50000	1
OOPS-2	Instructor2	4	10000	4
OOPS-3	Instructor3	2	25000	3
OOPS-4	Instructor4	3	40000	2

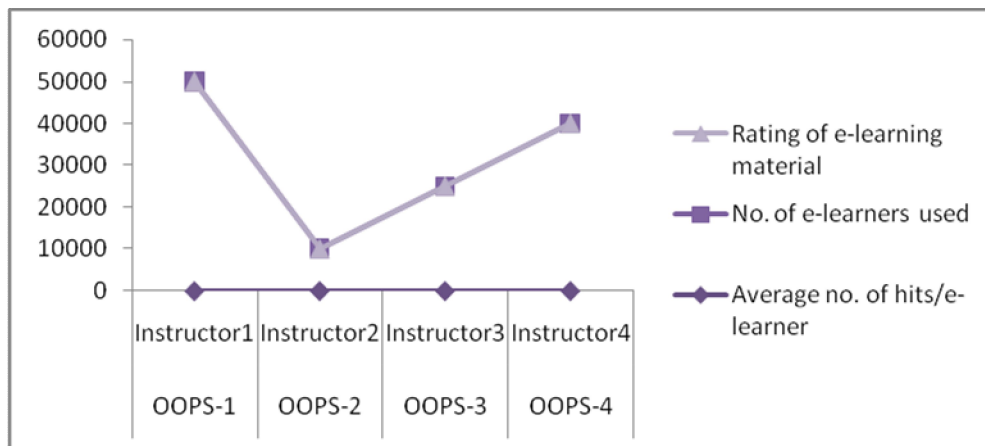


Fig 4. Graph on the Above Survey Giving Feedback

## 6. Result Interpretation

The agent based e-learning has considerably reduced the human intervention in the e-learning environment. The introduction of agents in the e-learning system has to a great extent brought the advantageous

characteristics of conventional teaching methods. The data presented through Table 1, provide details of the agents' monitoring the hit ratio and the average number of learners viewing a tutorial. This data helps to interpret the efficiency of each tutorial. The graph symbolizes which tutorial is more efficient. The highest point on the graph correspond to most effective tutorial, the tutorial with lowest number of



repeated hits by the same learner and which has maximum number of users. The agent introduction in the system has significantly enhanced the effectiveness of the system and reduced human intervention.

## 7. Conclusion and Future Works

Learning systems are designed to support learners and provide improved learning outcomes. In constantly changing world, there is a significant need to update existing materials in order to select the most appropriate tutorial. This paper describes the combination of computational intelligence of E-learning system and properties of intelligent agents. A set of E-learning agents that are capable of personalizing resources based on learner's potential, evaluating the student's performance, offering feedback to the tutor and reliable query-response system would improve the efficiency of e-learning environment.

Future work can be directed towards the introduction of newly developed evolutionary algorithms and introduction of more agents within the e-learning environment in order to enhance the training functionalities of the system. Efforts may be put in order to enhance the e-learning environment and introduce special features to the system.

## References

- [1] Namahn, "E-learning, a Research note".
- [2] S.Prakasam,Dr.R.M.Suresh, "An Agent-Based Intelligent System to Enhance E-Learning Through Mining Techniques", International Journal on Computer Society and Engineering-2010.
- [3] Khadida Harbouche, Mahiede Djoudi, "Agent-Based Design for E-learning Environment", Journal of Computer Science-2007, Science Publications.
- [4] Zhi Liu, Bo Chen, "Model and Implement an Agent Oriented E-learning System", IEEE Computer Society-2005.
- [5] Safiye Turgay, "A Multi-Agent System approach for Distance Learning Architecture", Turkish Online Journal of Education Technology-October 2005.
- [6] Markek Woda, Piotr Michalec, "Distance Learning System: Multi-Agent Approach", Journal of Digital Information Management-September 2005.
- [7] Dawn G.Gregg,"E-learning Agents", the Learning Organization vol. 14-2007.
- [8] Dorian Stoilescu, "Modalities of using Learning Objects for Intelligent Agents in E-learning", Interdisciplinary Journal of E-learning and Learning Objects-2008.
- [9] Osmar.R.Zaiane, "Building a recommender agents for e-learning systems", IEEE Computer Society-2002.
- [10] Sabin-Corneliu Buraga, "Agent-Oriented E-learning Systems", International Conference on Control Systems And Computer Science-2003.

# Linear Network Coding on Multi-Mesh of Trees (MMT) using All to All Broadcast (AAB)

Nitin Rakesh<sup>1</sup> and VipinTyagi<sup>2</sup>

<sup>1,2</sup>Department of CSE & IT, Jaypee University of Information Technology, Solan, H.P. 173215, India

## Abstract

We introduce linear network coding on parallel architecture for multi-source finite acyclic network. In this problem, different messages in diverse time periods are broadcast and every non-source node in the network decodes and encodes the message based on further communication. We wish to minimize the communication steps and time complexity involved in transfer of data from node-to-node during parallel communication. We have used Multi-Mesh of Trees (MMT) topology for implementing network coding. To envisage our result, we use all-to-all broadcast as communication algorithm.

**Keywords:** Coding, information rate, broadcasting.

## 1. Introduction

Shuo-Yen *et al.* [1] prove constructively that by linear coding alone, the rate at which a message reaches each node can achieve the individual max-flow bound. Also, provide realization of transmission scheme and practically construct linear coding approaches for both cyclic and acyclic networks. [1–6] shows that network coding is necessity to multicast two bits per unit time from a source to destinations. It also showed that the output flow at a given node is obtained as a linear combination of its input flows. The content of any information flowing out of a set of non-source nodes can be derived from the accumulated information that has flown into the set of nodes. Shuo-Yen *et al.* described an approach on network information flow and improved the performance in data broadcasting in all-to-all communication in order to increase the capacity or the throughput of the network.

[7] selected the linear coefficients in a finite field of opportune size in a random way. In this paper packetizing and buffering are explained in terms of encoding vector and buffering the received packets. It showed that each node sends packets obtained as a random linear combination of packets stored in its buffer and each node receives packets which are linear combinations of source packets and it stores them into a matrix. While Widmer *et al.* [8] gave an approach with energy efficient broadcasting in network coding. Subsequent work by Fragouli *et al.* [9] gave two heuristics and stated that each node in the graph

is associated with a forwarding factor. A source node transmits its source symbols (or packets) with some parameters bounded by this forwarding factor. And when a node receives an innovative symbol, it broadcast a linear combination over the span of the received coding vector. [10] deals with network coding of a single message over an acyclic network. Network coding over undirected networks was introduced by [11] and this work was followed by [12], [13]. The network codes that involve only linear mapping with a combination of global and local encoding mapping involves linear error-correction code [14], [15], [16] and [17] have also been presented.

We present an approach for parallel network in which network coding is employed to perform communication amid the nodes. The association among network coding and communication algorithm establishes a more efficient way to transfer data among the nodes. We consider parallel multi-source multicast framework with correlated sources on MMT architecture [18]. We use a randomized approach in which, other than the receiving nodes all nodes perform random linear mapping from inputs on outputs (see figure 1). In each incoming transmissions from source node, the destination node has knowledge of overall linear combination of that data set from source. This information is updated at each coding node, by applying the same linear mapping to the coefficient vectors as applied to the information signals. As an example, assume that in a directed parallel network ( $\check{N}$ ) the source node  $P_1$  (unique node, without any incoming at that instant of time) sends a set of two bits ( $\vec{d}_1, \vec{d}_2$ ) to node  $P_2, P_3$  and  $P_4, P_7$  (figure 1).

Network ( $\check{N}$ ) in figure 1 is used to show information multicast with network coding at each node. Node  $P_1$  multicast data set ( $\vec{d}_1, \vec{d}_2$ ) to destination nodes  $P_3$  and  $P_7$ . Any receiving non-destination node, truncates randomly chosen coefficient of finite fields with the received data before transferring to other nodes. The compressive transmission throughout network ( $\check{N}$ ) is heralded in following steps of Table 1.

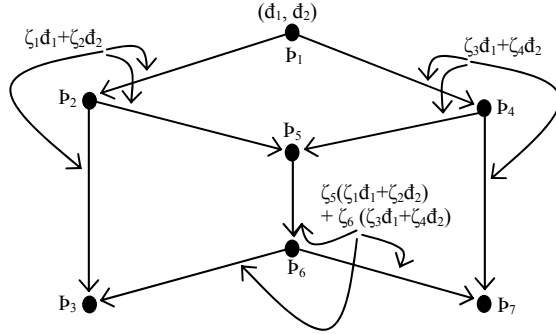


Fig. 1.A network ( $\tilde{N}$ ) used, as an example, to explain LNC with coefficient added at each data transfer from different nodes (the network has seven nodes  $P_1, P_2, \dots, P_7$  and nine edges  $P_1 P_2, P_1 P_4, P_2 P_5, P_4 P_5, P_2 P_3, P_5 P_6, P_4 P_3, P_6 P_3, P_6 P_7$  directed in this order).  $(d_1, d_2)$  is the set of data being multicast to destinations, and coefficients  $\zeta_1, \zeta_2, \dots, \zeta_6$  are randomly chosen elements of a finite field. Each link represents the data transmission.

Table 1: Compressive transmission in Network ( $\tilde{N}$ ).

S. No	Source Node	Destination Node	Coefficient Clubbed	Data to send further
1	$P_1$	$P_2$ and $P_4$	$(\zeta_1, \zeta_2$ and $\zeta_3, \zeta_4)$	$\zeta_1 d_1 + \zeta_2 d_2$ and $\zeta_3 d_1 + \zeta_4 d_2$
2	$P_2, P_4$	$P_5$	$(\zeta_5$ and $\zeta_6)$	$\zeta_5(\zeta_1 d_1 + \zeta_2 d_2) + \zeta_6(\zeta_3 d_1 + \zeta_4 d_2)$
3	$P_5$	$P_6$		$\zeta_5(\zeta_1 d_1 + \zeta_2 d_2) + \zeta_6(\zeta_3 d_1 + \zeta_4 d_2)$
4	$P_6$	$P_3, P_7$		$(d_1, d_2)$

This approach indicates that using network coding efficient multicast of diverse data in a network is possible. It is also right to say that the flow of information and flow of physical commodities are two different things [1]. So, the coding of information does not increase the information content. The capacity of a network to transmit information from the source to destination can become higher if the coding scheme becomes wider but it is limited to max-flow for a very wide coding scheme [1].

Now, as parallel networks contribute in data communication within several nodes in parallel, it is required to have higher capacity of such networks to transmit information. In sense to increase capacity of data communication in parallel networks we have implemented network coding on parallel architecture (MMT). To examine the performance of this network with network coding we have implemented this approach with existing All-to-All Broadcast algorithm (AAB) [19] on this architecture. In consecutive sections we have also shown that this approach has reduced the chance of error and increased the capacity of network to transmit data between nodes. For parallel transmission of information linearity of coding makes encoding (at source) and decoding (at receiving end) easy to contrivance. We do not address the problem which may or may not occur because of this approach, but have identified the possibilities of errors after implementation.

The remaining paper is organized in five sections. In section second our basic model and preliminaries are illustrated. In basic model, notions used for *linear-code multicast in parallel architecture* (LCM-PA) and definitions are explained. In section third, implementation of AAB on parallel network-MMT is explained. In section fourth, LNC is implemented using LCM-PA, it is illustrated using AAB algorithm on MMT [19]. The fifth section is used for results and simulations. In section sixth, we are concluding this paper and future scope of LCM-PA is given in this section. The basic definition, theorems and lemma used in this paper are explained in appendix.

## 2. Model and Preliminaries

A parallel network is represented as a directed graph  $G(V, E)$ , where  $V$  is the set of nodes in network and  $E$  is the set of links, such that, from node  $i$  to  $j$  for all  $(i, j) \in E$ , where node  $i$  and  $j$  are called the *origin* and *destination*, respectively, of link  $(i, j)$ , information can be sent noiselessly. Each link  $l \in E$  is associated with a nonnegative real number  $c_l$  representing its transmission capacity in bits per unit time. The origin and destination of a link  $l \in E$  are denoted as  $o(l)$  and  $d(l)$ , respectively, where  $o(l) \neq d(l) \forall l \in E$  is obtained as a coding function of information received at  $o(l)$ .

There are  $r$  discrete memoryless information source processes  $X_1, X_2, \dots, X_r$ , which are random binary sequences. The processors may change according to the parallel architecture. We denote the Slepian-Wolf region of the sources

$$\mathcal{R}_{SW} = \{(R_1, R_2, \dots, R_r) : \sum_{i \in S} R_i > H(X_S | X_{S^c}) \forall S \subseteq \{1, 2, \dots, r\}\}$$

Where,  $X_S = (X_{i_1}, X_{i_2}, \dots, X_{i_{|S|}})$ ,  $i_k \in S, k = 1 \dots |S|$ . Source process  $X_i$  is generated at node  $a(i)$ , and multi-cast to all nodes  $j \in b(i)$ , where  $a : \{1, \dots, r\} \rightarrow V$  and  $b : \{1, \dots, r\} \rightarrow 2^V$  are arbitrary mappings. For parallel architectures, we have considered the same approach to implement the network coding. In this paper, we consider the (multisource) multicast case where  $b(i) = \{\beta_1, \dots, \beta_d\}$  for all  $i \in [1, r]$ . The node  $a(1), \dots, a(r)$  are called *source nodes* and the  $\beta_1, \dots, \beta_d$  are called receiver nodes, or receivers. For simplicity, we assume subsequently that  $a(i) \neq \beta_j \forall i \in [1, r], j \in [1, d]$ . For data communication at different step, the source and destination nodes changes according to the flow of data in the algorithm. If the receiving node is able to encode the complete source information, than connection requirements are fulfilled. For parallel communication, we have used these sets of connection requirements for level of communication, which is encoded at each level (step of algorithm). To specify a multicast connection problem we used 1) a graph  $G(V, E)$ , 2) a set of multicast connection requirements, and 3) a set

of link capacity  $\{c_l \mid l \in E\}$ . To explain *linear-code multicast* (LCM) with parallel communication network, we present some terminologies, definitions and assumptions.

*Conventions:* 1) In MMT network, the edges e.g.,  $(P_i, P_j) \in (E)$  denotes that  $(P_i, P_j)$  is a bi-directed edge [18], but this edge may act as unidirectional depending on the algorithm. 2) The information unit is taken as a symbol in the base field, i.e., 1 symbol in the base field can be transmitted on a channel every unit time [1].

*Definitions:* 1) The communication in MMT network is interblock and intrablock [18]. LNC is implemented in blocks first and then in complete network. 2) A LCM on a communication network  $(G, o(l))$  is an assignment of vector space  $v(P_i)$  to every node  $P_i$  and a vector  $v(P_i, P_j)$  to every edge  $P_i, P_j$  [1] such that

$$v(o(l)) = \Omega;$$

$$v(P_i, P_j) \in v(P_i) \text{ for every edge } P_i, P_j;$$

for any collection  $d(l)$  of nonsource nodes in the network  $\langle \{v(P_i) : P_i \in d(l)\} \rangle = \langle \{v(P_i, P_j) : P_i \notin d(l), P_j \in d(l)\} \rangle$

*Assumptions:* 1) Each source process  $X_i$  has one bit per unit time entropy rate for independent source process, while larger rate sources are modeled as multiple sources.

2) Modeling of sources as linear combinations of independent source processes for linearly correlated sources. 3) Links with  $l \in E$  is supposed having a capacity  $c_l$  of one bit per unit time for both independent as well as linear correlated sources. 4) Both cyclic (networks with link delays because of information buffering at intermediate nodes; operated in a batched [2] fashion, burst [11], or pipelined [12]) and acyclic networks (networks whose nodes are delay-free i.e. zero-delay) are considered for implementation of LNC on parallel networks, by analyzing parallel network to be a cyclic or acyclic. 5) We are repeatedly using either of these terms processor and nodes, throughout the paper, which signify common significance.

The network may be analyzed to be acyclic or cyclic using scalar algebraic network coding framework [13]. Let us consider the zero-delay case first, by representing the equation  $Y_j = \sum_{\{i:a(i)=o(j)\}} a_{i,j} X_i + \sum_{\{l:d(l)=o(j)\}} f_{l,j} Y_l$ . The sequence of length- $u$  blocks or vectors of bits, which are treated as elements of a finite field  $F_q$ ,  $q = 2^u$ . The information process  $Y_j$  transmission on a link  $j$  is formed as a linear combination, in  $F_q$ , of link  $j$ 's inputs, i.e., source processes  $X_i$  for which  $a(i) = o(j)$  and random processes  $Y_l$  for which  $d(l) = o(j)$ . The  $i$ th output process  $Z_{\beta,i}$  at receiver node  $\beta$  is a linear combination of the information processes on its terminal links, represented as  $Z_{\beta,i} = \sum_{\{l:d(l)=\beta\}} b_{\beta,i,l} Y_l$ .

Memory is needed, for link delays on network for multicast, at receiver (or source) nodes, but a memoryless operation suffices at all other nodes [12]. The linear coding equation for unit delay links (considered) are

$$Y_j(t+1) = \sum_{\{i:a(i)=o(j)\}} a_{i,j} X_i(t) + \sum_{\{l:d(l)=o(j)\}} f_{l,j} Y_l(t). \\ Z_{\beta,i}(t+1) = \sum_{u=0}^{\mu} \check{A}_{\beta,i}(u) Z_{\beta,i}(t-u) + \sum_{\{l:d(l)=\beta\}} \sum_{u=0}^{\mu} \check{A}_{\beta,i,l}(u) Y_l(t-u)$$

where  $X_i(t)$ ,  $Y_j(t)$ ,  $Z_{\beta,i}(t)$ ,  $\check{A}_{\beta,i}(t)$ , and  $\check{A}_{\beta,i,l}(t)$  are the values of variables at  $t$  time and  $\mu$  represents the required memory. In terms of delay variable  $D$  these equation are as  $Y_j(D) = \sum_{\{i:a(i)=o(j)\}} D a_{i,j} X_i(D) + \sum_{\{l:d(l)=o(j)\}} D f_{l,j} Y_l(D)$ .

$$Z_{\beta,i}(D) = \sum_{\{l:d(l)=\beta\}} b_{\beta,i,l}(D) Y_l(D).$$

where

$$b_{\beta,i,l}(D) = \frac{\sum_{u=0}^{\mu} D^{u+1} \check{A}_{\beta,i,l}(u)}{1 - \sum_{u=0}^{\mu} D^{u+1} \sum_{u=0}^{\mu} \check{A}_{\beta,i}(u)}$$

$$\text{and } X_i(D) = \sum_{t=0}^{\infty} X_i(t) D^t$$

$$Y_j(D) = \sum_{t=0}^{\infty} Y_j(t) D^t, \quad Y_j(0) = 0$$

$$Z_{\beta,i}(D) = \sum_{t=0}^{\infty} Z_{\beta,i}(D)(t) D^t, \quad Z_{\beta,i}(0) = 0$$

The above given coefficients can be collected into  $r \times |E|$  matrices. These coefficients can be used from the transmission in parallel network. These matrices will be formed for both cyclic and acyclic cases.

$$A = \begin{cases} (a_{i,j}) \text{ in the acyclic delay-free case} \\ (D a_{i,j}) \text{ in the cyclic case with delay} \end{cases}$$

And  $B = (b_{\beta,i,l})$ , and the matrix  $|E| \times |E|$

$$F = \begin{cases} (f_{l,j}) \text{ in the acyclic delay-free case} \\ (D f_{l,j}) \text{ in the cyclic case with delay} \end{cases}$$

Now, let us consider an example of parallel network ( $\check{N}$ ) (MMT), in which processor  $P_i$  (unique processor, without any incoming at that instant of time) to node  $P_2$  and  $P_3$ , sends two bits,  $(\check{d}_1, \check{d}_2)$  as given in figure 2.

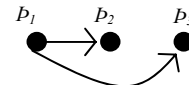


Fig. 2. A row of a block of MMT with  $n = 3$ , where  $n$  is the number of processors in MMT architecture. The detailed MMT architecture is given in figure 3 for more simplicity to the readers.

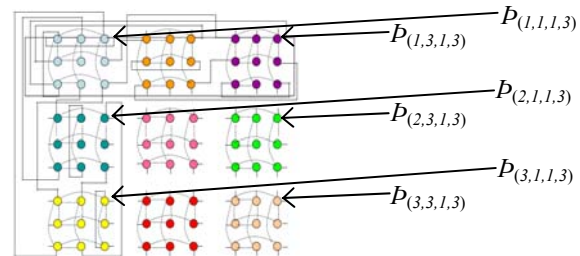


Fig. 3.  $3 \times 3$  Multi-Mesh of Trees (MMT) ( $\check{N}$ ). (All interblock links are not shown. The  $P_{(1,3,1,3)}$ ,  $P_{(2,3,1,3)}$ ,  $P_{(3,3,1,3)}$  are the processor index value which is used to identify individual processors through-out the architecture).

This linear coding is achieved with the LCM-PA ( $\psi$ , used to replace LCM-PA further for equations) is specified by

$$\psi(P_1, P_2) = \psi(P_1, P_3) \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

The matrix product of  $(\vec{d}_1 \vec{d}_2)$  with the column vector assigned by  $\psi$  is the data sent in a row of MMT. Further, for  $n$  number of processors the data received by other processors will be  $\vec{d}_1 + \vec{d}_2$ , where vector  $\vec{d}_1 + \vec{d}_2$  reduce to the exclusive-OR  $\vec{d}_1 \oplus \vec{d}_2$ . Also, for every  $\psi$  on a network, for all nodes  $P_3$  (which is the receiving processor) [1]

$$\dim(v(P_3) \leq \text{maxflow}(P_3).$$

This shows that  $\text{maxflow}(P_3)$  is an upper bound on the amount of information received at  $P_3$  when a LCM  $\psi$  is used [1].

### 3. Implementation of AAB on Parallel Network

In this section, we implement AAB on parallel network (MMT). For implementation, we are using AAB algorithm, which involves ten steps to completely transfer and receive information of all processors to all processors in MMT [19] and implement LNC using LCM-PA model in next section. We consider the MMT network with  $n = 8$ , where  $n$  is number of processors and in algorithm and we consider  $N = n^2 \times n^2$ ,  $\forall n \in \mathbb{N}$  and a *block* =  $n \times n = \text{row} \times \text{column}$ . The time taken to transfer and receive all information at each step of algorithm is listed in [19] involved in AAB algorithm.

For implementation of AAB on MMT network, first we state a reason for using this network. MMT network is better than other traditional parallel networks (we compared few of them e.g., Multi-Mesh (MM) [20, 21]) based on the topological properties of MMT, which is comparable regarding efficiency parameters. A comparison of these networks, based on some parameters, is given in figure 4 and a comparison between 2D Sort on MM and MMT for different values of processor is given in figure 5.

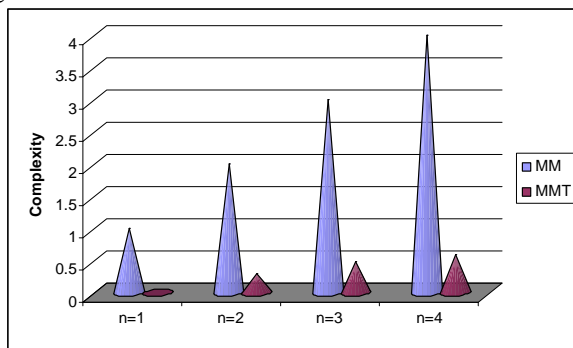


Fig. 4. Comparison of MMT and MM on the basis of Communication links, Solution of Polynomial Equations, One to All and Row & Column Broadcast.

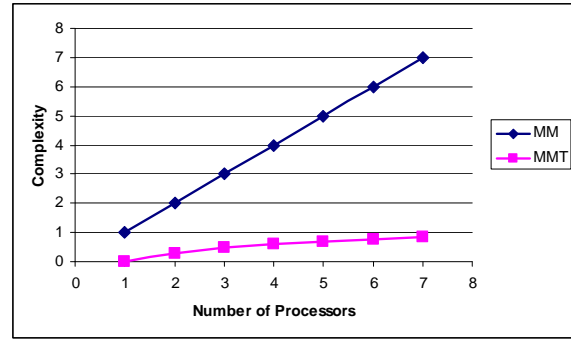


Fig. 5. A comparison between 2D Sort on MM and MMT for different values of processor.

Table 2 [18, 22] shows characteristics of various processor organizations based on some of the network optimization parameters. From all these network architectures MMT is more optimum network to be used.

Table 2: Characteristics of Various Processor Organizations.

Network	Nodes	Diameter	Bisection Width	Constant Number of Edges	Constant Edge Length
1-D mesh	$k$	$k-1$	1	Yes	Yes
2-D mesh	$k^2$	$2(k-1)$	$k$	Yes	Yes
3-D mesh	$k^3$	$3(k-1)$	$k^2$	Yes	Yes
Binary tree	$2^k - 1$	$2(k-1)$	1	Yes	No
4-ary hypertree	$2^k(2^k - 1)$	$2k$	$2^{k+1}$	Yes	No
Pyramid	$4k^2 - 1)/3$	$2\log k$	$2k$	Yes	No
Butterfly	$(k+1)2^k$	$2k$	$2^k$	Yes	No
Hypercube	$2^k$	$k$	$2^{k-1}$	No	No
Cube-connected cycles	$k2^k$	$2k$	$2^{k-1}$	Yes	No
Shuffle-exchange	$2^k$	$2k-1$	$\geq 2^{k-1}/k$	Yes	No
De Bruijn	$2^k$	$k$	$2^k/k$	Yes	No
MMT	$k^4$	$4\log k + 2$	$2(k-1)$	Yes	No
MM	$k^4$	$2k$	$2(k-1)$	No	No

Now, to demonstrate the algorithm, we consider  $N = 8^2 \times 8^2 = 4096$  nodes, as the size of network, where each *block* consists of  $8 \times 8$  i.e., *row*  $\times$  *column*. For clarity in explaining each step of algorithm, we have used either one row or one column, based on the algorithm, to show the flow of data in each step. For every step the data flow varies, so for each step different algorithms are used. Figure 6 shows first row in first block of the network and the connectivity between the processors is based on the topological properties of MMT [18]. We have considered that each processor is having a Working Array ( $WA$ ) which consist of the processor index ( $P_n$ ) and information associated with that processor ( $I_n$ ). The size of working array is based on the size of network used, i.e. for  $n = 8$ , the size of  $WA = 8$ .



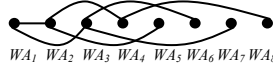


Fig. 6. Shows initial condition of processors containing WA(only one row of a block of  $8 \times 8$  MMT is shown)

The figure 7 (a) shows the position of data after completion of step 1 and figure 7 (b) shows the content of  $WA_1$  after step 1.



Fig. 7. (a) After Step 1

$P_1$	$P_2$	$P_3$	$P_4$	$P_5$	$P_6$	$P_7$	$P_8$
$I_1$	$I_2$	$I_3$	$I_4$	$I_5$	$I_6$	$I_7$	$I_8$

(b) Content of  $WA_1$  after Step 1

### Algorithm 1. Step 1 of AAB

- a. /\* This operation is common between all processors of each row of each block,
- b. Each node is represented by  $n_{(\alpha,\beta,i,j)}$ ; where  $\alpha, \beta$  are the block index and  $i, j$  are node index (see figure M)
- c. The transfer is conducted in order  $\frac{n_{(\alpha,\beta,i,j)}}{2 \times \text{Count of Iteration} - 1} < j \leq \frac{n_{(\alpha,\beta,i,j)}}{2 \times \text{Count of Iteration} - 1} * /$

1: Starting from each row of each block of network, such that the processor with greater index value will transfer data to lower index processors linked according to the topological properties of network.

2: repeat

3: Select nodes  $n_{(\alpha,\beta,i,\frac{N}{2}+1)}, n_{(\alpha,\beta,i,\frac{N}{2}+2)}, n_{(\alpha,\beta,i,\frac{N}{2}+3)}, \dots, n_{(\alpha,\beta,i,N)} \in N$  from each block of network such that at each transfer the block is divided in two parts (e.g. if  $N = 40$ , number of nodes in blocks will also be 40 and division will be 1 to 20 and 21 to 40<sup>th</sup> index position) and transfer message to remaining nodes  $n_{(\alpha,\beta,i,1)}, n_{(\alpha,\beta,i,2)}, n_{(\alpha,\beta,i,3)}, \dots, n_{(\alpha,\beta,i,\frac{N}{2})} \in N$  linked according to topological properties of this network.

Note: The message will be transferred from higher processor index to lower.

4: Select nodes  $n_{(\alpha,\beta,i,1)}, n_{(\alpha,\beta,i,2)}, n_{(\alpha,\beta,i,3)}, \dots, n_{(\alpha,\beta,i,\frac{N}{2})} \in N$  (other than the nodes from which message has already transferred) from each block of network such that at each transfer these nodes are divided in two parts (same as in 3; i.e.  $n_{(\alpha,\beta,i,1)}, n_{(\alpha,\beta,i,2)}, \dots, n_{(\alpha,\beta,i,\frac{N}{4})}$ , and  $n_{(\alpha,\beta,i,\frac{N}{4}+1)}, n_{(\alpha,\beta,i,\frac{N}{4}+2)}, \dots, n_{(\alpha,\beta,i,\frac{N}{2})} \in N$ ).

Now  $n_{(\alpha,\beta,i,\frac{N}{4}+1)}, n_{(\alpha,\beta,i,\frac{N}{4}+2)}, \dots, n_{(\alpha,\beta,i,\frac{N}{2})}$  will transfer respective messages to  $n_{(\alpha,\beta,i,1)}, n_{(\alpha,\beta,i,2)}, \dots, n_{(\alpha,\beta,i,\frac{N}{4})}$ , linked according to topological properties of this network.

5: until all nodes have finished transmitting and forwarding.

### Algorithm 2. Step 2 of AAB

- a. /\* This operation is common between all root processors of each row of each block,
- b. Root processors of each row of a block are identified as in figure B,
- c. The transfer of information of all root processors of respective rows is conducted according to connectivity. \*/

1: Starting from each row of each block. The root nodes of respective rows will transfer data to connected nodes of that

row.

2: repeat

3: until all nodes have received the information of root processors.

After the completion of step 2 the position of data in a row is shown in figure 8. The data from the root node of a row of all blocks of network receives the complete information of that row as the content of  $WA_1$ .

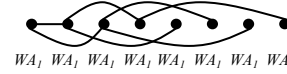


Fig. 8. After Step 2

### Algorithm 3. Step 3 of AAB

- a. /\* This operation is common between all root processors of each column of each block,
- b. The transfer is conducted in order  $\frac{n_{(\alpha,\beta,i,j)}}{2 \times \text{Count of Iteration} - 1} < i \leq \frac{n_{(\alpha,\beta,i,j)}}{2 \times \text{Count of Iteration} - 1} * /$

1: Starting from each column of each block of network, such that the processor with greater index value will transfer data to lower index processors linked according to the topological properties of network.

2: repeat

3: Select nodes  $n_{(\alpha,\beta,i,\frac{N}{2}+1)}, n_{(\alpha,\beta,i,\frac{N}{2}+2)}, n_{(\alpha,\beta,i,\frac{N}{2}+3)}, \dots, n_{(\alpha,\beta,i,N)} \in N$  from each block of network such that at each transfer the block is divided in two parts (e.g. if  $N = 40$ , number of nodes in blocks will also be 40 and division will be 1 to 20 and 21 to 40<sup>th</sup> index position) and transfer message to remaining nodes  $n_{(\alpha,\beta,i,1)}, n_{(\alpha,\beta,i,2)}, n_{(\alpha,\beta,i,3)}, \dots, n_{(\alpha,\beta,i,\frac{N}{2})} \in N$  linked according to topological properties of this network.

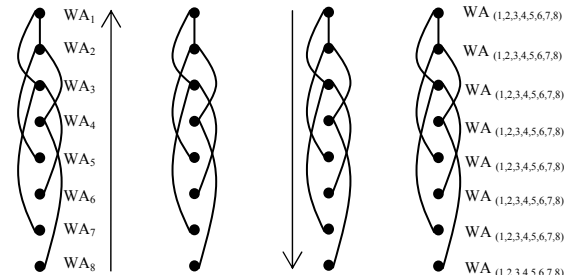


Fig. 9. a) Step 3 b) After Step 3 c) Step 4 d) After Step 4

Figure 9 shows the Step 3 and 4 in which the communication is performed in each column of each block of the network. After the completion of step 4 each column of each block of network consists of complete information of respective column.

### Algorithm 4. Step 4 of AAB

- a. /\* This operation is common between all root processors of each column of each block,
- b. Root processors of each column of each block are identified as in figure C,
- c. The transfer of information of all root processors of respective columns is conducted according to connectivity. \*/

1: Starting from each column of each block. The root nodes of respective columns will transfer data to connected nodes of that

- 
- row.  
 2: repeat  
 3: until all nodes have received the information of root processors.
- 

#### Algorithm 5. Step 5 of AAB (Interblock Communication)

/\* The step is performed using the °horizontal interblock links of this network which transfers the information of all the blocks of respective rows to the root processors of respective block with processor index ( $j = N$ ) \*/

- 
- 1: Starting from each blocks of each rows the information is communicated to the root processors of respective block in such a manner that the processor index  $n_{(\alpha,\beta,i,j=N)}$ .
  - 2: In one communication step this information is broadcasted to every root processor of respective block of respective row. This step is performed on entire network.  
*Note: At the end of this step every root processor contains the information of complete block from which this information is broadcasted.*
- 

#### Algorithm 6. Step 6 of AAB (Interblock Communication)

- a. /\* This step uses algorithm 3 for communicating the information received after step 5 (algorithm 5).
- b. This operation is common between all root processors of each column of each block,
- c. The transfer is conducted in order  $\frac{n_{(\alpha,\beta,i,j)}}{2 \times \text{Count of Iteration} - 1} < i \leq \frac{n_{(\alpha,\beta,i,j)}}{2 \times \text{Count of Iteration} - 1} *$

- 
- 1: Starting from each column of each block of network, such that the processor with greater index value will transfer data to lower index processors linked according to the topological properties of network.
  - 2: repeat
  - 3: Select nodes  $n_{(\alpha,\beta,i,\frac{N}{2}+1)}, n_{(\alpha,\beta,i,\frac{N}{2}+2)}, n_{(\alpha,\beta,i,\frac{N}{2}+3)}, \dots, n_{(\alpha,\beta,i,N)} \in N$  from each block of network such that at each transfer the block is divided in two parts and transfer message to remaining nodes  $n_{(\alpha,\beta,i,1)}, n_{(\alpha,\beta,i,2)}, n_{(\alpha,\beta,i,3)}, \dots, n_{(\alpha,\beta,i,\frac{N}{2})} \in N$  linked according to topological properties of this network.
- 

#### Algorithm 7. Step 7 of AAB (Interblock Communication)

/\* °One-to-all broadcast is used in the block\*/

To transfer the information of a block in a row to other block of respective rows the one-to-all broadcast algorithm is used.  
*Note: At the end of this step, complete blocks of each row have information of all processors in that row.*

---

#### Algorithm 8. Step 8 of AAB (Interblock Communication)

/\* The step is performed using the horizontal interblock links of this network which transfers the information of all the blocks of respective columns to the root processors of respective block with processor index ( $t = N$ ) \*/

- 
- 1: Starting from each blocks of each columns the information is communicated to the root processors of respective block in such a manner that the processor index  $n_{(\alpha,\beta,i=N,j)}$ .
  - 2: In one communication step this information is broadcasted to every root processor of respective block of respective column.
- 

This step is performed on entire network.

*Note: At the end of this step every root processor contains the information of complete block from which this information is broadcasted.*

---

#### Algorithm 9. Step 9 of AAB

- a. /\* This operation is common between all processors of each row of each block,
- b. Each node is represented by  $n_{(\alpha,\beta,i,j)}$ ;
- c. The transfer is conducted in order  $\frac{n_{(\alpha,\beta,i,j)}}{2 \times \text{Count of Iteration} - 1} < i \leq \frac{n_{(\alpha,\beta,i,j)}}{2 \times \text{Count of Iteration} - 1} *$

- 
- 1: Starting from each row of each block of network, such that the processor with greater index value will transfer data to lower index processors linked according to the topological properties of network.
  - 2: repeat
  - 3: Select nodes  $n_{(\alpha,\beta,i,\frac{N}{2}+1)}, n_{(\alpha,\beta,i,\frac{N}{2}+2)}, n_{(\alpha,\beta,i,\frac{N}{2}+3)}, \dots, n_{(\alpha,\beta,i,N)} \in N$  from each block of network such that at each transfer the block is divided in two parts and transfer message to remaining nodes  $n_{(\alpha,\beta,i,1)}, n_{(\alpha,\beta,i,2)}, n_{(\alpha,\beta,i,3)}, \dots, n_{(\alpha,\beta,i,\frac{N}{2})} \in N$  linked according to topological properties of this network.  
*Note: The message will be transferred from higher processor index to lower.*
  - 4: Select nodes  $n_{(\alpha,\beta,i,1)}, n_{(\alpha,\beta,i,2)}, n_{(\alpha,\beta,i,3)}, \dots, n_{(\alpha,\beta,i,\frac{N}{2})} \in N$  (other than the nodes from which message has already transferred) from each block of network such that at each transfer these nodes are divided in two parts (same as in 3; i.e.  $n_{(\alpha,\beta,i,1)}, n_{(\alpha,\beta,i,2)}, \dots, n_{(\alpha,\beta,i,\frac{N}{4})}$ , and  $n_{(\alpha,\beta,i,\frac{N}{4}+1)}, n_{(\alpha,\beta,i,\frac{N}{4}+2)}, \dots, n_{(\alpha,\beta,i,\frac{N}{2})} \in N$ ). Now  $n_{(\alpha,\beta,i,\frac{N}{4}+1)}, n_{(\alpha,\beta,i,\frac{N}{4}+2)}, \dots, n_{(\alpha,\beta,i,\frac{N}{2})}$  will transfer respective messages to  $n_{(\alpha,\beta,i,1)}, n_{(\alpha,\beta,i,2)}, \dots, n_{(\alpha,\beta,i,\frac{N}{4})}$ , linked according to topological properties of this network.
  - 5: until all nodes have finished transmitting and forwarding.
- 

#### Algorithm 10. Step 10 of AAB

/\* °AAB is used in the block\*/

Select block from each column to transfer information of a block in a column to other block of respective columns for this AAB is used.

*Note: At the end of this step, all the processors of each block contains information of all processors of the network.*

---

## 4. Implementing LNC on AAB using MMT

In this section we implement network coding for each step to make the communication faster and increase the rate of information transmitted from each node. We consider network as delay-free (acyclic) and  $o(l) \neq d(l)$ . The algorithm results are analyzed later with  $n=8$  processors. For each step independent and different algorithms are used (see section IV) and linear coding is implemented with each algorithm. According to algorithm 1, data from

all processors are transferred with  $n = 8$  and  $count = 1$  to 2 i.e.,  $(8/(2 \times 1 - 1)) < j \leq (8/2 \times 1) = 8 < j \leq 4$ , which means the processors  $P_1, P_2, P_3$  and  $P_4$  will receive data from  $P_5, P_6, P_7$  and  $P_8$ , shown in figure 10.

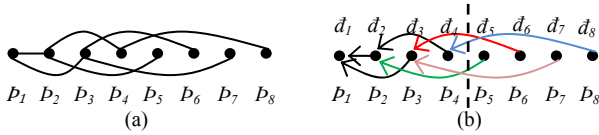


Fig. 10. (a) Shows the indexing of processors with respect to nodes in the figure. (b) Shows the direction of flow of data in step 1 of AAB algorithm on MMT,  $P_1, P_2, P_3$  and  $P_4$  are the processor receiving data and  $P_5, P_6, P_7$  and  $P_8$  are the sending processors. The dotted line distinguishes between the receiving and sending processors in first iteration of step 1.

**Step 1:** Linear coding is implemented on  $P_1, P_2$  and  $P_3$  processors, as these are receiving a set of data form source processors  $P_5, P_6, P_7$  and  $P_8$  in first iteration. Processor  $P_8$  is source and  $P_4$  is its destination;  $P_7$  and  $P_6$  are sources and  $P_3$  is their destination; lastly in  $\log n$  iteration i.e. (3 iteration for  $n = 8$ ),  $P_1$  will receive data from  $P_2$  and  $P_3$ . After implementation of LNC according to LCM-PA on these sources and destinations, step 1 will work as in figure 11. During first iteration of AAB on MMT, LCM-PA will work as in figure 11 (a). Data from  $P_5, P_6, P_7$  and  $P_8$  is sent to  $P_2, P_3, P_3$  and  $P_4$  respectively. So, the complete set of data from all processors reached processor  $P_1$ , i.e. after execution of step 1 all data, in a row, will reach its root processors, but due to LCM-PA the data reached  $P_1$  will have time complexity of  $(\log n - 1)$ , as one step is reduced during transfer of the data using LCM-PA model.

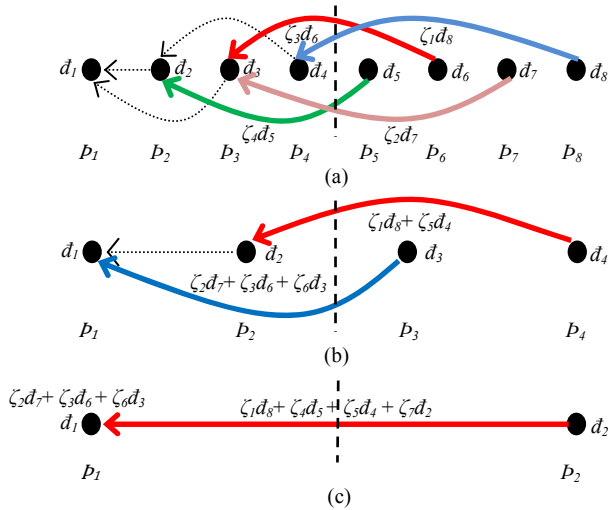


Fig. 11. (a) Iteration first of step 1; data from processors  $P_5, P_6, P_7$  and  $P_8$  is sent to processors to  $P_4, P_3, P_3$  and  $P_2$  respectively. (b) Iteration second of step 1; data from processors  $P_4$  and  $P_3$  is sent to processors to  $P_2$  and  $P_1$  respectively. (c) Iteration third of step 1; data from processors  $P_2$  is sent to processors  $P_1$ .

**Step 2:** The root processors of each row, ( $P_1$ : root processor of first block and first row) will broadcast the

data (from  $P_1$ :  $\zeta_1 \bar{d}_8 + \zeta_2 \bar{d}_7 + \zeta_3 \bar{d}_6 + \zeta_4 \bar{d}_5 + \zeta_5 \bar{d}_4 + \zeta_6 \bar{d}_3 + \zeta_7 \bar{d}_2$ ) to all the processors of respective row using intrablock links transfer, see figure 12.

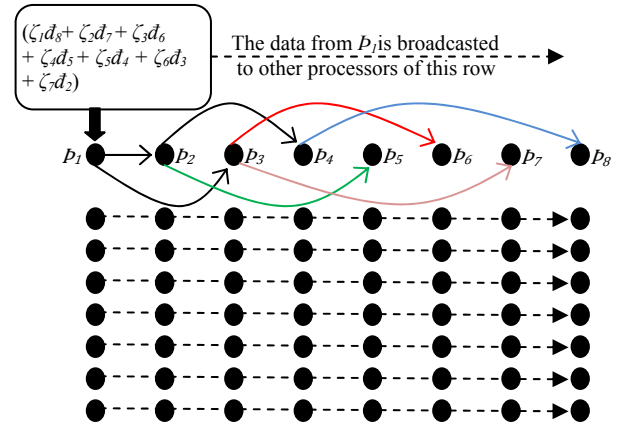


Fig. 12. The data from each row root processor is broadcasted to other processors of respective row in each block.

The time complexity for this step will be reduced by  $n$  i.e.,  $n(\log n - 1)$ . This step is a broadcasting step in each block with intrablock links of MMT. At the end of this step, complete data from root processor is received by other processors of that row. LCM-PA is applied at the same level as in step 1, but the size of data increases to  $n$ .

**Step 3:** This step is similar to step 1, but in this step the data is broadcasted in column-wise order of each block. Linear coding is implemented on  $P_{11}, P_{12}$  and  $P_{13}$  processors, as these are receiving a set of data form source processors  $P_{15}, P_{16}, P_{17}$  and  $P_{18}$  in first iteration. Processor  $P_{18}$  is source and  $P_{14}$  is its destination;  $P_{17}$  and  $P_{16}$  are sources and  $P_{13}$  is their destination; lastly in  $\log n$  iteration i.e. (3 iteration for  $n = 8$ ),  $P_{11}$  will receive data from  $P_{12}$  and  $P_{13}$ . After implementation LCM-PA on these sources and destinations, step 3 works as in figure 13.

**Step 4:** In this step all the root processors of each column and each block, ( $P_{11}$ : root processor of first block and first column) will broadcast the data (from  $P_{11}$ :  $\zeta_1 \bar{d}_{18} + \zeta_2 \bar{d}_{17} + \zeta_3 \bar{d}_{16} + \zeta_4 \bar{d}_{15} + \zeta_5 \bar{d}_{14} + \zeta_6 \bar{d}_{13} + \zeta_7 \bar{d}_{12}$ ) to all the processors of respective column using intrablock links transfer, see figure 14. The time complexity of this step is reduced by  $n^2$  i.e.,  $n^2(\log n - 1) = n^2 \log n - n^2$ . The coefficient value ( $\zeta_i$ ) in step 4 is different from the coefficient value in step 1.

**Step 5:** After step 4, each processors of respective columns contains information of all processors of that column. The step 5, perform the interblock communication using the horizontal interblock links which transfers this information (of all the blocks of respective rows) to the root processors (of respective block), and this requires one communication step (CS) [19]. The time complexity of this step will be same as of AAB i.e. 1CS.

**Step 6:** Using step 3, for transferring information of all the processors in the column at the processors with  $P\_ID$  ( $j=n$ ), so the WA of all the processors is transferred in the

column in the order  $n/(2count-1) < j \leq n/(2count)$ .  
 Time complexity of step 6:  $n^3 \log n$ .

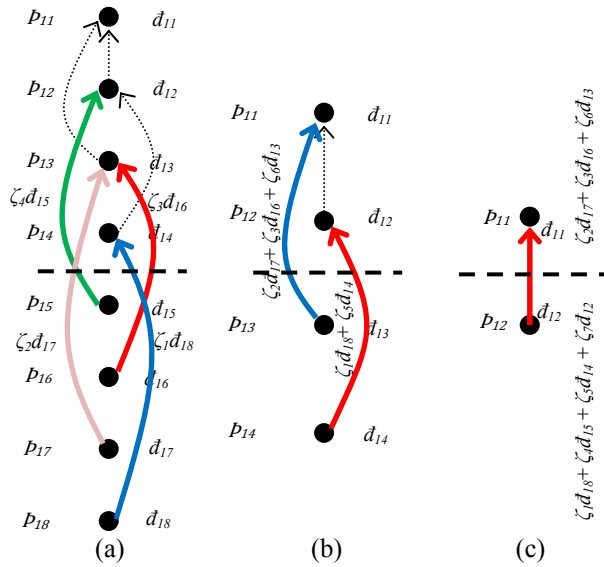


Fig. 13. (a) Iteration first of step 3; data from processors  $P_{15}, P_{16}, P_{17}$  and  $P_{18}$  is sent to processors to  $P_{14}, P_{13}, P_{12}$  and  $P_{11}$  respectively. (b) Iteration second of step 3; data from processors  $P_{14}$  and  $P_{13}$  is sent to processors to  $P_{12}$  and  $P_{11}$  respectively. (c) Iteration third of step 3; data from processors  $P_{12}$  is sent to processors  $P_{11}$ .

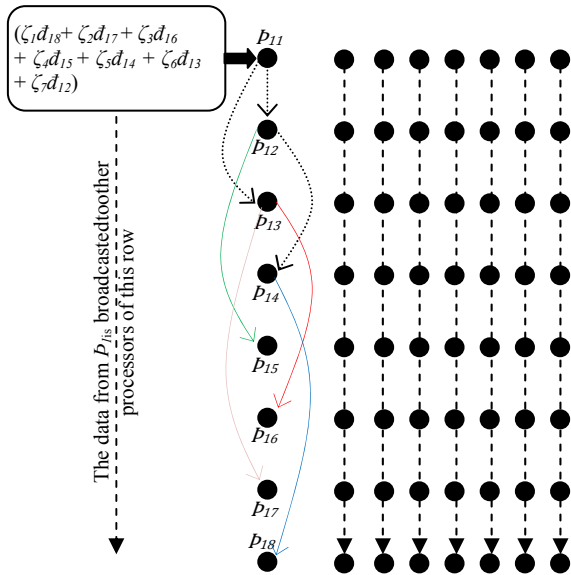


Fig. 14. The data from each column root processor is broadcasted to other processors of respective column in each block.

**Step 7:** Call one-to-all algorithm [19] in the block to transfer the INFO of other blocks (of respective rows) in  $n^3 \log n$  time. At the end of this step, complete blocks of each row have INFO of all the processors in that row. Time complexity of step 7:  $n^3 \log n$ .

**Step 8:** This step performs the interblock communication using horizontal link transfer that transfers the INFO (of all the blocks of respective column) to the root processors

(of respective block) with P\_ID ( $i=n$ ), and this requires one communication step. Time complexity of step 8:  $1CS$ .

**Step 9:** Using step 1 transfer of INFO of all the processors with P\_ID ( $i=n$ ). Time complexity of step 9:  $n^4 \log n$ .

**Step 10:** Call AAB algorithm in the block to transfer the INFO of other blocks that column in the block with  $n^4 \log n$  time complexity. Time complexity of step 10:  $n^4 \log n$ .

At the end of this step all, the processors of each block have the INFO of all processors of other blocks.

## 5. Results and Simulations

The implementation of linear coding using AAB on MMT enables the sharing of data between multiple processors, at a time unit, more convenient and easy. As the algorithm becomes more complex, the involvement of processors also increases. For parallel architectures, important issue is to make these architectures more processor utilitarian, otherwise the processors in these architectures are idle, and all are not in use at every step of algorithms. Also, the involvement of coefficients used to broadcast data is high, compared to coefficients involvement after implementation of LCM-PA with AAB on MMT. This makes the algorithm less complex as fewer amounts of coefficients are used for broadcasting data using linear coding. While broadcasting the data in AAB, the time involved to communicate and deliver/receive data from different processors is more. The fall of time complexity at different number of processors shows that the architecture is possible with a set of processors having a combination which makes the algorithm to be implemented with positive results.

The algorithm starts with the execution of each step in the order defined (as step 1... step 10), as the execution of each step starts the involvement of each processors also increases to broadcast data. In parallel processing the algorithm starts with active processor and involves other processors as it progresses [22]. Figure 15 illustrate the involvement of processors with average percentage of iteration in each step.

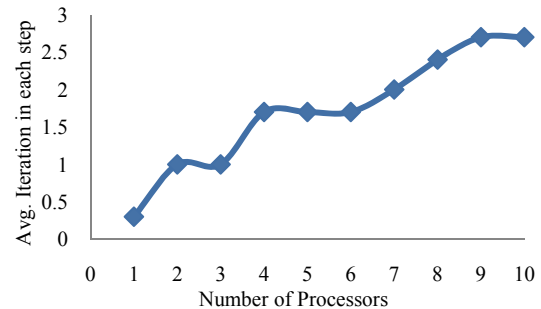


Fig. 15. Involvement of processors at different steps of algorithm.

Based on the above result in figure 15, as the iterations increases the involvement of processors also increases. The algorithm with LCM-PA approach, utilizes the maximum number of processors compared to without LCM-PA approach. So the utilization of processors in parallel architectures is also increases while using linear coding.

## 6. Conclusion and Future Work

We have presented a LCM-PA, model of linear coding, on parallel architecture with efficient implementation of our approach on AAB algorithm on MMT, with comparative time complexity after implementation with LCM-PA. Our model is network independent and can be implemented on any parallel architecture with assumptions to be common as we have used in section second. Future work includes extensions to this approach and analyzing the complexity aspects by implementing with other parallel algorithms (e.g. Multi-Sort [23]). In addition, to make the extension of this approach with LCM-PA model it is needed to be implemented with other parallel algorithms to make vision of research more clear.

### Appendix

Here we provide the proof of all theorems, definitions and terms used with main text. The definitions used in this paper are defined by other authors but for readers convenience they are elaborated with proof in this section.

**Definition 1 (Horizontal intrablock links).** The processors in row  $i$  of each block  $B(\alpha, \beta)$  are connected to form a binary tree rooted at  $(\alpha, \beta, i, 1)$ ,  $1 \leq i \leq n$ . That is, for  $j = 1$  to  $\lfloor n/2 \rfloor$  processor  $P(\alpha, \beta, i, j)$  is directly connected to the processors  $P(\alpha, \beta, i, 2j)$  and  $P(\alpha, \beta, i, 2j + 1)$ , whenever they exist.

**Proof.** If this network is used for  $N$  number of processors than this type of link exists. Suppose  $N = 4$ , then total number of processors in the network are  $N^4 = 256$  processors, which are divide in four rows and four columns and each row and column consists of four block, and each block consists of four rows and four columns. Now according to definition 1, the processors of block  $B(1, 1)$  are connected in order:

$P(1, 1, 1, 1) \rightarrow P(1, 1, 1, 2)$  and  $P(1, 1, 1, 3)$ ;  
 $P(1, 1, 1, 2) \rightarrow P(1, 1, 1, 4)$ ; //as  $P(1, 1, 1, 5)$  does not exist.  
 $P(1, 1, 2, 1) \rightarrow P(1, 1, 2, 2)$  and  $P(1, 1, 2, 3)$ ;  
 $P(1, 1, 2, 2) \rightarrow P(1, 1, 2, 4)$ ;  
 $P(1, 1, 3, 1) \rightarrow P(1, 1, 3, 2)$  and  $P(1, 1, 3, 3)$ ;  
 $P(1, 1, 3, 2) \rightarrow P(1, 1, 3, 4)$ ;  
 $P(1, 1, 4, 1) \rightarrow P(1, 1, 4, 2)$  and  $P(1, 1, 4, 3)$ ;  
 $P(1, 1, 4, 2) \rightarrow P(1, 1, 4, 4)$ ;

As the values of  $i$  and  $j$  changes the number of connecting horizontal link also varies.<sup>o</sup>

**Definition 2 (Vertical intrablock links).** The processors in column  $j$  of each block  $B(\alpha, \beta)$  are also used to form a binary tree rooted at  $(\alpha, \beta, 1, j)$ ,  $1 \leq j \leq n$ . That is, for  $i = 1$  to  $\lfloor n/2 \rfloor$  processor  $P(\alpha, \beta, i, j)$  is directly connected to the processors  $P(\alpha, \beta, 2i, j)$  and  $P(\alpha, \beta, 2i + 1, j)$ , whenever they exist.

**Proof.** If this network is used for  $N$  number of processors than this type of link exists. Suppose  $N = 4$ , then total number of processors in the network are  $N^4 = 256$  processors, which are divide in four rows and four columns and each row and column consists of four block, and each block consists of four rows and four columns. Now according to definition 2, the processors of block  $B(1, 1)$  are connected in order:

$P(1, 1, 1, 1) \rightarrow P(1, 1, 2, 1)$  and  $P(1, 1, 3, 1)$ ;  
 $P(1, 1, 2, 1) \rightarrow P(1, 1, 4, 1)$ ; //as  $P(1, 1, 5, 1)$  does not exist.  
 $P(1, 1, 1, 2) \rightarrow P(1, 1, 2, 2)$  and  $P(1, 1, 3, 2)$ ;  
 $P(1, 1, 2, 2) \rightarrow P(1, 1, 4, 2)$ ;  
 $P(1, 1, 1, 3) \rightarrow P(1, 1, 2, 3)$  and  $P(1, 1, 3, 3)$ ;  
 $P(1, 1, 2, 3) \rightarrow P(1, 1, 4, 3)$ ;  
 $P(1, 1, 1, 4) \rightarrow P(1, 1, 2, 4)$  and  $P(1, 1, 3, 4)$ ;  
 $P(1, 1, 2, 4) \rightarrow P(1, 1, 4, 4)$ ;

As the values of  $i$  and  $j$  changes the number of connecting horizontal link also varies.<sup>o</sup>

**Definition 3 (Horizontal interblock links).**  $\forall \alpha, 1 \leq \alpha \leq n$ , the processor  $P(\alpha, \beta, i, 1)$  is directly connected to the processor  $P(\alpha, i, \beta, n)$ ,  $1 \leq i, \beta \leq n$ . It can be noted that for  $\beta = i$ , these links connect two processors within the same block.

**Proof.** These are the links between the boundary or corner processors of different blocks. If this network is used for  $N$  number of processors than this type of link exists. Suppose  $N = 4$ , according to definition 3, the processors for  $1 \leq \alpha \leq n$  are connected in order:

$P(1, 1, 1, 1) \rightarrow P(1, 1, 1, 4)$ ;  $P(1, 1, 2, 1) \rightarrow P(1, 2, 1, 4)$ ;  
 $P(1, 1, 3, 1) \rightarrow P(1, 3, 1, 4)$ ;  $P(1, 1, 4, 1) \rightarrow P(1, 4, 1, 4)$ ;  
 $P(1, 2, 1, 1) \rightarrow P(1, 1, 2, 4)$ ;  $P(1, 3, 1, 1) \rightarrow P(1, 1, 3, 4)$ ;  
 $P(1, 4, 1, 1) \rightarrow P(1, 1, 4, 4)$ ;  $P(2, 1, 2, 1) \rightarrow P(2, 2, 1, 4)$ ;

As the values of  $\beta$ , and  $i$  changes the number of connecting horizontal links also varies.<sup>o</sup>

**Definition 4 (Vertical interblock links).**  $\forall \beta, 1 \leq \beta \leq n$ , the processor  $P(\alpha, \beta, 1, j)$  is directly connected to the processor  $P(j, \beta, n, \alpha)$ ,  $1 \leq j, \alpha \leq n$ . It can be noted that for  $\alpha = j$ , these links connect two processors within the same block.

**Proof.** These are the links between the boundary or corner processors of different blocks. If this network is used for  $N$  number of processors than this type of link exists. Suppose  $N = 4$ , according to definition 3, the processors for  $1 \leq \beta \leq n$  are connected in order:

$P(1, 1, 1, 1) \rightarrow P(1, 1, 4, 1)$ ;  $P(1, 1, 1, 2) \rightarrow P(2, 1, 4, 1)$ ;



$P(1, 1, 1, 3) \rightarrow P(3, 1, 4, 1); P(1, 1, 1, 4) \rightarrow P(4, 1, 4, 1);$   
 $P(2, 1, 1, 1) \rightarrow P(1, 1, 4, 2); P(3, 1, 1, 1) \rightarrow P(1, 1, 4, 3);$   
 $P(4, 1, 1, 1) \rightarrow P(1, 1, 4, 4);$

As the values of  $\alpha$  and  $i$  changes the number of connecting vertical links also varies.°

**Definition 5 (Directed Graph).** A parallel network in any of the phase of communication is said to be directed based on the flow of data with respect to the algorithm.

**Proof.** A parallel network is said to be directed, when the flow of data is decided based on some parameters. As MMT network is bidirectional, in some part of communication it is using a specific orientation for communication while in some parts it may be reverse, based on the algorithm used to decide the communication. As an example consider algorithm 1 in which the communication is performed from greater processor index to lesser processor index, so the direction is different from algorithm 2 in which the root processor transfers data to other processors of respective rows.°

## References

- [1] R. Ahlswede, N. Cai, S.-Y. Li, and R. Yeung, "Network information flow", *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1204-1216, July 2000.
- [2] P. A. Chou, Y. Wu, and K. Jain, "Practical network coding", in *Proc. 41st Annu. Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Sep. 2003.
- [3] A. Argawal and M. Charikar, On the advantage of network coding for improving network throughput, in Proc. 2004 IEEE Information Theory Workshop, (2004).
- [4] A. Rasala-Lehman, Network coding. Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science (2005).
- [5] S. -Y. R. Li and R. W. Yeung, On the Theory of Linear Network Coding, submitted to IEEE Trans. Inform. Theory.
- [6] Z. Li and B. Li, Network coding in undirected networks, in Proc. 38th Annual Conference on Information Sciences and Systems, (Princeton, NJ) (2004) 17-19.
- [7] G. Kramer and S. A. Savari, Cut sets and information flow in networks of two-way channels, in Proc. 2004 IEEE International Symposium on Information Theory, (2004).
- [8] J. Widmer, C.Fragouli, and J.-Y. Le Boudec, "Low-complexity energy efficient broadcasting in wireless ad-hoc networks using network coding". in *Proc. Workshop on Network Coding, Theory, and Applications*, Apr.2005.
- [9] C. Fragouli, J. Widmer and J. -Y. L. Boudec, "A network coding approach to energy efficient broadcasting", in *Proc. INFOCOM06* Apr. 2006.
- [10] P. A. Chou, Y. Wu, and K. Jain, "Practical network coding", in *Proc. 41st Annu. Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Sep. 2003.
- [11] S.-Y. R. Li, R. W. Yeung, and N. Cai, "Linear network coding", *IEEE Tran. Inf. Theory*, vol. 49, no. 2, pp. 371-381, Feb. 2003.
- [12] P. Sanders, S.Egner, and L.Tolhuizen, "Polynomial time algorithms for network information flow", in *Proc. 15th ACM Symp. Parallel Algorithms and Architectures*, San Diego, CA, pp. 286-294, 2003
- [13] R. Koetter and M. Medard, "An algebraic approach to network coding", *IEEE/ACM Trans. Netw.*, vol. 11, no. 5, pp. 782-795, Oct. 2003.

- [14] E. R. Berlekamp, *Block coding for the binary symmetric channel with noise-less, delayless, feedback in Error Correcting Codes*, (H. B. Mann, ed.) Wiley: New York, 1968.
- [15] S. Lin and D. J. Costello Jr., *Error control coding: Fundamentals and applications*. Prentice-Hall, 1983.
- [16] R. E. Blahut, *Theory and practice of error control codes*. Addison-Wesley: Massachusetts, 1983.
- [17] S. B. Wicker, *Error control systems for digital communication and storage*. Englewood Cliffs: Prentice Hall, 1995.
- [18] P.K. Jana, "Multi-Mesh of Trees with its parallel algorithms", *Journal of System Architecture*, ELSEVIER, vol. 50, issue 4, pp. 193-206, March 2004.
- [19] Nitin Rakesh and Nitin, "Analysis of All to All Broadcast on Multi Mesh of Trees Using Genetic Algorithm", *International Workshop on Advances in Computer Networks, VLSI, ANVIT 2009*, St. Petersburg, Russia.
- [20] D. Das, B.P. Sinha, "Multi-mesh an efficient topology for parallel processing", *Proc. of the Ninth International Parallel Processing Symposium*, Santa Barbara CA, April 25-28, 1995, pp. 17-21.
- [21] D. Das, M. Dey, B.P. Sinha, "A New Network Topology with Multiple Mesh", *IEEE Trans. on Computer*, Vol. 48, No. 5, May 1999, pp.536-551.
- [22] Michael J. Quinn, "Parallel Computing: Theory and Practice", *Tata Mcgraw Hill*, New York, edition 2, 1994.
- [23] Nitin Rakesh and Nitin, "Analysis of Multi-Sort Algorithm on Multi-Mesh of Trees (MMT) Architecture", *Journal of Supercomputing*, Springer, March 2010, DOI: 10.1007/s11227-010-0404-4, pp. 1-38.

**Nitin Rakesh** is Sr. Lecturer in the Department of Computer Science and Engineering & Information Technology, Jaypee University of Information Technology (JUIT), Wakanaghat, Solan-173215, Himachal Pradesh, India. He was born on October 30, 1982, in Agra, India. In 2004, he received the Bachelor's Degree in Information Technology and Master's Degree in Computer Science and Engineering from Jaypee University of Information Technology, Noida, India in year 2007. Currently he is pursuing his doctorate in Computer Science and Engineering and his topic of research is parallel and distributed systems. He is a member of IEEE, IAENG and is actively involved in research publication. His research interest includes Interconnection Networks & Architecture, Fault-tolerance & Reliability, Networks-on-Chip, Systems-on-Chip, and Networks-in-Packages, Network Algorithms, Parallel Algorithms, Fraud Detection. Currently he is working on Efficient Parallel Algorithms for advanced parallel architectures.

**Dr. Vipin Tyagi** is Associate Prof. in Department of Computer Science and Engineering at Jaypee University of Information Technology, Wakanaghat, India. He has about 20 years of teaching and research experience. He is an active member of Indian Science Congress Association and President of Engineering Sciences Section of the Association. He is a Life Fellow of the Institution of Electronics and Telecommunication Engineers and a senior life member of Computer Society of India. He is member of Academic-Research and Consultancy committee of CSI. He is elected as Executive Committee member of Bhopal Chapter of CSI and M.P. and CG chapter of IETE. He is a Fellow of Institution of Electronics and Telecommunication Engineers, life member of CSI, Indian Remote Sensing Society, CSTA, ISCA and IEEE, IAENG. He has published more than 50 papers in various journals, advanced research series and has attended several national and international conferences in India and abroad. He is Principal Investigator of research projects funded by DRDO, MP Council of Science and Technology and CSI. He has been a member of Board of Studies, Examiner Member of senate of many Universities. His research interests include Parallel Computing, Image Processing and Digital Forensics.

# Minimization of Call Blocking Probability by Using an Adaptive Heterogeneous Channel Allocation Scheme for Next Generation Wireless Handoff Systems

Debabrata Sarddar<sup>1</sup>, Arnab Raha<sup>1</sup>, Shubhajeet chatterjee<sup>2</sup>, Ramesh Jana<sup>1</sup>, Shaik Sahil Babu<sup>1</sup>, Prabir Kr Naskar<sup>1</sup>, Utpal Biswas<sup>3</sup>, M.K. Naskar<sup>1</sup>.

1. Department of Electronics and Telecommunication Engg, Jadavpur University, Kolkata – 700032.

2. Department of Electronics and Communication Engg, Institute of Engg. & Managment college, saltlake, Kolkata-700091.

3. Department of Computer Science and Engg, University of Kalyani, Nadia, West Bengal, Pin- 741235.

## Abstract

Nowadays IEEE 802.11 based wireless local area networks (WLAN) have been widely deployed for business and personal applications. The main issue regarding wireless network technology is handoff or hand over management. The minimization of handoff failure due to call blocking is an important issue of research. For the last few years plenty of researches had been done to reduce the handoff failure. Here we also propose a method to minimize the handoff failure by using an adaptive heterogeneous channel allocation scheme.

**Keywords:** IEEE 802.11, Handoff failure, GPS (Global Positioning System), Channel allocation, Neighbor APs.

## 1. Introduction

For last few years handoff becomes a burning issue in wireless communication. Every base station has a limited number of channels. Thus a proper channel distribution is required to perform the handoff successfully.

### 1.1 Handoff

When a MS moves out of reach of its current AP it must be reconnected to a new AP to continue its operation. The search for a new AP and subsequent registration under it constitute the handoff process which takes enough time (called handoff latency) to interfere with proper functioning of many applications.

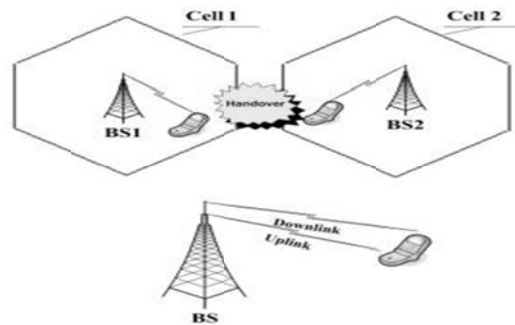


Figure 1. Handoff process

Three strategies have been proposed to detect the need for hand off[1]:

*mobile-controlled-handoff (MCHO)*:The mobile station(MS) continuously monitors the signals of the surrounding base stations(BS)and initiates the hand off process when some handoff criteria are met.

*network-controlled-handoff (NCHO)*:The surrounding BSs measure the signal from the MS and the network initiates the handoff process when some handoff criteria are met.

*mobile-assisted-handoff (MAHO)*:The network asks the MS to measure the signal from the surrounding BSs.the network make the handoff decision based on reports from the MS.

Handoff can be of many types:

*Hard & soft handoff*: Originally hard handoff was used where a station must break connection with the old AP before joining the new AP thus resulting in large handoff delays. However, in soft handoff the old connection is maintained until a new one is established thus significantly reducing packet loss .

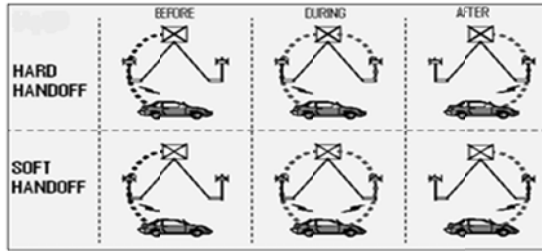


Figure 2. Hard handoff & Soft handoff

In NGWS(next generation wireless system),two types of handoff scenarios arise: horizontal handoff, vertical handoff[2][3].

➤ *Horizontal Handoff*: When the handoff occurs between two BSs of the same system it is termed as horizontal handoff. It can be further classified into two:

- *Link layer handoff* : Horizontal handoff between two BSs that are under the same foreign agent(FA).
- *Intra system handoff* : Horizontal handoff between two BSs that belong to two different FAs and both FAs belong to the same gateway foreign agent (GFA) and hence to the same system.

➤ *Vertical Handoff* : When the handoff occurs between two BSs that belong to two different GFAs and hence to two different systems it is termed as vertical handoff .

Call admission control (CAC) and network resource allocation are the key issues of concern. CAC determines the condition for accepting or rejecting a new call depending upon the availability of sufficient network resources. On the other hand, the network resource allocation decides how to accept incoming connection requests. This is where we are going to apply our method.

### 1.2 Channel distribution

IEEE802.11b and IEEE802.11g operates in the 2.4GHz ISM band and use 11 of the maximum 14 channels available and are hence compatible due to use of same frequency channels. The channels (numbered 1to14) are spaced by 5MHz with a bandwidth of 22MHz, 11MHz above and below the centre of the channel. In addition there is a guard band of 1MHz at the base to accommodate out-of-band emissions below 2.4GHz. Thus a transmitter set at channel one transmits signal from 2.401GHz to 2.423GHz and so on to give the standard channel frequency distribution.

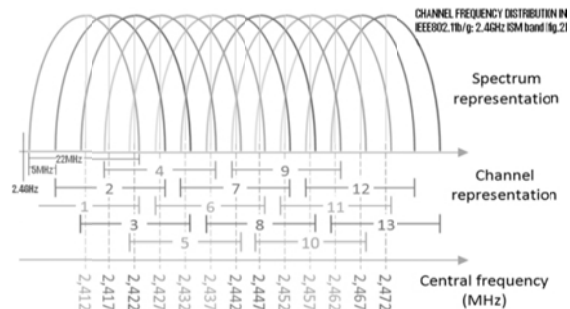


Figure 3.Channel distribution

Many dynamic allocations of channel have been proposed by different authors and all these mechanisms will improve the performance of wireless network. However for practical reason channel allocation is done in a static manner.

In section II we take you through the various works that have already been done to achieve this and in section III We introduce a new method to minimize the call blocking probability by using an adaptive heterogeneous channel allocation scheme for next generation wireless handoff system. This is followed by performance evaluation of our proposed technique using simulations in section IV after which in section V we propose a few areas in which further improvement can be made. Finally, we provide an extensive list of references that has helped us tremendously in our work.

## 2. Related works

In last few years, many researches had been done to develop a user friendly channel allocation. The simplest way of channel allocation is “Guard channel” allocation where the handoff call is given more priority than the new calls by reserving a fixed number of channels for them[1]. In [2], only the new voice calls are buffered in queue whereas in [3], both new call and handoff call are allowed to be queued. Author of [4] proposed a handoff scheme with two level priority reservation. Higher priority is given to the handoff call because termination of ongoing call is more annoying than the new one [5]. All of the above researches are based on voice cellular system. But due to the rapid development in wireless communication, the effect of non-real-time service needs to be taken in consideration [6]. Author in [7] proposed a method where only data service handoff requests are allowed to be queued where as a two dimensional traffic model for cellular mobile system is proposed in [8]. Some algorithms also proposed for multimedia users with fixed bandwidth requirement in [9], [10], [11], [12]. In [13] author used a two dimensional Markov chain to propose a new approximation approach that reduces the computational complexity. Authors of [14], [15] propose a dynamic

channel allocation i.e. no fixed channel among the cells where all channels are kept in a central pool and will be assigned dynamically when the new calls will arrive. For choosing any one channel from the pool where more than one channels are available, a new method is proposed in [16]. In [17], authors proposed a non-preemptive prioritization scheme for access control in cellular networks where as a dynamic buffering is used to minimize the traffic congestion in mobile networks [18].

### 3. Proposed Works

In case of queuing systems the channels registered for hand-off calls alone and the space for both new generating calls and hand-off calls are fixed in nature. We require the demarcation between these two spaces to be adaptive to the requirements of the nature of the area in which these cells belong.

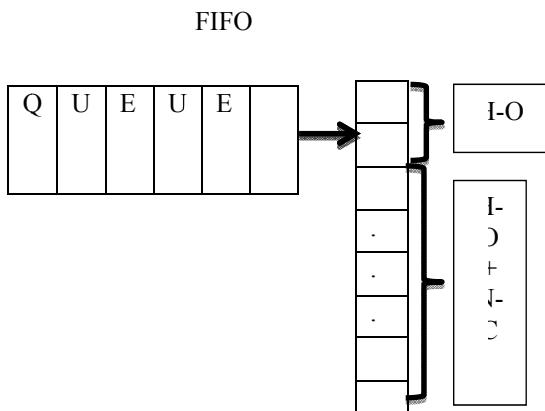


Figure 4. Queue

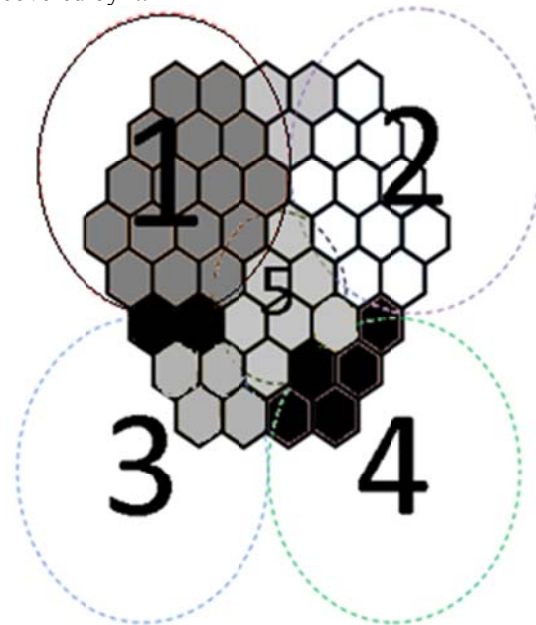
H-O => Channels reserved for Hand-off only.  
 H-O+ N-C=> Channels reserved for Hand-off and New-calls generated within the hexagonal cell.

Now the total number of channels in a particular cell is divided among the different types of calls like new calls, hand-off calls, data calls etc. In our case we are mainly interested with new calls and hand-off calls. If we can devise a method for an optimized and systematic way of dividing the number of channels channel into channels reserved for hand-off and those for new-calls and hand-off both we can reduce the call blocking probability or in other sense the hand-off failure.

Here we assume two kinds of arrival rates: a)  $\lambda_{n-c}$  : arrival rate of new-calls b)  $\lambda_{h-o}$ : arrival rate of hand-off calls.

Although call termination rates ( $\mu_{n-c}$  and  $\mu_{h-o}$ ) play an important role in determining the call blocking probabilities and thereby in determining the hand-off failure probability, but in our case for determination of the fractions of total channels devoted to only Hand-off and

that to both hand-off and new calls are calculated only on the basis of the call arrival rates and their types. Here we assume some discrete fractional allocations based upon the type of call arrival and their rates. In this paper the demarcation lies on the previous call arrival rate history of an area that is the hexagonal cells covered by it.



AREA ID	COLOR	RELATION
1	Grey	$\lambda_{n-c} \ll \lambda_{h-o}$
2	White	$\lambda_{n-c} < \lambda_{h-o}$
3	White	$\lambda_{n-c} > \lambda_{h-o}$
4	Black	$\lambda_{n-c} \gg \lambda_{h-o}$
5	White	$\lambda_{n-c} = \lambda_{h-o}$

Figure 5. Cell Cluster

For the yellow, violet and orange colored cells, the values of  $\lambda_{n-c}$ ,  $\lambda_{h-o}$  are not well-defined i.e. they may be varying widely with time and so no definite relation can be ascertained.

In this case we assume that to be partitioned in to some sub-divisions which have different relations between  $\lambda_{n-c}$ ,  $\lambda_{h-o}$  and hence the subareas covered by the hexagonal cells can be assumed to be heterogeneous. Till date most of the present work in the literature is based upon homogeneous cells and uniform nature of subareas covered by those homogeneous cells.

Our scheme proposes different channel allocation schemes for the different cases as shown above.

When  $\lambda_{n-c} \ll \lambda_{h-o}$ , it is evident that the channels allocated for (H-O+N-C) as denoted earlier should be much greater

than that for only H-O . So we can assign some discrete weights to represent this fraction. This will ensure that the blocking probability is as minimum as possible. The total number of channels may be determined by the following expression.

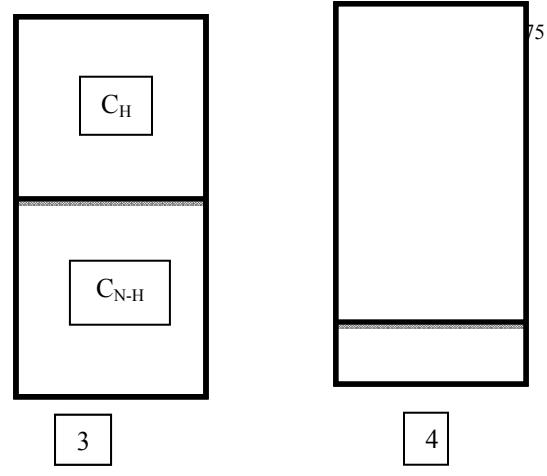
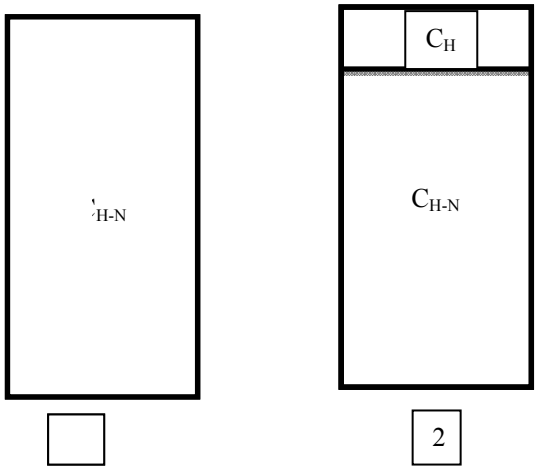
Let,  
 $C_T$ =total number of channels;  
 $C_{H-N}$ = number of channels reserved for both hand-off and new calls generated within the cells;  
 $C_H$ = number of channels reserved only for hand-off.  
 $W_{H-N}$ =weightage on  $C_{H-N}$   
 $W_H$ =weightage on  $C_H$   
 Here we assume  $W_{H-N} + W_H = 1$ .

Determination of the values of  $W_{H-N}$ ,  $W_H$ .  
 $W_{H-N} = \lambda_{n-c} / (\lambda_{n-c} + \lambda_{h-o}) \dots\dots\dots (1)$   
 $W_H = \lambda_{h-o} / (\lambda_{n-c} + \lambda_{h-o}) \dots\dots\dots (2)$   
 Equation (2) is not so significant in this case because suppose for the case  $\lambda_{n-c} = 0$  , it doesn't really make any effect if we take

$W_{H-N} = W_H \cdot 1 \dots\dots\dots (3)$   
 hand-off calls will be processed in any case.

Thereby  
 $C_{H-N} = W_{H-N} * C_T = \lambda_{n-c} / (\lambda_{n-c} + \lambda_{h-o}) * C_T \dots\dots\dots (4)$   
 $C_H = W_H * C_T = \lambda_{h-o} / (\lambda_{n-c} + \lambda_{h-o}) * C_T \dots\dots\dots (5)$   
 Which reaffirms our assumption that:  $C_T = C_{H-N} + C_H \dots\dots\dots (6)$

Now channel allocation can be as varied as the following:



- 1 represents  $\lambda_{n-c} = \lambda_{h-o} = 0$
- 2 represents  $\lambda_{n-c} \gg \lambda_{h-o}$
- 3 represents  $\lambda_{n-c} = \lambda_{h-o} \neq 0$
- 4 represents  $\lambda_{n-c} \ll \lambda_{h-o}$  (although this case is not important)

### 4. Simulation Results

We simulate our proposed method by using the above conception. For justifying the practicability of our method in real models we made an artificial environment where we are going to apply our method. At first we have consider a case where the number of channels reserved for both hand-off and new calls are much greater than the number of channels reserved for handoff calls(25%). Corresponding result is shown in Figure.6. Where we can see up to 25% of the channel, handoff probability is maximum and no call dropping occurs at here.

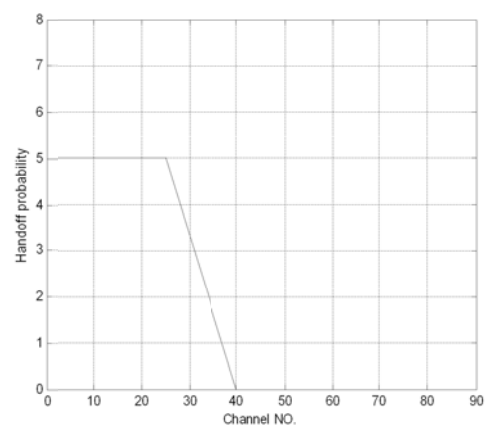


Figure.6

Next, we consider a case where the number of channels reserved for both the hand-off and new calls are equal to the number of channels reserved for handoff calls(50%) and the simulation result shown in below. Here we can see up to 50% of the channel allocation there is no call dropping probability.



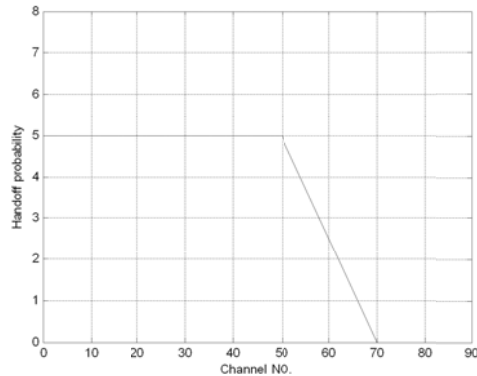


Figure.7

At last, we consider the number of channels reserved for both hand-off and new calls are much smaller than the number of channels reserved for handoff calls (75%). Here also we can see up to 75% of the channel allocation there is no call dropping probability.

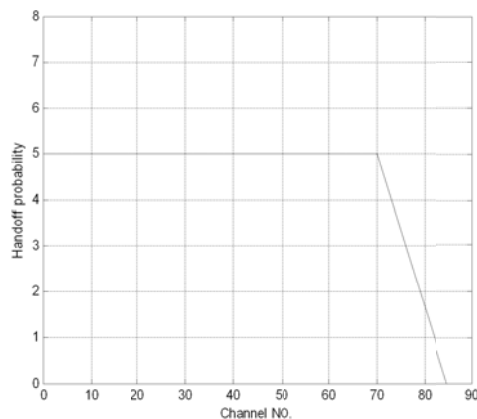


Figure.8

## 5. Conclusion

Our proposed method aims at reducing handoff time by reducing the number of APs to be scanned which is accomplished by fitting a trend equation to the motion of the MS. This in turn reduces the number of channels to be scanned which brilliantly reduces the handoff failure as is clear from the simulation presented in the above section. However the proposed algorithm may prove erroneous if the motion of the MS is too much random to be used for prediction purposes. Future works in this field may include research on more refined algorithms regarding channel allocation. Error estimation method may also be improved. It is worth mentioning here that although the proposed work has been presented considering honeycomb structures yet our algorithm would work in a similar manner for other cell structures and neighbor AP locations. Minor changes would be introduced depending on the network topology.

## References

- [1] Yi-Bing Lin Imrich Chalmatc, "Wireless and Mobile Network Architectures," pp. 17.
- [2] AKYILDIZ, I. F., XIE, J., and MOHANTY, S., "A survey on mobility management in next generation all-IP based wireless systems," IEEE Wireless Communications, vol. 11, no. 4, pp. 16-28, 2004.
- [3] STEMM, M. and KATZ, R. H., "Vertical handoffs in wireless overlay networks," ACM/Springer Journal of Mobile Networks and Applications(MONET), vol. 3, no. 4, pp. 335-350, 1998.
- [4] Lin Y.B and Chlamtac I , "Wireless and MobileNetwork Architecture", John Wiley and Sons Inc., 2001, pp.60-65.
- [5] Guerin R, "Queuing Blocking System with Two Arrival Streams and Guard Channels", IEEE Transactions on Communications, 1998, 36:153-163.
- [6] Zeng A. A, Mukumoto K. and Fukuda A., "Performance Analysis of Mobile Cellular Radio System with Priority Reservation Handoff Procedure", IEEE VTC-94, , Vol 3, 1994, pp. 1829-1833.
- [7] Zeng A. A, Mukumoto K. and Fukuda A., "Performance Analysis of Mobile Cellular Radio System with Two-level Priority Reservation Procedure", IEICE Transactions on Communication, Vol E80-B, No 4, 1997, pp. 598-607.
- [8] Jabbari B. & Tekinay S., "Handover and Channel Assignment in Mobile Cellular Networks", IEEE Communications Magazine, 30 (11),1991, pp.42-46.
- [9] Goodman D. J, "Trends in Cellular and Cordless Communication", IEEE Communications Magazine, Vol. 29, No. 6, 1991, pp.31-40.
- [10] Zeng Q.A and Agrawal D.P, "Performance Analysis of a Handoff Scheme in Integrated Voice/Data Wireless Networks", Proceedings of IEEE VTC-2000, pp. 1986-1992.
- [11] Pavlidou F.N, "Two-Dimensional Traffic Models for Cellular Mobile Systems", IEEE Transactions on Communications, Vol 42, No 2/3/4, 1994, pp. 1505-1511.
- [12] Evans J. and Everitt D., "Effective Bandwidth Based Admission Control for Multiservice CDMA Cellular Networks", IEEE Trans. Vehicular Tech., 48 (1),1999, pp. 36-46.
- [13] Choi S. and Shin K.G , "Predictive and Adaptive Bandwidth Reservation for Handoffs in QoS-Sensitive Cellular Networks", In ACM SIGCOMM'98 Proceedings, 1998, pp. 155-166.
- [14] ] Levine D.A, Akyildz I.F, and Naghshineh M , "A Resource Estimation and Call Admission Algorithm for Wireless Multimedia Networks using the Shadow Cluster Concept". IEEE/ACM Trans. On Networking, 5 (1),1997, 525-537.
- [15] Lu S and Bharghavan V, "Adaptive Resource Management Algorithms for Indoor Mobile Computing Environments", In ACM SIGCOMM'96 Proceedings, 231-242.
- [16] Yuguang Fang and Yi Zhang, "Call Admission Control Schemes and Performance Analysis in Wireless Mobile Networks", IEEE Transactions on Vehicular Technology, Vol. 51, No. 2, pp. 371-382, March 2002.

- [17] Kazunori O. and Fumito K., "On Dynamic Channel Assignment in Cellular Mobile Radio Systems", IEEE International Symposium on Circuits and Systems, 1991, 2:938-941.
- [18] Scott Jordan and Asad Khan, "Optimal Dynamic Allocation in Cellular Systems", IEEE Transactions on Vehicular Technology, 42 (2), 1994, pp. 689-697.
- [19] Cox D.C and Reudink D.O, "Dynamic Channel Assignment in Two Dimension Large-Scale Mobile Radio Systems", The Bell System Technical Journal, 1972, 51:1611-1628.
- [20] ] Novella Bartolini, Handoff and Optimal Channel Assignment in Wireless Networks", Mobile Networks and Applications, 6, 2001, pp. 511-524.
- [21] ] Dutta A, Van den berg E., Famolari D., Fajardo V., Ohba Y., Taniuchi K. & Kodama T., "Dynamic Buffering Control Scheme for Mobile Handoff", Proceedings of 17th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'07), 2007, pp. 203-214.



**Debabrata Sarddar** is currently pursuing his PhD at Jadavpur University. He completed his M.Tech in Computer Science & Engineering from DAVV, Indore in 2006, and his B.Tech in Computer Science & Engineering from Regional Engineering College, Durgapur in 2001. His research interest includes wireless and mobile system.



**Arnab Raha** is presently pursuing B.E. (3rd Year) in Electronics and Telecommunication Engg. at Jadavpur University. His research interest includes wireless sensor networks, advanced embedded systems and wireless communication systems.



**Shubhajeet Chatterjee** is presently pursuing B.Tech Degree in Electronics and Communication Engg. at Institute of Engg. & Management College, under West Bengal University Technology. His research interest includes wireless sensor networks and wireless communication systems.



**Ramesh Jana** is presently pursuing M.Tech (2nd year) in Electronics and Telecommunication Engg. at Jadavpur University. His research interest includes wireless sensor networks, fuzzy logic and wireless communication systems



**Prabir kr Naskar** is currently pursuing his Master of Engineering in Computer Technology from Jadavpur University. He completed his B.Tech in Computer Science and Engineering, from College of Engineering & Leather Technology affiliated to West Bengal University of Technology in 2006. His fields of interest include wireless sensor networks and computer networking.



**SHAIK SAHIL BABU**

is pursuing Ph.D in the Department of Electronics and Telecommunication Engineering under the supervision of Prof. M.K. NASKAR at Jadavpur University, KOLKATA. He did his Bachelor of Engineering in Electronics and Telecommunication Engineering from Muffa Kham Jah College of Engineering and Technology, Osmania University, Hyderabad, and Master of Engineering in Computer Science and Engineering from Thapar Institute of Engineering and Technology, Patiala, in Collaboration with National Institute of Technical Teachers' Training and Research, Chandigarh.



**Utpal Biswas** received his B.E, M.E and PhD degrees in Computer Science and Engineering from Jadavpur University, India in 1993, 2001 and 2008 respectively. He served as a faculty member in NIT, Durgapur, India in the department of Computer Science and Engineering from 1994 to 2001. Currently, he is working as an associate professor in the department of Computer Science and Engineering, University of Kalyani, West Bengal, India. He is a co-author of about 35 research articles in different journals, book chapters and conferences. His research interests include optical communication, ad-hoc and mobile communication, semantic web services, E-governance etc.



**Mrinal Kanti Naskar** received his B.Tech. (Hons) and M.Tech degrees from E&ECE Department, IIT Kharagpur, India in 1987 and 1989 respectively and Ph.D. from Jadavpur University, India in 2006.. He served as a faculty member in NIT, Jamshedpur and NIT, Durgapur during 1991-1996 and 1996-1999 respectively. Currently, he is a professor in the Department of Electronics and Tele-Communication Engineering, Jadavpur University, Kolkata, India where he is in charge of the Advanced Digital and Embedded Systems Lab. His research interests include ad-hoc networks, optical networks, wireless sensor networks, wireless and mobile networks and embedded systems. He is an author/co-author of the several published/accepted articles in WDM optical networking field that include "Adaptive Dynamic Wavelength Routing for WDM Optical Networks" [WOCN,2006], "A Heuristic Solution to SADM minimization for Static Traffic Grooming in WDM uni-directional Ring Networks" [Photonic Network Communication, 2006], "Genetic Evolutionary Approach for Static Traffic Grooming to SONET over WDM Optical Networks" [Computer Communication, Elsevier, 2007], and "Genetic Evolutionary Algorithm for Optimal Allocation of Wavelength Converters in WDM Optical Networks" [Photonic Network Communications,2008].

# On-Demand Multicasting in Ad-hoc Networks: Performance Evaluation of AODV, ODMRP and FSR

Rajendiran. M<sup>1\*</sup> Srivatsa. S. K<sup>2</sup>

<sup>1</sup> Department of Computer Science and Engineering, Sathyabama University, Chennai, Tamilnadu 600119, India

<sup>2</sup> Department of Computer Science and Engineering, St. Josephs College of Engineering, Chennai, Tamilnadu 600119, India

## Abstract

Adhoc networks are characterized by connectivity through a collection of wireless nodes and fast changing network topology. Wireless nodes are free to move independent of each other which makes routing much difficult. This calls for the need of an efficient dynamic routing protocol. Mesh-based multicast routing technique establishes communications between mobile nodes of wireless adhoc networks in a faster and efficient way.

In this article the performance of prominent on-demand routing protocols for mobile adhoc networks such as ODMRP (On Demand Multicast Routing Protocol), AODV (Adhoc on Demand Distance Vector) and FSR (Fisheye State Routing protocol) was studied. The parameters viz., average throughput, packet delivery ration and end-to-end delay were evaluated. From the simulation results and analysis, a suitable routing protocol can be chosen for a specified network. The results show that the ODMRP protocol performance is remarkably superior as compared with AODV and FSR routing protocols.

**Keywords:** MANET, Multicast Routing, ODMRP, AODV, FSR.

## 1. Introduction

One of the basic internet tasks is routing between various nodes. It is nothing other than establishing a path between the source and the destination. However in large and complex networks routing is a difficult process because of the possible intermediate hosts it has to cross in reaching its final destination. In order to reduce the complexity, the network is considered as a collection of sub domains and each domain is considered as a separate entity. This helps routing easy [1]. However basically there are three routing protocols in ad hoc networks namely proactive, reactive and hybrid routing protocols. Of these reactive routing protocols establish and maintain routes based on demand.

The reactive routing protocols (e.g. AODV) usually use distance-vector routing algorithms that keep only information about next hops to adjacent neighbors and

costs for paths to all known destinations [2]. The reactive routing protocols (e.g. AODV) usually use distance-vector routing algorithms that keep only information about next hops to adjacent neighbors and costs for paths to all known destinations [2].

On the other hand hybrid routing protocols combine the advantages of both proactive and reactive protocols. Reliable multicast in mobile network was proposed by Prakash et al. [3]. In their solution the multicast message is flooded to all the nodes over reliable channels. The nodes then collectively ensured that all mobile nodes belonging to the multicast group get the message. If a node moves from one cell to another while a multicast is in progress, delivery of the message to the node was guaranteed.

Tree-based multicast routing provides fast and most efficient way of routing establishment for the communications of mobile nodes in MANET [4]. The authors described a way to improve the throughput of the system and reduce the control overhead. When network load increased, MAODV ensures network performance and improves protocol robustness. Its PDR was found to be effective with reduced latency and network control overhead. On Demand Multicast Routing Protocol is a multicast routing protocol(ODMRP) designed for ad hoc networks with mobile hosts [5]. Multicast is nothing but communication between a single sender and multiple receivers on a network and it transmits a single message to a select group of recipients [6]. Multicast is commonly used in streaming video, in which many megabytes of data are sent over the network. The major advantage of multicast is that it saves bandwidth and resources [7]. Moreover multicast data can still be delivered to the destination on alternative paths even when the route breaks. It is an extension to Internet architecture supporting multiple clients at network layers. The fundamental motivation behind IP multicasting is to save

network and bandwidth resource via transmitting a single copy of data to reach multiple receivers. Single packets are copied by the network and sent to a specific subset of network addresses. These addresses point to the destination. Protocols allowing point to multipoint efficient distribution of packets are frequently used in access grid applications. It greatly reduces the transmission cost when sending the same packet to multiple destinations.

A primary issue in managing multicast group dynamics is the routing path built for data forwarding. Most existing ad hoc multicasting protocols can be classified as tree-based or mesh-based. The tree-based protocol, a tree-like data forwarding path is built with the root at the source of the multicast session. The mesh-based protocol [eg. ODMRP], in contrast, provide multiple routes between any pair of source and destination, intended to enrich the connectivity among group members for better resilience against topology changes.

## 2. Literature Survey

A lot of work has been done to evaluate the performance of routing protocols in ad hoc networks. Thomas Kunz et al. [8] compared AODV and ODMRP in Ad-Hoc Networks. Yadav et al. [9] studied the effects of speed on the Performance of Routing Protocols in Mobile Ad-hoc Networks. Corson et al.[10] discussed the Routing protocol in MANET with performance issues and evaluation considerations. Guangyu et.al. [11] presented the application layer routing as Fisheye State Routing in Mobile Ad Hoc Networks. In view of need to evaluate the performance of ODMRP with other common routing protocols used now days, simulation based experiments were performed by evaluating Packet Delivery Ratio, End to End delay and average throughput. Many researchers have evaluated multicast routing performance under a variety of mobility patterns [12-13].

The fisheye State Routing (FSR) algorithm for ad hoc networks introduces the notion of multi-level “scope” to reduce routing update overhead in large networks [14]. A node stores the link state for every destination in the network. It periodically broadcasts the link state update of a destination to its neighbors with a frequency that depends on the hop distance to that destination. Pei et al. [15] studied the routing accuracy of FSR and identified that it was comparable with an ideal Link State. FSR is more desirable for large mobile networks where mobility is high and the bandwidth is low. It has proved as a flexible solution to the challenge of maintaining accurate routes in ad hoc environments.

## 3. Experimental Setup

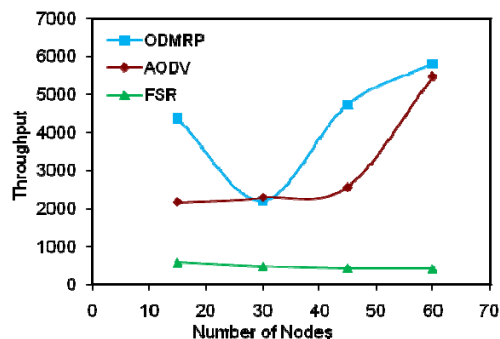
Evaluation of the performance of different routing techniques such as ODMRP, AODV and FSR was carried out through simulation using the GloMoSim v2.03 simulator [16]. The channel capacity of mobile hosts was set at 2Mbps. For each simulation, 60 nodes were randomly placed over a square field whose length and width is 1000 meters. Nodes communicate using MAC and CSMA for the routing protocols ODMRP, AODV and FSR. Each multicast source uses a Constant Bit Rate (CBR) flow. These parameters were chosen from “config.in” file within the simulator. Based on the requirements the values were adjusted and then it was executed. Monitored parameters were average throughput, end to end delay and packet delivery ratio (PDR).

## 4. Results and Discussion

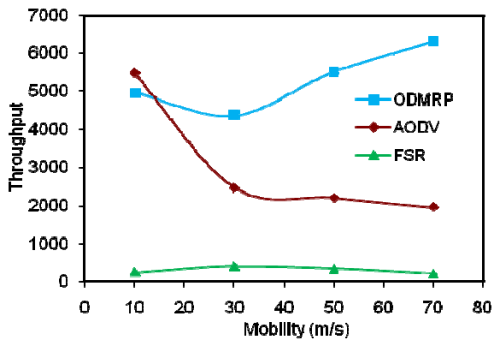
The performance of the three routing protocols, i.e. ODMRP, AODV and FSR were evaluated under varying simulation conditions. The evaluation of performance was done on the basis of monitored parameters, average throughput, end to end delay and packet delivery ratio.

### 4.1 Average Throughput

Average throughput signifies the rate of packets communicated per unit time. The average throughput at a unit time (simulation time of 200 seconds) under varying number of nodes and mobility for all the simulated routing protocols are indicated in the Figure 1 (a-b). It can be observed that under most of nodal conditions the throughput of ODMRP is 4276.25 which are remarkably higher to throughput of AODV (3125.50) and throughput of FSR (487.25).



(a) under varying nodes

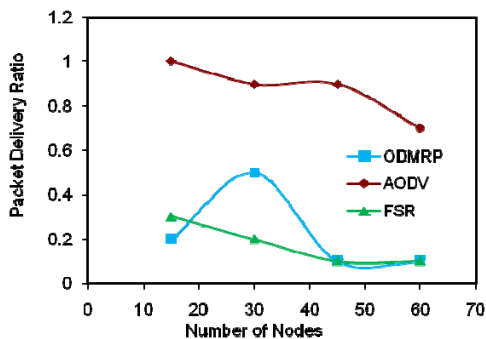


(b) under varying mobility

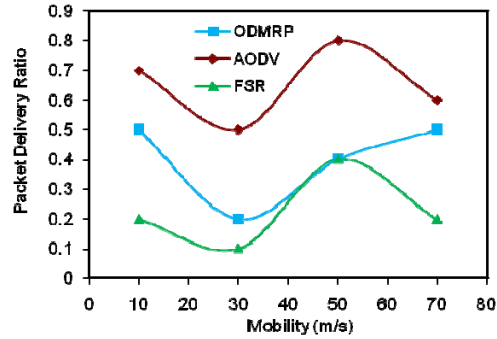
Figure 1 Average throughput under various input conditions

The FSR topology maintains up-to-date information received from neighboring nodes. The topology information is exchanged between neighbors via Unicast. Each node maintains network topology map for distance calculations and when network size increases, the amount of periodic routing information could become large. However the routing packets are not flooded. FSR captures pixels near the focal point with high detail. The details decrease as the distance from the focal point increase. When the mobility increases the routes to remote destinations become less accurate. The route table size still grows linearly with network size [14]. Hence throughput of FSR could here be lower than AODV and ODMRP.

Similarly for different mobility conditions too, ODMRP routing protocol displays increased performance as compared to the other two. The ODMRP average throughput with node mobility is 5276.75 bytes per simulation time as against AODV's 3024.00 and FSR's 298.75. The same reasons as stated for the improved performance of ODMRP under differing number of nodes can be given here too. The same behavior is experienced in the previous studies too under similar conditions [12].



(a) under varying nodes



(b) under varying mobility

Figure 2 Packet delivery ratio under various input conditions

It can be observed that the PDR of AODV routing protocol is higher than the ODMRP and Fisheye state routing protocols. Higher the PDR, higher is the number of legitimate packets delivered without any errors. This shows that AODV exhibits a better delivery system as compared with the other two. The reasons for the higher PDR ratio of AODV can be attributed to its good performance in large networks with low traffic and low mobility. It discovers routes on-demand, and effectively uses available bandwidth. Also it is highly scalable and minimizes broadcast and transmission latency. Its efficient algorithm provides quick response to link breakage in active routes.

Moreover the ability of a routing algorithm to cope with the changes in routes is identified by varying the mobility. In this too the PDR of AODV protocol is higher as compared to the other two. The same reasons for the better PDR ratio of AODV under changing number of nodes can be given here too.

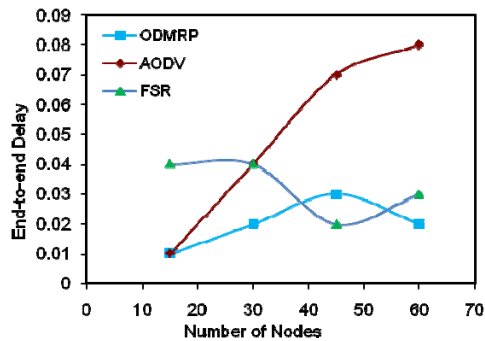
#### 4.3 End-to-End Delay

The total latency between the source and destination experienced by a legitimate packet is given by end-to-end delay. It is calculated by summing up the time periods experienced as processing, packet, transmission, queuing and propagation delays. The speed of delivery is an important parameter in the present day competitive circumstances.

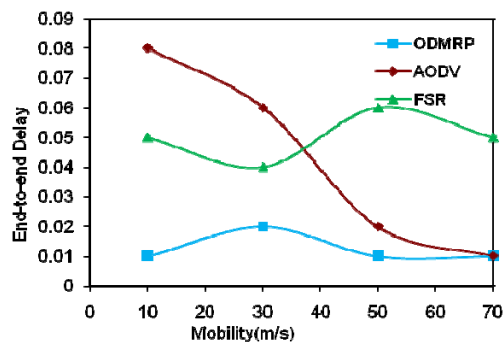
Higher end- to -end delay values imply that the routing protocol is not fully efficient and causes a congestion in the network. The values of end- to- end delay for the protocols ODMRP, AODV and FSR simulated at different number of nodes and differing mobility values are indicated in Figure 3. As against the other two protocols studied ODMRP exhibits lesser values of end-to-end delay. This implies that for ad hoc networks, the multicast



routing protocol ODMRP exhibits a better performance than AODV and FSR.



(a) under varying nodes



(b) under varying mobility

Figure 3 End-to-End Delay under various input conditions

## 5. Conclusions

Performance of the various routing protocols such as ODMRP, AODV and FSR were evaluated in this study. The following conclusions were drawn.

- Both under varying number of nodes and differing values of mobility Average throughput is higher for the routing protocol ODMRP. The maximum throughput of ODMRP is 43% higher than the maximum of AODV and FSR under varying nodes condition.
- AODV has a higher ratio of legitimate packet delivery as compared with the other routing protocols evaluated, ODMRP and FSR. The maximum packet delivery of AODV is 38% higher than the maximum of ODMRP and FSR under varying nodes condition.
- ODRMP performs better in avoiding network congestion as compared to AODV and FSR. The better your paper looks, the better the Journal looks. Thanks for your cooperation and contribution.

## Acknowledgments

The authors gratefully acknowledge Dr.P.Chinnadurai, Secretary & Correspondent, Panimalar Engineering College, Chennai for his encouragement and continuous support. They also appreciate the help rendered by Dr.L.Karthikeyan.

## References

- [1] Nadjib Badache, Djamel Djenouri and Abdelouahid Derhab “Mobility Impact on Mobile Ad hoc Routing Protocols” In ACS/IEEE International Conf. on AICCSA’03, July 2003.
- [2] Ian D.Chakeres and Elizabeth M.Belding-Royer “AODV Routing Protocol Implementation Design” International Conf. on Distributed Computing Sysmtes(ICDCSW’04) IEEE, vol.7 2004
- [3] Ravi Prakash, Andre Schiper and Mansoor Mohsin “Reliable Multicast in Mobile Networks” Proc. of IEEE 2003(WCNC)
- [4] Weiliang Li and Jianjun Hao “Research on the Improvement of Multicast Ad Hoc On-demand Distance Vector in MANETS” IEEE Vol.1 2010
- [5] M.Gerla et al., “On-demand multicast routing protocol (ODMRP) for ad hoc networks”. Internet draft,<draft-ietf-manet-odmrp-04.txt>,(2000)
- [6] Shapour Joudi Begdillo, Mehdi Asadi and Haghghat.A.T. “Improving Packet Delivery Ratio in ODMRP with Route Discovery”, International Jour. Of Computer Science and Network Security, Vol.7 No.12 Dec 2007.
- [7] Gu Jian and Zhang Yi, “A Multi-Constrained multicast Routing Algorithm based on Mobile Agent for Ad Hoc network” International Conference on Communications and Mobile Computing, IEEE 2010.
- [8] Thomas Kunz, and Ed Cheng, “On Demand Multicasting in Ad hoc Networks: Comparing AODV and ODMRP”, Proc, of the 22nd IEEE International Conf. on Distributed Computing Systems(ICDCS’02),Vol-2, pp 1063-6927(2002)
- [9] Narendra Singh Yadav and R.P.Yadav, “The Effects of Speed on the Performance of Routing Protocols in Mobile Ad-hoc Networks”, Int. Journal of Electronics, Circuits and Systems, Vol. 1, No.2, pp 79-84 (2009)
- [10] S. Corson, J. Macker, “Mobile ad hoc networking (MANET) :Routing protocol performance issues and evaluation considerations”, Internet Draft(1999)
- [11] Guangyu pei, Mario Gerla, Tsu-Wei Chen, “Fisheye State Routing in Mobile Ad Hoc Networks”, Proc. Of IEEE ICC’00 (2000)
- [12] Yudhvirsingh, Yogesh Chaba, Monika Jain and Prabha Rani “Performance Evaluation of On-Demand Multicasting Routing Protocols in Mobile Adhoc Networks” IEEE International Conf. on Recent Trends in Information,Telecomm and Computing 2010.
- [13] Samir R.Das Charles E.Perkins and Elizabeth M.Royer “Performance Comparison of Two On-demand Routing Protocols for Ad Hoc Networks” IEEE INFOCOM 2000.
- [14] Mario Gerla, Xiaoyan Hong, Guangyu Pei, “Fisheye State Routing Protocol (FSR) for Ad Hoc Networks”, INTERNET-DRAFT-<draft-ietf-manet-fsr-03.txt> (2002)
- [15] Mehran Abolhasan and Tadeusz Wysocki “Displacement-based Route update strategies for proactive routing protocols

- in mobile ad hoc networks” International Workshop on the Internet, Telecommunications and Signal processing(2003).
- [16] UCLA Parallel computing laboratory, University of California, About GloMoSim, September 2004  
<http://pcl.cs.ucla.edu/projects/glomosim/>.

# Enhanced Stereo Matching Technique using Image Gradient for Improved Search Time

Pratibha Vellanki<sup>1</sup>, Dr. Madhuri Khambete<sup>2</sup>

<sup>1</sup> ENTC Department, Cummins College of Engineering for Women, Pune University,  
Pune, India.

<sup>2</sup> Principal, Cummins College of engineering for Women, Pune University,  
Pune, India.

## Abstract:

Stereo matching algorithms developed from local area based correspondence matching involve extensive search. This paper presents a stereo matching technique for computation of a dense disparity map by trimming down the search time for the local area based correspondence matching. We use constraints such as epipolar line, limiting the disparity, uniqueness and continuity to obtain an initial dense disparity map. We attempt to improvise this map by using color information for matching. A new approach has been discussed which is based on the extension of the continuity constraint for reducing the search time. We use correspondence between rows and gradient of image to compute the disparity. Thus we achieve a good trade off between accuracy and search time.

**Keywords:** *Color stereo matching, Image Gradient, Continuity constraint, Epipolar geometry, Stereo Vision.*

## I. Introduction:

Stereo correspondence for dense disparity estimation has been one of the most researched topics in computer vision. Dense surface information is required in 3D reconstruction. The determination of corresponding matches of an object between the left image and the right image is called correspondence. If this correspondence is solved for each pixel then it results in a dense disparity computation. The motivation behind this dense matching is that almost all the image pixels can be matched. Correspondence is an essential problem in dense stereo matching. For the computation of a reliable dense disparity map the stereo algorithm must preserve discontinuities in depth and also avoid gross errors.

Traditional dense matching techniques are divided into two types; local window based matching and global optimization method. Local window based matching compares intensity similarity of neighborhood of the corresponding points to be matched. A cost parameter is used to decide the best match. In this approach the selection of appropriate window size is critical to achieve a smooth and detailed disparity map. The 2<sup>nd</sup> approach is the global optimization algorithm which optimizes a certain disparity function and the smoothness constraint item to solve the matching problem.

The technique of matching points by correlation uses two windows: a fixed window centered at the pixel of interest in the reference image and a slippery window that browses the search zone [1]. It is important to select two optimal parameters, which are: window size  $n \times n$  and cost parameter. The window size selection depends on the local variation in texture and disparity. Generally a small window is used for unwanted smoothing, but in areas with low texture it doesn't have enough intensity variation for reliable matching. On the other hand if the disparity varies within the window then intensity values may not correspond due to projective distortions [2].

The design of the cost parameter decides the speed of implementing the stereo algorithm. The cost parameters generally used in area matching are Sum of absolute differences (SAD), Sum of squared differences (SSD), Zero mean normalized cross correlation (ZNCC), Zero mean sum of absolute differences (ZSAD). SAD and SSD are the most popular functions due to their simplicity [3]. We have used SAD as the cost function in a  $3 \times 3$  window.

A. The matching constraints:

Some of the common constraints used for matching in stereo correspondence are as explained below:

- Epipolar constraint: Corresponding points must lie on corresponding epipolar lines.
- Continuity constraint: Disparity tends to vary slowly across a surface
- Uniqueness constraint: A point in one image should have at the most one corresponding match in the other image.
- Ordering constraint: the order of features along epipolar lines is the same.
- Occlusion constraint: discontinuity in one eye corresponds to occlusion in other eye and vice versa.

The motive of this paper is to improve area based correspondence in two aspects: accuracy and search time. To improve accuracy we have proposed to use color information for matching. The color makes matching less sensitive to occlusion considering the fact that occlusion often causes color discontinuities [4]. Thus all the images used in our algorithm are color images.

To make the stereo algorithm fast we have proposed to use inter-row dependency as an assumption. This assumption is based on the fact that, the disparities on the current row will be similar to their neighbors in the previous row unless the top face of a new surface or a new object (i.e. discontinuity) is starting exactly at the pixel of interest. Based on this assumption we have modified the area based matching algorithm to obtain results in less search time.

Even though a general problem of finding correspondences between images involves the search within the whole image, once a pair of stereo images is rectified so that the epipolar lines are horizontal scan lines, a pair of corresponding edges in the right and left images should be searched for only within the same horizontal scanlines. Thus we have used rectified images as inputs to our algorithm.

## II. Color Information for matching:

There are many motivations behind using color information in stereo correspondence. Firstly, chromatic information is precisely obtained from CCD sensors of digital cameras. Secondly, recent developments in this area have proved that chromatic information plays an important role in human stereopsis. Thirdly, it is obvious that a red pixel cannot match with a green or blue pixel even if their intensities are same. Thus color information will potentially improve the performance of the matching algorithm.

The color space used here is RGB and the metric used is MSE. For color images we use MSE, defined as:

$$MSE_{color}(x, y, d) = \quad (1)$$

$$(1/n^2) \sum_{i=-k}^k \sum_{j=-k}^k \text{dist}(C_L(x+i, y+j), C_R(x+i, y+j+d))$$

$$\text{dist}(c^1, c^2) = (R^1 - R^2)^2 + (G^1 - G^2)^2 + (B^1 - B^2)^2 \quad (2)$$

In eq(1) and eq(2)  $d$  is the disparity and  $C^1$  and  $C^2$  are two points corresponding to the left and right images  $C_L$  and  $C_R$ , defined as:

$$C^1 = (R^1, G^1, B^1), \quad C^2 = (R^2, G^2, B^2) \quad (3)$$

The MSE is calculated using a 3 x 3 window and the left and right color spaces are defined as:

$$C_L(x, y) = (R_L(x, y), G_L(x, y), B_L(x, y)) \quad (4)$$

$$C_R(x, y) = (R_R(x, y), G_R(x, y), B_R(x, y)) \quad (5)$$

## III. Inter-row dependency with gradient information :

As explained in the introduction the proposed algorithm is based on the assumption that in an image generally there is a background and there are objects placed on the background. Thus it is obvious that the column discontinuities are more than the row discontinuities. Based on this explanation we have modified the program such that the search zone for the pixel match depends on the disparity of its neighbor in the row just above. Except if there exists a column discontinuity then the algorithm will search the complete search zone for the perfect match. This is where the image gradient comes into picture. The column discontinuity is detected by computing the gradient in the column direction.

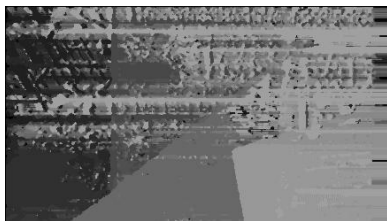
The search window used without using inter-row dependency is 20. As the maximum value of true disparity for the test images used is not more than 20, the purpose is solved. With inter-row dependency, we limit the research window to -5 to +5 range of the disparity of the pixel just above the reference pixel. This reduces the research window by 50%. The search zone is further reduced when the gradient in the column direction is used, being -3 to +3 range of the disparity of the pixel just above the reference pixel. The objective to reduce the search time is thus satisfied with this method.

#### IV. Results and Conclusion:

The stereo pair images used as inputs were obtained from middlebury university database. The images are rectified and thus satisfy the epipolar constraint and the intensity assumption. After obtaining the disparity map, median filtering is used to find the disparity of unmatched pixels. Median filter also discards any singular errors and makes the disparity map smooth. The results are as follows:



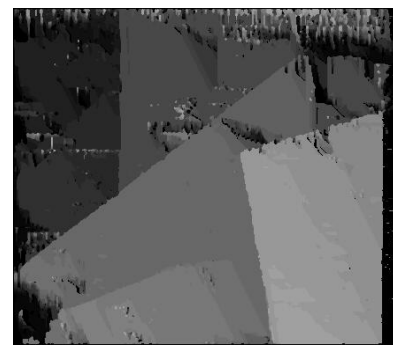
a) The right image of the stereo pair



b) Output of traditional local area based correspondence algorithm for grayscale images



c) Output of traditional local area based correspondence algorithm for color images



d) Output of inter-row dependency algorithm

The following is the table of results which comprises of the percentage of matched pixels for the traditional



methods: local area correspondence algorithm for grey and for color images and our proposed algorithm: inter-row dependency algorithm respectively; for a set of 5 images.

**Table I**

Images	Grey	Color	Inter-row
Barn2	90.68%	97.02%	90.13%
Poster	89.70%	96.37%	88.48%
Venus	88.38%	96.44%	88.47%
Tsukuba	89.79%	93.93%	85.72%
Sawtooth	92.76%	96.98%	91.65%

The average search time for a 383 x 434 image in case of local area based correspondence algorithm is 220 seconds while in case if our inter row dependency algorithm is 30 seconds. From the table I, the percentage of matched pixels of our algorithm is almost the same as the traditional algorithm for grayscale images. From these factors we can conclude that our algorithm achieves good trade off between accuracy and search time.

- [1] Mohammed Rziza, Ahmed Tamtaoui, Luce Morin and Driss Aboutajdine, "Estimation and segmentation of a Dense Disparity Map for 3D Reconstruction" IEEE Transaction, 2000.
- [2] D. Scharstein and R. Szeliski. "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms" *International Journal of Computer Vision*, 47(1/2/3):7-42, April-June 2002.
- [3] Jinhai Cai, "Fast Stereo Matching: Coarser to Finer with Selective Updating", Proceedings of Image and Vision Computing New Zealand 2007, pp. 266-270, Hamilton, New Zealand, December 2007.
- [4] Hajar Sadeghi, Payman Moallem and S. Amirhassn Monadjemi, "Feature Based Dense Stereo Matching using Dynamic Programming and Color", International Journal of Information and Mathematical Sciences, 2008.
- [5] Jiang Ze-tao, Zheng Bi-na, Wu Min and Chen zhong-xiang, "A 3D Reconstruction Method Based on Images Dense Stereo Matching", IEEE Proceedings of International Conference on Genetic and Evolutionary Computing, 2009.
- [6] Lu Yang, Rongben Wang, Pingshu Ge, Fengping Cao, "Research on Area-Matching, Algorithm Based on Feature-Matching Constraints" IEEE Proceedings of 2009 fifth International Conference on Natural Computation.

## References:

# Analyzing the Impact of Scalability on QoS-aware Routing for MANETs

Rajneesh Kumar Gujral<sup>1</sup>, Manpreet Singh<sup>2</sup>

<sup>1</sup>Assoc. Professor, Computer Engineering Department, M. M. Engineering College, M. M. University, Ambala, Haryana, India - 133207.

<sup>2</sup> Professor, Computer Engineering Department, M. M. Engineering College, M. M. University, Ambala, India - 133207.

## Abstract

Mobile Ad hoc networks (MANETs) are self-created and self organized by a collection of mobile nodes, interconnected by multi-hop wireless paths in a strictly peer to peer fashion. Scalability of a routing protocol is its ability to support the continuous increase in the network parameters (such as mobility rate, traffic rate and network size) without degrading network performance. The goal of QoS provisioning is to achieve a more deterministic network behaviors, so that information carried by the network can be better delivered and network resources can be better utilized. In this paper, we are going to analyze the impact of scalability on various QoS Parameters for MANETs routing protocols one proactive protocol (DSDV) and two prominent on-demand source initiated routing protocols. The performance metrics comprises of QoS parameters such as packet delivery ratio, end to end delay, routing overhead, throughput and jitter. The effect of scalability on these QoS parameters is analyzed by varying number of nodes, packet size, time interval between packets and mobility rates.

**Keywords:** MANETs, Scalability, QoS, Routing Protocols.

## 1. Introduction

Mobile Ad hoc networks (MANETs) are self-created and self organized by a collection of mobile nodes, interconnected by multi-hop wireless paths in a strictly peer to peer fashion [1]. The increase in multimedia, military application traffic has led to extensive research focused on achieving QoS guarantees in current networks. The goal of QoS provisioning is to achieve a more deterministic network behaviors, so that information carried by the network can be better delivered and network resources can be better utilized. The QoS parameters differ from application to application e.g., in case of multimedia application bandwidth, delay jitter and delay are the key QoS parameters [2]. After receiving a QoS service request, the main challenges is routing with scalable performance in deploying large scale MANETs. Scalability can refer to the capability of a system to increase total throughput

under an increased load [3]. Many protocols have been proposed but a few comparisons have been made with respect to scalability. The routing protocols Dynamic Source Routing (DSR), Ad hoc On-demand Distance Vector (AODV) and Temporally Ordered Routing Algorithm (TORA) protocol had been analyzed theoretically and through simulation using an Optimized Network Engineering Tools (OPNET) by varying node density and number of nodes [4].

The effect of scalability of a network on Genetic Algorithm based Zone Routing Protocols by varying the number of node is analyzed in [5]. In [6], simulation have been conducted to investigate scalability of DSR, AODV and LAR routing protocols using prediction based link availability model. Simulation results of the modified DSR (MDSR) as proposed in [7] has less overhead and delay as compared to conventional DSR irrespective of network size. In [8] simulation based comparative study of AODV, DSR, TORA and DSDV was reported which highlighting that DSR and AODV achieved good performance at all mobility speed whereas DSDV and TORA perform poorly under high speeds and high load conditions respectively. In [9] showed the proactive protocols have the best end-to-end-delay and packet delivery fraction but at the rate of higher routing load. In [10] three routing protocols were evaluated in a city traffic scenarios and it was shown that AODV outperforms both DSR and the proactive protocol FSR. In [11] simulation study of AODV, DSR and OLSR was done which shown that AODV and DSR outperform OLSR at higher speeds and lower number of traffic streams and OLSR generates the lowest routing load. In [12] more limited study was conducted which favoring DSR in terms of packet delivery fraction and routing overhead whereas OLSR shows the lowest end-to-end delay at lower network loads. In [13] simulation based performance comparison on DSDV, AODV and DSR is

done on the basis of Packet delivery ratio, Throughput, End to End delay & routing overhead by varying packet size, time interval between packet sending & mobility of nodes on 25 nodes using NS2.34. In [14] author performed realistic comparison between two MANETs protocols namely AODV (reactive protocol) and DSDV (proactive protocol). It is analyzed that the performance of AODV protocol is better than the DSDV protocol in term of PDF, Average end-to-end delay, packet loss and routing overhead by taking fixed number of nodes and varying number of nodes which helps in improving scalability of MANETs. In [15] author evaluated the scalability of on-demand ad hoc routing protocols by taking of up to 10,000 nodes. To improve the performance of on-demand protocols in large networks, five modification combinations have been separately incorporated into an on-demand protocol, and their respective performance has been studied. It has been shown that the use of local repair is beneficial in increasing the number of data packets that reach their destinations. Expanding ring search and query localization techniques seem to further reduce the amount of control overhead generated by the protocol, by limiting the number of nodes affected by route discoveries. While the performance improvements of the modifications have only been demonstrated with the AODV protocol. In [16] author proposed an effective and scalable AODV (called as AODV-ES) for Wireless Ad hoc Sensor Networks (WASN) by using third party reply model, n-hop local ring and time-to-live based local recovery. The above said work goal is to reduce time delay for delivery of the data packets, routing overhead and improve the data packet delivery ratio. The resulting algorithm "AODV-ES" is then simulated by NS-2 under Linux operating system. The performance of routing protocol is evaluated under various mobility rates and found that the proposed routing protocol is better than AODV. In [17] moreover, most of current routing protocols assume homogeneous networking conditions where all nodes have the same capabilities and resources. Although homogenous networks are easy to model and analysis, they exhibits poor scalability compared with heterogeneous networks that consist of different nodes with different resources. The author studies simulations for DSR, AODV, LAR1, FSR and WRP in homogenous and heterogeneous networks. The results showed that these which all protocols perform reasonably well in homogenous networking conditions, their performance suffer significantly over heterogonous networks

In this paper, the impact of scalability on QoS Parameters such as packet delivery ratio, end to end delay, routing overhead, throughput and jitter has been analyzed by varying number of nodes, packet size, time interval between packets & mobility rates. The rest of paper is organized as follow. In section 2, gives an overview of routing protocols, section 3 describe the performance

metrics, Section 4 simulation results and analysis are discussed and section 5 concludes the paper.

## 2. Overview of Routing Protocols

Routing protocols for MANETs have been classified according to the strategies of discovering and maintaining routes into three classes: proactive, reactive and Hybrid [18]

**Destination-Sequenced Distance Vector (DSDV):** DSDV is a table-driven routing [9] scheme for MANETs. The Destination-Sequenced Distance-Vector (DSDV) Routing Algorithm is based on the idea of the classical Bellman-Ford Routing Algorithm with certain improvements. Every mobile station maintains a routing table that lists all available destinations, the number of hops to reach the destination and the sequence number assigned by the destination node. The sequence number is used to distinguish stale routes from new ones and thus avoid the formation of loops.

**Dynamic Source Routing (DSR):** is an on-demand protocol designed to restrict the bandwidth consumed by control packets in ad hoc wireless networks by eliminating the periodic table-update messages required in the table-driven approach [19]. The major difference between this and other on-demand routing protocols is that it is beaconless and hence does not require periodic hello packet (beacon) transmission, which are used by a node to inform its neighbors of its presence. The basic approach of this protocol (and all other on-demand routing protocols) during the route construction phase is to establish a route by flooding Route Request packets in the network. The destination node, on receiving a Route Request packet, responds by sending a Route Reply packet back to the source, which carries the route traversed by the Route Request packet received.

**Ad hoc On-demand Distance Vector (AODV):** AODV routing protocol is also based upon distance vector, and uses destination numbers to determine the freshness of routes. AODV minimizes the number of broadcasts by creating routes on-demand as opposed to DSDV that maintains the list of the entire routes. To find a path to the destination, the source broadcasts a route request packet. The neighbors in turn broadcast the packet to their neighbors till it reaches an intermediate node that has recent route information about the destination or till it reaches the destination. A node discards a route request packet that it has already seen. The route request packet uses sequence numbers to ensure that the routes are loop free and to make sure that if the intermediate nodes reply to route requests, they reply with the latest information only.

## 3. QoS Based Performance Metrics

The performance metrics includes the following QoS parameters such as PDR (Packet Delivery Ratio),

Throughput, End to End Delay, Routing overhead and Jitter.

**Packet Delivery Ratio (PDR):** also known as the ratio of the data packets delivered to the destinations to those generated by the CBR sources. This metric characterizes both the completeness and correctness of the routing protocol also reliability of routing protocol.

$$PDR = \frac{\sum_{i=1}^n CBR_{re}ce}{\sum_{i=1}^n CBR_{se}nt} * 100$$

**Average End to End Delay:** Average End to End delay is the average time taken by a data packet to reach from source node to destination node. It is ratio of total delay to the number of packets received.

$$Avg\_End\_to\_End\_Delay = \frac{\sum_{i=1}^n (CBR_{re}cetime - CBR_{se}ntime)}{\sum_{i=1}^n CBR_{re}ce} * 100$$

**Throughput:** Throughput is the ratio of total number of delivered or received data packets to the total duration of simulation time.

$$Throughput = \frac{\sum_{i=1}^n CBR_{re}ce}{simulation\ time}$$

**Normalized Protocol Overhead/ Routing Load:** Routing Load is the ratio of total number of the routing packets to the total number of received data packets at destination.

$$Routing\_Load = \frac{\sum_{i=1}^n RTRPacket}{\sum_{i=1}^n CBR_{re}ce}$$

**Jitter:** Jitter describes standard deviation of packet delay between all nodes.

#### 4. Simulation Results and Analysis

The performance of QoS parameters on routing protocols AODV, DSR and DSDV is simulated using NS-2.34. The parameters used for simulation and different scenario on which they are analyzed are shown in Table 1 and Table 2 respectively. The positioning and communication among nodes is represented in Figure 1.

Table 1 Simulation Parameters

Parameters	Value
No of Node	25,50,75,100
Simulation Time	10 sec
Environment Size	1200x1200
Traffic Size	CBR (Constant Bit Rate)
Packet Size	500 and 1000 bytes
Queue Length	50
Source Node	Node 0
Destination Node	Node 2
Mobility Model	Random Waypoint
Antenna Type	Omni directional
Simulator	NS-2.34
Mobility speed	1000,2000 m/s
Packet Interval	0.015,0.15 ns

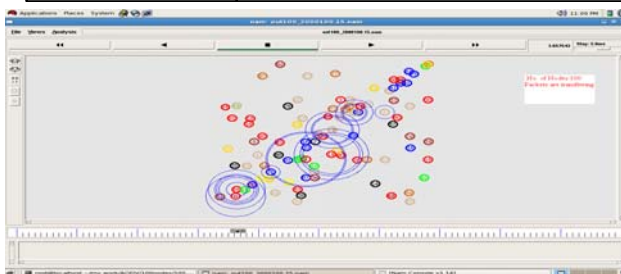


Figure 1. (Simulation Showing Packets transferring)

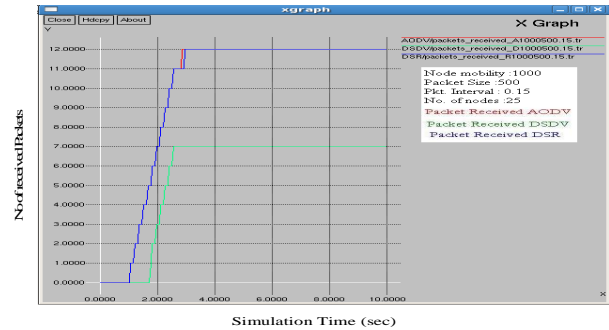


Figure 2(Packets Received when number of nodes=25 packet size=500 bytes, interval=0.15 sec Mobility=1000)

In scenario 01, Figure 2 shows that packet received in AODV and DSR is higher as compared to DSDV. The result in Table 3 shows that PDR, throughput, end to end delay is same in AODV and DSR is better than DSDV. Routing load is minimum in AODV. Jitter is less in DSDV as compared to AODV and DSR but throughput and PDR is also very low.

Table 3 (Performance Matrix number of nodes=25 packet size=500 bytes, interval=0.15 sec Mobility=1000)

Table-3	Packets Sent/Received	PDR	End-End Delay	Throug hput	Routi ng Load	Jitter (sec)
AODV	60/12	20.00	1.84	1.33	7.08	140.67
DSDV	60/7	11.66	2.07	0.77	8.57	106.87
DSR	60/12	20.00	1.85	1.33	20.41	147.88

Table -4 (Performance Matrix number of nodes=50 packet size=500 bytes , interval=0.15 sec Mobility=1000)

Table-4	Packets Sent/Received	PDR	End - End	Throug hput	Rou ting Loa	Jitter (sec)
AODV	60/56	93.33	5.61	6.22	5.08	155.88
DSDV	60/6	10.00	2.13	0.66	10.0	100.02
DSR	60/51	85.00	5.83	5.66	7.60	176.09

Table 2 shows different parameters taken for different simulation scenarios

Scenario no	No of nodes	Packets Size (bytes)	Packets Interval	Mobility(m/sec)
01	25,50,75,100	500	0.15 sec	1000
02	25,50,75,100	500	0.015 sec	1000
03	25,50,75,100	1000	0.15 sec	1000
04	25,50,75,100	1000	0.015 sec	1000
05	25,50,75,100	500	0.15 sec	2000
06	25,50,75,100	500	0.015 sec	2000
07	25,50,75,100	1000	0.15 sec	2000
08	25,50,75,100	1000	0.015 sec	2000

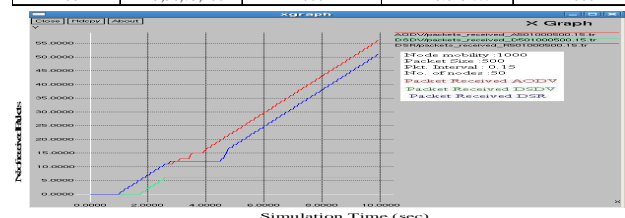


Figure 3 (Packets Received when number of nodes=50 packet size=500 bytes, interval=0.15 sec Mobility=1000)

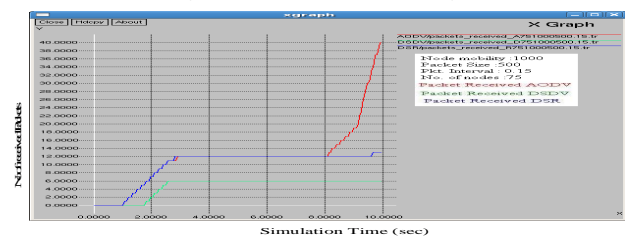


Figure 4 (Packets Received when number of nodes=75 packet size=500 bytes, interval=0.15 sec Mobility=1000)

Table 5(Performance Matrix number of nodes=75 packet size=500 bytes, interval=0.15 sec Mobility=1000)

Table-5	Packets Sent/Received	PDR	End-End Delay	Throug hput	Routin g Load	Jitter (sec)
AODV	60/42	70.00	7.16	4.66	7.80	281.66
DSDV	60/6	10.00	2.13	0.66	10.00	100.00
DSR	60/14	23.33	2.97	1.55	53.50	631.54

Table 6 (Performance Matrix number of nodes=100 packet size=500 bytes, interval=0.15 sec Mobility=1000)

Table-6	Packets Sent/Received	PDR	End - End	Throug hput	Routin g Load	Jitter (sec)
AODV	60/47	78.33	7.36	5.22	6.27	246.87
DSDV	60/7	11.66	2.06	0.77	8.57	107.03
DSR	60/31	51.66	6.10	3.44	18.61	363.61

Figure 3, 4 and 5 shows that number of packets received in AODV is more as compared to DSR and DSDV when numbers of nodes are scalable from 50, 75 and 100. AODV having that highest PDR and throughput with minimum routing load and jitter from DSR. We have also analyzed that in DSDV Jitter, end to end delay is low as compared to AODV and DSR but throughput, number of packets received and PDR is very low. The overall performance of AODV is best as four QoS parameters out of six has favourable results as indicated in Table 4, Table 5 and Table 6.

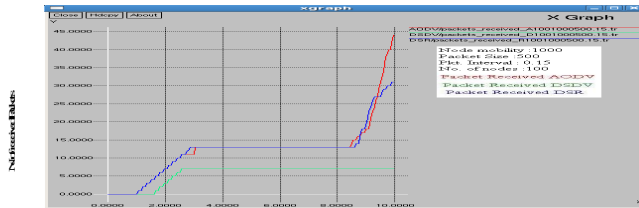


Figure 5 (Packets Received number of nodes=100 packet size=500 bytes, Interval=0.15 sec Mobility=1000)

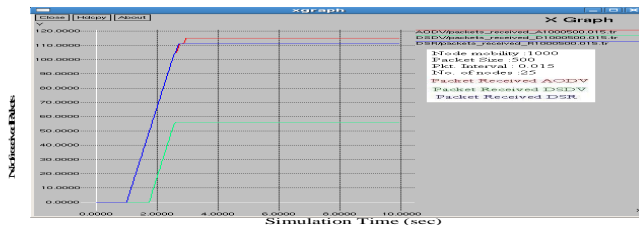


Figure 6 (Packets Received when number of nodes=25 packet size=500 bytes, interval=0.015 sec Mobility=1000)

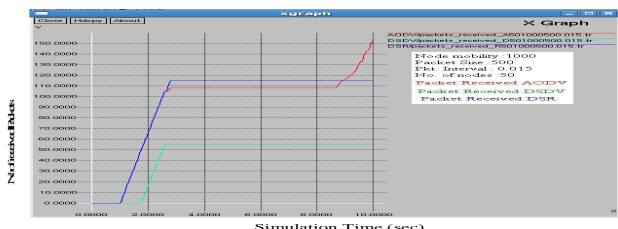


Figure 7 (Packets Received when number of nodes=50 packet size=500 bytes, interval=0.015 sec Mobility=1000)

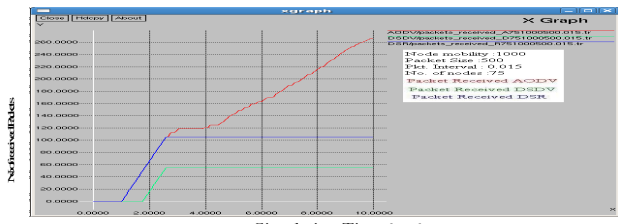


Figure 8 (Packets Received when number of nodes=75 packet size=500 bytes, interval=0.015 sec Mobility=1000)

Table 7 (Performance Matrix number of nodes=25 packet size=500 bytes, interval=0.015 sec Mobility=1000)

Table-7	Packets Sent/Received	PDR	End-End Delay	Throug hput	Routin g Load	Jitter
AODV	600/115	19.16	1.87	12.77	8.95	17.17
DSDV	600/56	9.33	2.15	6.22	10.71	14.46
DSR	600/111	18.50	1.83	12.33	15.98	15.57

Table 8 (Performance Matrix number of nodes=50 packet size=500 bytes, interval=0.015 sec Mobility=1000)

Table-8	Packets Sent/Received	PDR	End - End	Throug hput	Routin g Load	Jitter
AODV	600/154	25.66	4.09	17.	11.16	73.43
DSDV	600/55	9.16	2.16	6.11	10.90	14.24
DSR	600/115	19.16	1.86	12.77	15.08	16.81

Table 9 (Performance Matrix number of nodes=75 packet size=500 bytes, interval=0.015 sec Mobility=1000)

Table-9	Packets Sent/Received	PDR	End-End Delay	Throug hput	Routin g Load	Jitter
AODV	600/266	44.33	4.74	29.55	6.12	92.84
DSDV	600/55	9.16	2.16	6.11	10.90	14.27
DSR	600/105	17.50	1.78	11.66	16.39	14.66

Table 10 (Performance Matrix number of nodes=100 packet size=500 bytes, interval=0.015 sec Mobility=1000)

Table-10	Packets Sent/Received	PDR	End - End	Throug hput	Routin g Load	Jitter
AODV	600/208	34.66	4.64	23.11	8.45	89.00
DSDV	600/64	10.66	2.09	7.11	9.37	14.35
DSR	600/113	18.83	1.97	12.55	14.94	45.8

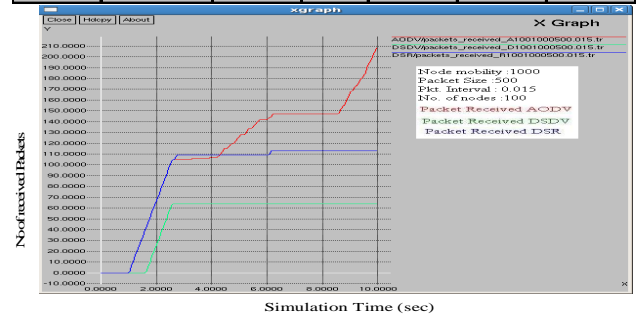


Figure 9(Packets Received number of nodes=100 packet size=500 bytes, interval=0.015 sec Mobility=1000)

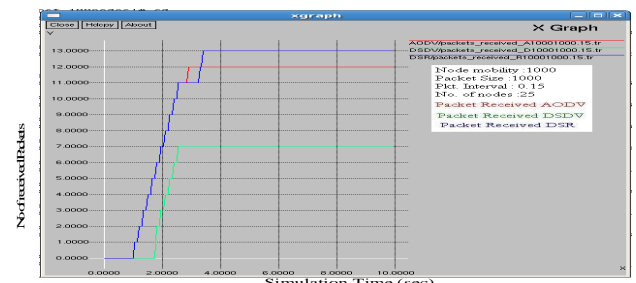


Figure 10(Packets Received number of nodes=25 packet size=1000 bytes, interval=0.15 sec Mobility=1000)

In scenario 02, Figure 6, 7, 8 and 9 shows that number of packets received in AODV is more as compared to DSR and DSDV, when numbers of nodes are scalable from 25, 50, 75 and 100. AODV is also having the highest PDR and throughput with minimum routing load and jitter relative to DSR. We have also analyzed that in DSDV, Jitter, end to end delay is low as compared to AODV and DSR but throughput, number of packets received and PDR is also on lower side. The overall performance of AODV is better, as four QoS parameters out of six has favourable results as indicated in Table 7, Table 8, Table 9 and Table 10.

Table 11(Performance Matrix number of nodes=25 packet size=1000 bytes, interval=0.15 sec Mobility=1000)

Table-11	Packets Sent/Received	PDR	End-End Delay	Throug hput	Routi ng Loa	Jitter
AODV	60/12	20.00	1.85	1.33	7.08	141.34
DSDV	60/7	11.66	2.08	0.77	8.57	106.66
DSR	60/13	21.66	1.99	1.44	23.15	156.70

Table 12(Performance Matrix number of nodes=50 packet size=1000 bytes, interval=0.15 sec Mobility=1000)

Table-12	Packets Sent/Received	PDR	End - End	Throug hput	Routi ng Loa	Jitter
AODV	60/54	90.00	5.75	6.00	5.20	202.22
DSDV	60/6	10.00	2.13	0.66	10.0	100.02
DSR	60/59	98.33	5.76	6.55	7.61	176.60

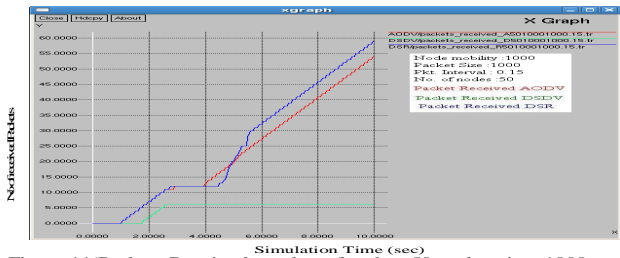


Figure 11(Packets Received number of nodes=50 packet size=1000 bytes, interval=0.15 sec Mobility=1000)

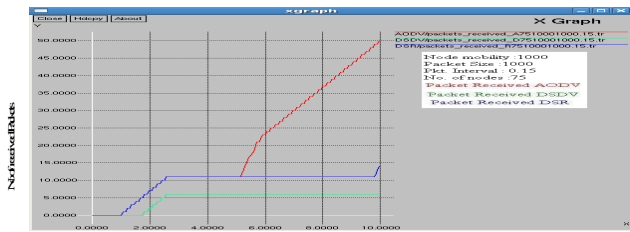


Figure 12(Packets Received number of nodes=75 packet size=1000 bytes, interval=0.15 sec Mobility=1000)

Table 13 (Performance Matrix number of nodes=75 packet size=1000 bytes, interval=0.15 sec Mobility=1000)

Table-13	Packets Sent/Received	PDR	End-End Delay	Throug hput	Routi ng Loa	Jitter
AODV	60/50	83.33	6.01	5.55	6.44	175.20
DSDV	60/6	10.00	2.13	0.66	10.00	100.00
DSR	60/14	23.33	3.50	1.55	39.57	626.18

Table 14(Performance Matrix number of nodes=100 packet size=1000 bytes, interval=0.15 sec Mobility=1000)

Table-14	Packet s Sent/Received	PDR	End-End Delay	Throug hput	Routin g Loa	Jitter
AODV	60/47	78.33	5.91	5.22	5.68	190.106
DSDV	60/7	11.66	2.06	0.77	8.57	107.03
DSR	60/31	51.66	5.74	3.44	15.32	411.98

In scenario 03, Figure 10 and 11 shows number of packets received in DSR are more in comparison with AODV and DSDV, when numbers of nodes are 25 and 50. The performance of DSR is also better for other QoS parameters with these numbers of nodes as depicted in Table 11 and Table 12. Figure 12 and 13 shows the number of received packets and performance of DSR degrades when number of nodes are increased to 75 and 100 as shown in Table 13 and Table 14.

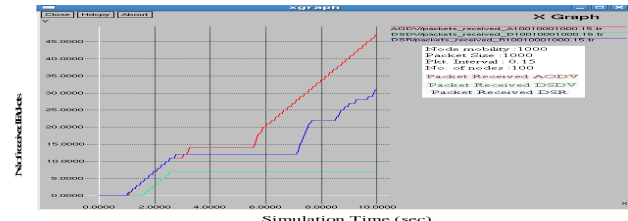


Figure 13(Packets Received number of nodes=100 packet size=1000 bytes, interval=0.15 sec Mobility=1000)

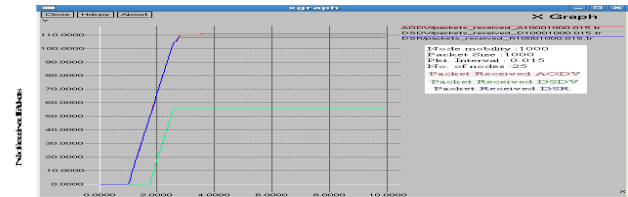


Figure 14(Packets Received number of nodes=25 packet size=1000 bytes, interval=0.015 sec Mobility=1000)

Table 15(Performance Matrix number of nodes=25 packet size=1000 bytes, interval=0.015 sec Mobility=1000)

Table-15	Packets Sent/Received	PDR	End-End Delay	Throug hput	Routing Loa	Jitter
AODV	600/111	18.50	1.84	12.33	6.46	22.62
DSDV	600/56	9.33	2.15	6.22	10.71	14.39
DSR	600/108	18.00	1.81	12.00	12.89	14.20

Table 16 (Performance Matrix number of nodes=50 packet size=1000 bytes, interval=0.015 sec Mobility=1000)

Table-16	Packets Sent/Received	PDR	End - End	Throug hput	Routing Loa	Jitter
AODV	600/142	23.66	3.05	15.77	8.76	89.68
DSDV	600/55	9.16	2.16	6.11	10.90	14.16
DSR	600/105	17.50	1.79	11.66	12.86	14.79

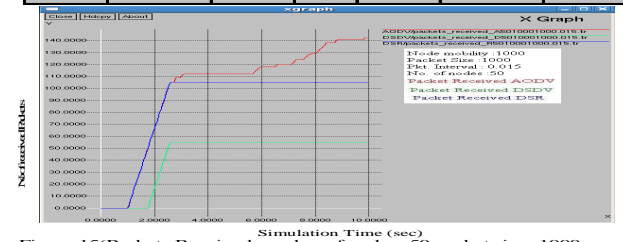


Figure 15(Packets Received number of nodes=50 packet size=1000 bytes, interval=0.015 sec Mobility=1000)

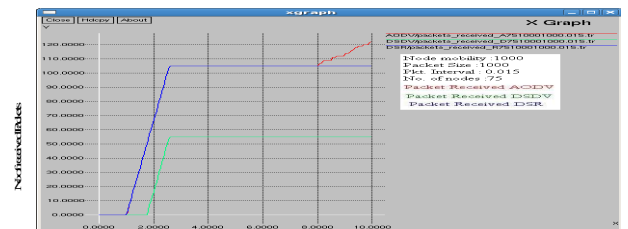


Figure 16(Packets Received number of nodes=75 packet size=1000 bytes, interval=0.015 sec Mobility=1000)

Table 17(Performance Matrix number of nodes=75 packet size=1000 bytes, interval=0.015 sec Mobility=1000)

Table-17	Packets Sent/Received	PDR	End-End Delay	Throug hput	Routi ng Loa	Jitter
AODV	600/123	20.50	2.84	13.66	10.31	83.63
DSDV	600/55	9.16	2.16	6.11	10.90	14.20
DSR	600/105	17.50	1.78	11.66	12.36	14.62

Table 18 (Performance Matrix number of nodes=100 packet size=1000 bytes, interval=0.015 sec Mobility=1000)

Table-18	Packets Sent/Received	PDR	End-End Delay	Throug hput	Routin g Loa	Jitter
AODV	600/171	28.50	3.63	19.00	6.65	72.67
DSDV	600/64	10.66	2.09	7.11	9.37	14.29
DSR	600/110	18.33	1.83	12.22	11.97	12.19



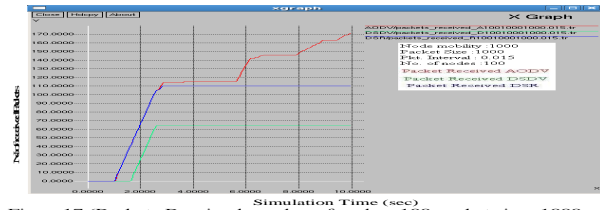


Figure 17 (Packets Received number of nodes=100 packet size=1000 bytes, interval=0.015 sec Mobility=1000)

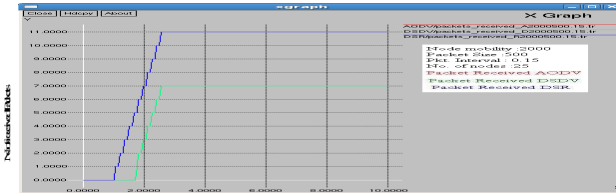


Figure 18 (Packets Received number of nodes=25 packet size=500 bytes, interval=0.15 sec Mobility=2000)

In scenario 04, Figure 14, 15, 16 and 17 shows that number of packets received in AODV is more as compared to DSR and DSDV, when numbers of nodes are scalable from 25, 50, 75 and 100. AODV is also having the highest PDR and throughput with minimum routing load and jitter relative to DSR. We have also analyzed that in DSDV, Jitter, end to end delay is low as compared to AODV and DSR but throughput, number of packets received and PDR is also on lower side. The overall performance of AODV is better, as four QoS parameters out of six has favourable results as indicated in Table 15, Table 16, Table 17 and Table 18.

Table 19 (Performance Matrix number of nodes=25 packet size=500 bytes, interval=0.15 sec Mobility=2000)

Table-19	Packets Sent/Received	PDR	End-End Delay	Throug hput	Routin g Load	Jitter
AODV	60/11	18.33	1.75	1.22	20.00	122.72
DSDV	60/7	11.66	2.07	0.77	8.57	106.87
DSR	60/11	18.33	1.75	1.22	8.09	122.72

Table 20 (Performance Matrix number of nodes=50 packet size=500 bytes, interval=0.15 sec Mobility=2000)

Table-20	Packets Sent/Received	PDR	End-End Delay	Throug hput	Routin g Load	Jitter
AODV	60/58	96.66	5.55	6.44	5.81	150.60
DSDV	60/6	10.00	2.13	0.66	10.00	100.02
DSR	60/52	86.66	5.82	5.77	8.05	171.75

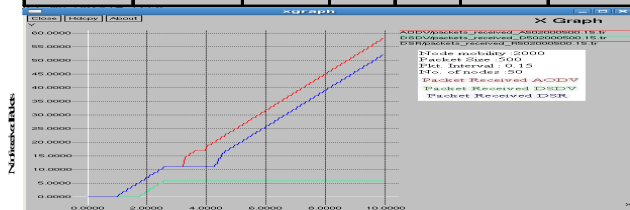


Figure 19 (Packets Received number of nodes=50 packet size=500 bytes, interval=0.15 sec Mobility=2000)

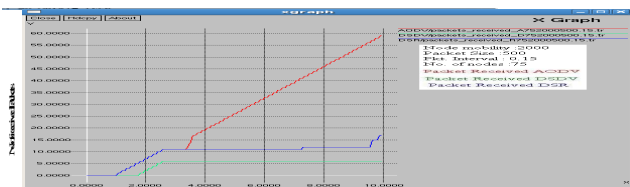


Figure 20 (Packets Received number of nodes=75 packet size=500 bytes, interval=0.15 sec Mobility=2000)

Table 21 (Performance Matrix number of nodes=75 packet size=500 bytes, interval=0.15 sec Mobility=2000)

Table-21	Packets Sent/Received	PDR	End-End Delay	Throug hput	Routin g Load	Jitter
AODV	60/59	98.33	5.53	6.55	5.89	148.04
DSDV	60/6	10.00	2.13	0.66	10.00	100.00
DSR	60/17	28.33	4.41	1.88	12.52	513.63

Table 22 (Performance Matrix number of nodes=100 packet size=500 bytes, interval=0.15 sec Mobility=2000)

Table-22	Packets Sent/Received	PDR	End-End Delay	Throug hput	Routin g Load	Jitter
AODV	60/58	96.66	5.57	6.44	5.36	150.50
DSDV	60/7	11.66	2.06	0.77	8.57	107.03
DSR	60/12	20.00	3.41	1.53	60.08	712.08

In scenario 05, Figure 18 shows, when number of nodes 25 the number of packets received in AODV and DSR equal, so its QoS parameters are almost same as depicted in Table 19. Figure 19, 20 and 21 shows when numbers of nodes are scalable from 50, 75 and 100 the number of received packets and performance of DSR degrades. The overall performance of AODV is best as four QoS parameters out of six has favourable results as indicated in Table 20, Table 21 and Table 22.

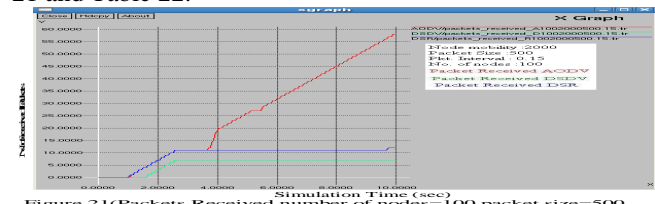


Figure 21 (Packets Received number of nodes=100 packet size=500 bytes, interval=0.15 sec Mobility=2000)

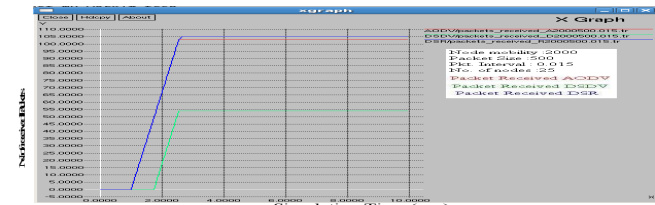


Figure 22 (Packets Received number of nodes=25 packet size=500 bytes, interval=0.015 sec Mobility=2000)

Table 23 (Performance Matrix number of nodes=25 packet size=500 bytes, interval=0.015 sec Mobility=2000)

Table-23	Packets Sent/Received	PDR	End-End Delay	Throug hput	Routin g Load	Jitter
AODV	600/103	17.16	1.77	11.44	12.01	14.69
DSDV	600/54	9.00	2.13	6.00	11.11	14.43
DSR	600/105	17.50	1.78	11.66	11.09	14.69

Table 24 (Performance Matrix number of nodes=50 packet size=500 bytes, interval=0.015 sec Mobility=2000)

Table-24	Packets Sent/Received	PDR	End-End Delay	Throug hput	Routin g Load	Jitter
AODV	600/104	17.33	1.77	11.55	17.00	15.50
DSDV	600/53	8.83	2.14	5.88	14.32	14.21
DSR	600/131	21.83	2.89	14.55	13.91	15.50

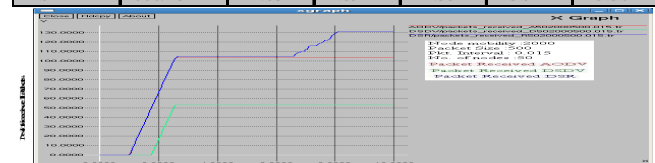


Figure 23 (Packets Received number of nodes=50 packet size=500 bytes, interval=0.015 sec Mobility=2000)

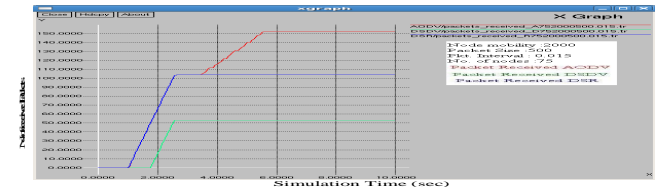


Figure 24 (Packets Received number of nodes=75 packet size=500 bytes, interval=0.015 sec Mobility=2000)

Table 25 (Performance Matrix number of nodes=75 packet size=500 bytes, interval=0.015 sec Mobility=2000)

Table-25	Packets Sent/Received	PDR	End-End Delay	Throughput	Routing Load	Jitter
AODV	600/152	25.33	2.64	16.88	11.38	38.67
DSDV	600/53	8.83	2.14	5.88	11.32	14.25
DSR	600/104	17.33	1.77	11.55	16.54	38.67

Table 26(Performance Matrix number of nodes=100 packet size=500 bytes, interval=0.015 sec Mobility=2000)

Table-26	Packets Sent/Received	PDR	End-End Delay	Throughput	Routing Load	Jitter
AODV	600/230	38.33	4.30	25.55	6.23	80.04
DSDV	600/62	10.33	2.07	6.88	9.67	14.33
DSR	600/104	17.33	1.77	11.55	17.09	80.04

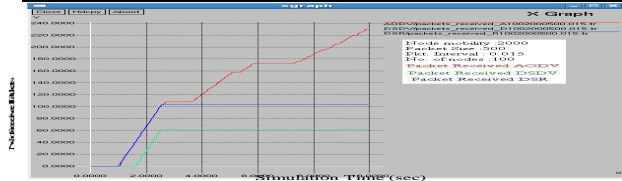


Figure 25(Packets Received number of nodes=100 packet size=500 bytes, interval=0.015 sec Mobility=2000)

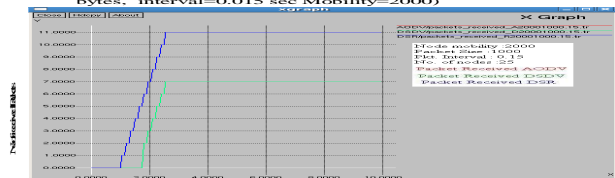


Figure 26(Packets Received number of nodes=25 packet size=1000 bytes, interval=0.15 sec Mobility=2000)

In scenario 06, Figure 22 and 23 shows number of packets received in DSR are more in comparison with AODV and DSDV, when numbers of nodes are 25 and 50. The performance of DSR is also better for other QoS parameters with these numbers of nodes as depicted in Table 23 and Table 24. Figure 24 and 25 shows the number of received packets and performance of DSR degrades when number of nodes are increased to 75 and 100 as shown in Table 25 and Table 26.

Table 27 (Performance Matrix number of nodes=25 packet size=1000 bytes, interval=0.15 sec Mobility=2000)

Table-27	Packets Sent/Received	PDR	End-End Delay	Throughput	Routing Load	Jitter
AODV	60/11	18.33	1.76	1.22	20.36	122.7
DSDV	60/7	11.66	2.08	0.77	8.57	106.6
DSR	60/11	18.33	1.76	1.22	9.18	122.7

Table 28 (Performance Matrix number of nodes=50 packet size=1000 bytes, interval=0.15 sec Mobility=2000)

Table-28	Packets Sent/Received	PDR	End-End Delay	Throughput	Routing Load	Jitter
AODV	60/58	96.66	5.57	6.44	5.81	151.02
DSDV	60/6	10.00	2.13	0.66	10.00	100.02
DSR	60/36	60.00	6.75	4.00	17.75	595.09

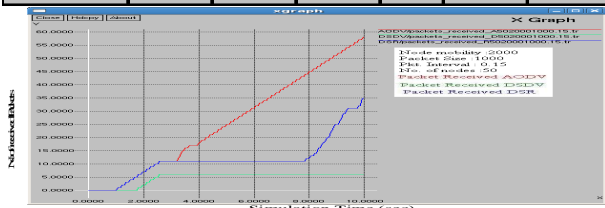


Figure 27(Packets Received number of nodes=50 packet size=1000 bytes, interval=0.15 sec Mobility=2000)

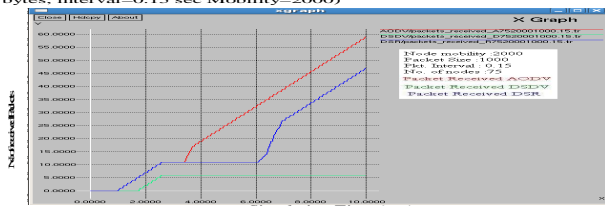


Figure 28(Packets Received number of nodes=75 packet size=1000 bytes, interval=0.15 sec Mobility=2000)

Table 29 (Performance Matrix number of nodes=75 packet size=1000 bytes, interval=0.15 sec Mobility=2000)

Table-29	Packets Sent/Received	PDR	End-End Delay	Throughput	Routing Load	Jitter
AODV	60/59	98.33	5.56	6.55	5.89	148.46
DSDV	60/6	10.00	2.13	0.66	10.00	100.00
DSR	60/47	78.33	6.25	5.22	9.36	212.20

Table 30(Performance Matrix number of nodes=100 packet size=1000 bytes, interval=0.15 sec Mobility=2000)

Table-30	Packets Sent/Received	PDR	End-End Delay	Throughput	Routing Load	Jitter
AODV	60/58	96.66	5.61	6.44	5.36	150.84
DSDV	60/7	11.66	2.06	0.77	8.57	107.03
DSR	60/22	36.66	5.43	2.44	20.90	618.23

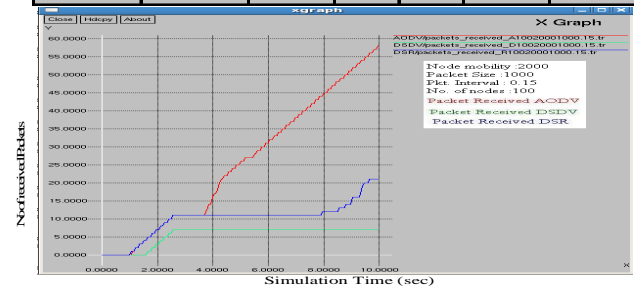


Figure 29(Packets Received number of nodes=100 packet size=1000 bytes, interval=0.15 sec Mobility=2000)

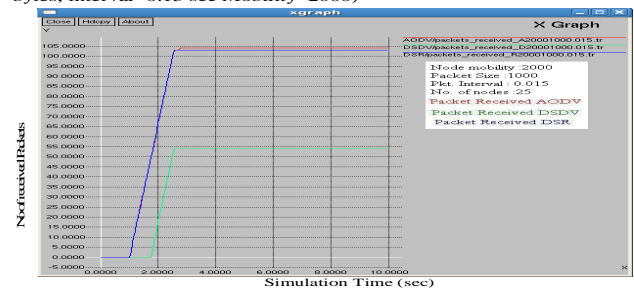


Figure 30(Packets Received number of nodes=25 packet size=1000 bytes, interval=0.015 sec Mobility=2000)

In scenario 07, Figure 26 shows, when number of nodes 25 the number of packets received in AODV and DSR equal, so its QoS parameters are almost same as depicted in Table 27. Figure 27, 28 and 29 shows when numbers of nodes are scalable from 50, 75 and 100 the number of received packets and performance of DSR degrades. The overall performance of AODV is best as four QoS parameters out of six has favourable results as indicated in Table 28, Table 29 and Table 30.

Table 31 (Performance Matrix number of nodes=25 packet size=1000 bytes, interval=0.015 sec Mobility=2000)

Table-31	Packets Sent/Received	PDR	End-End Delay	Throughput	Routing Load	Jitter
AODV	600/104	17.33	1.78	11.55	6.08	16.09
DSDV	600/54	9.00	2.14	6.00	11.11	14.36
DSR	600/103	17.16	1.77	11.44	11.51	14.64

Table 32 (Performance Matrix number of nodes=50 packet size=1000 bytes, interval=0.015 sec Mobility=2000)

Table-32	Packets Sent/Received	PDR	End-End Delay	Throughput	Routing Load	Jitter
AODV	600/194	32.33	4.07	21.5	6.41	158.26
DSDV	600/53	8.83	2.15	5.88	11.32	14.14
DSR	600/103	17.16	1.77	11.4	12.93	14.78

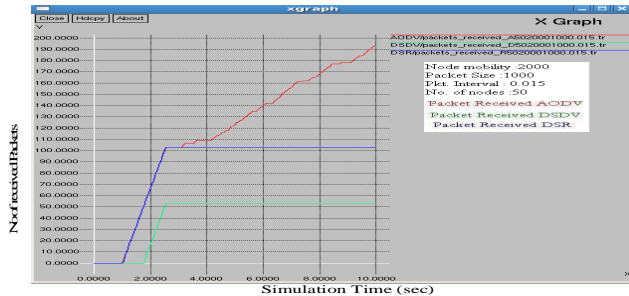


Figure 31(Packets Received number of nodes=50 packet size=1000 bytes, interval=0.015 sec Mobility=2000)

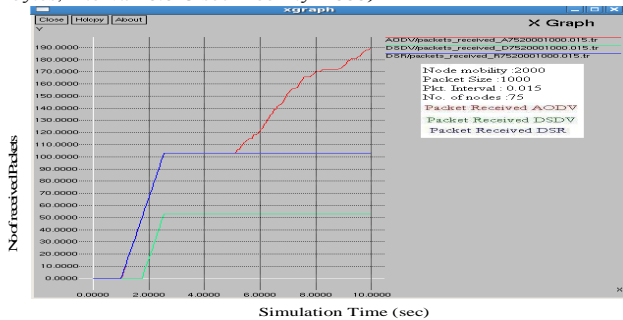


Figure 32(Packets Received number of nodes=75 packet size=1000 bytes, interval=0.015 sec Mobility=2000)

Table 33 (Performance Matrix number of nodes=75 packet size=1000 bytes, interval=0.015 sec Mobility=2000)

Table-33	Packets Sent/Received	PDR	End-End Delay	Througput	Routin g Load	Jitter
AODV	600/191	31.83	4.25	21.22	6.86	80.96
DSDV	600/53	8.83	2.15	5.88	11.32	14.17
DSR	600/103	17.16	1.77	11.44	13.17	14.61

Table 34(Performance Matrix number of nodes=100 packet size=1000 bytes, interval=0.015 sec Mobility=2000)

Table-34	Packets Sent/Received	PDR	End - End	Througput	Routin g Load	Jitter
AODV	600/182	30.33	3.83	20.22	7.03	98.19
DSDV	600/62	10.33	2.08	6.88	9.67	14.27
DSR	600/103	17.16	1.77	11.44	12.89	14.79

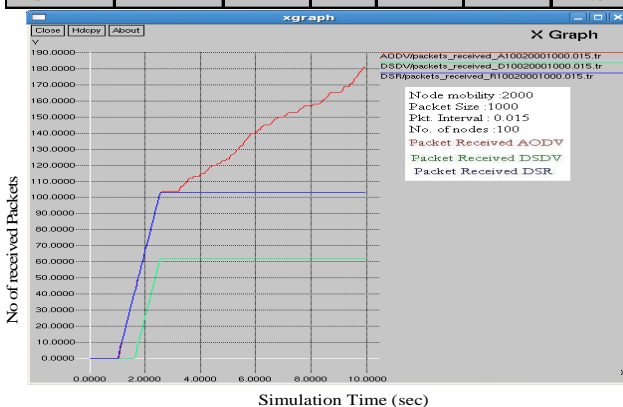


Figure 33(Packets Received number of nodes=100 packet size=1000 bytes, interval=0.015 sec Mobility=2000)

In scenario 08, Figure 30 shows, when number of nodes 25 the number of packets received in AODV and DSR equal, so its QoS parameters are almost same as depicted in Table 31. Figure 31, 32 and 33 shows when numbers of nodes are scalable from 50, 75 and 100 the number of received packets and performance of DSR degrades. The overall performance of AODV is best as four QoS

parameters out of six has favourable results as indicated in Table 32, Table 33 and Table 34.

## 5 Conclusions

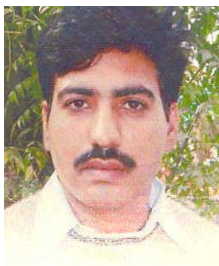
As observed by simulation from eight different scenarios, the AODV protocol is QoS-aware routing protocols under the effect of scalability in terms of variation in number of nodes, mobility rate and packet intervals. With the increase in network size, the performance of DSR decreases due to increase in packet-header overhead size as data and control packets in DSR typically carry complete route information. We have also analyzed that the performance of DSDV has not been affected by varying number of nodes, packet size, and mobility rate, its overall performance is less than AODV and DSR Protocol.

## References

- [1] Nadia Qasim, Fatin Said, Hamid Aghvami , “ Mobile Ad Hoc Networks Simulations Using Routing Protocols for Performance Comparisons”, Proceedings of the World Congress on Engineering 2008 Vol I WCE 2008, London, U.K, July 2 - 4, 2008.
- [2] Mr. Rajneesh Gujral, Dr. Anil Kapil “Secure QoS Enabled On-Demand Link-State Multipath Routing in MANETS” Proceeding of BAIP 2010, pp. 250-257 SPRINGER LNCS-CCIS, Trivandrum, Kerala, India, March 26-27, 2010.
- [3] C. Siva Ram Murthy & B.S. Manoj, Ad-hoc Wireless networks, Pearson
- [4] N.Adam, M.Y.Ismail, J. Abdullah, “Effect of node density on performances of three Manet routing protocols”, in International conference on Electronic devices, Systems and Applications 2010.
- [5] P. Sateesh Kumar, S. Ramachandram, "Scalability of Network Size on Genetic Zone Routing Protocol for MANETs," icacte, International Conference on Advanced Computer Theory and Engineering, pp.763-767, 2008.
- [6] Prashant Singh, Daya Krishan Lobiyal, "Scalability of Routing in MANET," iccsn, Second International Conference on Communication Software and Networks, pp.325-328, 2010.
- [7] M. Tamilarasi, V.R. Shyam Sunder, Udara Milinda Haputhanthri, Chamath Somathilaka, Nannuri Ravi Babu, S. Chandramathi, T.G. Palanivelu, "Scalability Improved DSR Protocol for MANETs," iccima, vol. 4, pp.283-287, 2007.
- [8] J Broch, D. Maltz, D. Johnson, Y.C.Hu, and J. Jetcheva, “ A performance comparison of multihop wireless ad hoc network routing protocols,” MOBICOM'98, pp.85-97, October 1998.
- [9] S.R. Das, R. Castaneda, J. Yan, R. Sengupta, “Comparative Performance Evaluation of Routing Protocols for Mobile Ad hoc Networks,” Proceedings of the international Conference On Computer Communications and Networks (ICCCN 1998), pp.153-161, 1998.
- [10] S.Jaap, M.Bechler and L. Wolf, “Evaluation of Routing Protocols for Vehicular Ad Hoc network in city traffic scenarios,” International Conference on ITS Telecommunications, France, 2005.
- [11] T. Clausen, P. Jacquet, L. Viennot, “ Comparative study of Routing Protocols for Mobile Ad Hoc Networks,” The first

Annual Mediterranean Ad Hoc Networking Workshop, September 2002.

- [12] J. Novatnack, L. Greenwald, H. Arora “Evaluating Ad hoc Routing Protocols With Respect to Quality of Service,” Wireless and Mobile Computing Networking and Communications WiMob’2005.
- [13] Mr. Rajneesh Gujral, Dr. Anil Kapil “Comparative Performance Analysis of QoS-Aware Routing on DSDV, AODV and DSR Protocols in MANETs” Communications in Computer and Information Science, 1, volume 101, Information and Communication Technologies, Part 3, pp. 610-615, 2010.
- [14] Asha Ambhaikar, Debashmita Mitra, Rajesh Deshmukh “Performance of MANET Routing Protocol for Improving Scalability” International Journal of Advanced Engineering & Application, pp. no- 15-18 Jan 2011.
- [15] Sung-Ju Lee, Elizabeth M. Belding-Royer, and Charles E. Perkins “Scalability Study of the Ad hoc On Demand Distance Vector Routing Protocol” International Journal of Network Management Vol. 13, pp no.97–114, 2003.
- [16] Debajyoti Mishra, Ashima Rout and Srinivas Sethi. Article: An Effective and Scalable AODV for Wireless Ad hoc Sensor Networks. International Journal of Computer Applications 5(4):33–38, August 2010.
- [17] Huda AlAmri, Mehran Abolhasan, Tadeusz A. Wysocki “Scalability of MANET routing protocols for heterogeneous and homogenous networks” Journal of Computers & Electrical Engineering-CEE, vol. 36, no. 4, pp. 752-765, 2010.
- [18] Abolhasan, M., T. Wysocki, and E. Dutkiewicz, “A review of routing protocols for mobile ad hoc networks” Elsevier journal of Ad hoc networks, 12(1): pp. 1-22, 2004.
- [19] David .B Johnson, David. A. Maltz, and Josh Broch, “Dynamic Source Routing protocol for Multihop Wireless Ad Hoc Networks,” In Ad Hoc Networking, edited by Charles E. Perkins Addison-Wesley., chapter 5, pp. 139-172., 2001.



**First Author** Rajneesh Kumar Gujral is working as Assoc. Professor in Department of Computer Science and Engineering, M.M Engineering College, M. M. University Mullana, Ambala. He obtained his BE (Computers) in 1999 from Punjab Technical University (PTU), Jalandhar. He also obtained his MTECH (IT) in 2007 from University School of Information Technology, GGSIP University Delhi. He has about 10 publications in International journals and Conferences. His research areas are Wireless communications which include Mobile Ad hoc and sensor based networks, Network Security and computer communication networks etc.



**Second Author** Dr. Manpreet Singh is working as Professor and Head of Department of Computer Science and Engineering, M.M engineering college, M. M. University Mullana, Ambala, India. He obtained his Ph.D., M.Tech. and B. Tech. from Kurukshetra University. He has about 25 publications in International journals and Conference. His research areas are Distributed Computing, Grid Computing, Ad hoc and sensor based networks, Distributed Database etc.



# Improving Data Association Based on Finding Optimum Innovation Applied to Nearest Neighbor for Multi-Target Tracking in Dense Clutter Environment

E.M.Saad<sup>1</sup>, El.Bardawiny<sup>2</sup>, H.I.ALI<sup>3</sup> and N.M.Shawky<sup>4</sup>

<sup>1</sup> Electronics and Communication Engineering Department, Helwan University,  
Cairo, Egypt

<sup>2</sup> Radar Department, M.T.C College  
Cairo, Egypt

<sup>3</sup> Electronics and Communication Engineering Department, Helwan University  
Cairo, Egypt

<sup>4</sup> Electronics and Communication Engineering Department, Helwan University  
Cairo, Egypt

## Abstract

In this paper, a new method, named optimum innovation data association (OI-DA), is proposed to give the nearest neighbor data association the ability to track maneuvering multi-target in dense clutter environment. Using the measurements of two successive scan and depending on the basic principle of moving target indicator (MTI) filter, the proposed algorithm avoids measurements in the gate size of predicted target position that are not originated from the target and detects the candidate measurement with the lowest probability of error. The finding of optimum innovation corresponding to the candidate valid measurement increases the data association performance compared to nearest neighbor (NN) filter. Simulation results show the effectiveness and better performance when compared to conventional algorithms as NNKF and JPDAF.

**Keywords:** Data Association, Multi-Target Tracking (MTT), Moving Target Indicator (MTI) Filter, Nearest Neighbor Kaman Filter (NNKF), Joint Probabilistic Data Association Algorithm (JPDA).

## 1. Introduction

Multiple-target tracking (MTT) is an essential component of surveillance systems. Real-world sensors; e.g., radar, sonar, and infrared (IR) sensors often report more than one measurement that may be from a given target. These may be either measurements of the desired target or Clutter

measurements. Clutter refers to detections or returns from nearby objects, clouds, electromagnetic interference, acoustic anomalies, false alarms, etc. A general formulation of the problem assumes an unknown and varying number of targets that are continuously moving in a given region. The states of these targets and the noisy measurements that are sampled by the sensor at regular time intervals (scan periods) are provided to the tracking system. When tracking a target in clutter, it is possible to have more than one measurement at any time since a measurement may have originated from either the target, clutter, or some other source. It is impossible to associate the target with a measurement perfectly. The performance of a tracking filter, however, relies heavily on the use of the correct measurement. In addition to the detection probability is not perfect and the targets may go undetected at some sampling intervals. A primary task of the MTT system is data association that is responsible for deciding on each scan which of the received multiple measurements that lie in the specified gate size of the predicted target position should update with the existing tracking target. The secondary goal is estimation of the number of targets and their position (states) based on the measurements originating from the targets of interest. In general, data association between measurements and targets is needed, but this is difficult to realize because of measurement error, false alarms, and missed targets. Due to the data association result is crucial for overall tracking process; a gating process is used to reduce the number of candidate

measurements to be considered. In data association process, the gating technique [1] in tracking a maneuvering target in clutter is essential to make the subsequent algorithm efficient but it suffers from problems since the gate size itself determines the number of valid included measurements. Another problem in case of tracking multiple targets, data association becomes more difficult because one measurement can be validated by multiple tracks in addition to a track validating multiple measurements as in the single target case. To solve these problems, the important of an alternative approaches known as nearest neighbor data association (NNDA) [2-5], probabilistic data association (PDA) [6,7], joint probabilistic data association (JPDA) [7,8], and multiple hypothesis Tracking (MHT) [9], etc. has been used to track multiple targets by evaluating the measurement to track association probabilities with different methods to find the state estimate [10-12]. NNDA that depends only on choosing the nearest valid measurement to the predicted target position, has been used in real work widely because of its low calculation cost, but it readily miss-tracks in dense cluttered environment. PDA, JPDA and MHT need prior knowledge and some of them have large calculation amount [13-16]. We propose here an extended algorithm applied to conventional NNDA to be able to track the multi-target in dense clutter environment. This proposed algorithm is more accurate to choose the true measurement originated from the target with lower probability of error and less sensitivity to false alarm targets in the gate region size than NNDA algorithm. Depending on the basic principle of moving target indicator (MTI) filter used in radar signal processing [16-20] which get rid from the fixed targets and the targets that moving with lower velocity and their moving distance lower than specified certain threshold value, the proposed algorithm reduces the number of candidate measurements in the gate by MTI filtering method that compares the moving distance measure for each measurement in the current gate at the update step to all previous measurement in the same gate at the predicted step and then avoids any measurement in the current gate moves a distance less than the threshold value due to comparison. Thus, decreasing the number of candidate measurements in the current gate lead to decreasing the probability of error in data association process. The main key to detect the moving or fixed false target is the innovation parameter that measure the moving distance between the current measurement and the predicted target position. By calculating this parameter for all measurement in the current gate compared with the scanned previous measurement in the same gate, the optimum innovation of the candidate measurement is obtained. This is called optimum innovation data association (OI-DA) method which is combined with NNDA algorithm to apply the proposed algorithm in multi

tracking targets in presence of various clutter densities. Simulation results showed better performance when compared to the two conventional NNKF, JPDA algorithm.

## 2. Background

### 2.1 Kalman Filter Theory

Based on Kalman filter estimation [21], we list the filter model. The dynamic state and measurement model of target  $t$  can be represented as follows

$$x^t(k) = A^t(k-1)x^t(k-1) + w^t(k-1) \quad t = 1, 2, \dots, T \quad (1)$$

$$z^t(k) = H^t(k)x^t(k) + v^t(k) \quad t = 1, 2, \dots, T \quad (2)$$

Where  $x^t(k-1)$  is the  $n \times 1$  target state vector. This state can include the position and velocity of the target in space  $x = (x, y, \dot{x}, \dot{y})'$ , The initial target state,  $x^t(0)$  for  $t = 1, 2, \dots, T$ , is assumed to be Gaussian With mean  $m_0^t$  and known covariance matrix  $p_0^t$ . Where the unobserved signal (hidden states)  $\{x^t(k) : k \in N\}$ ,  $x^t(k) \in X$  be modeled as a Markov process of transition probability  $p(x^t(k) | x^t(k-1))$  and initial distribution  $p(x^t(0)) = N(x^t(0); m_0^t, p_0^t)$ .  $z^t(k)$  is the  $m \times 1$  measurement vector,  $A^t(k-1)$  denotes state transition matrix,  $H^t(k)$  denotes measurement matrix,  $w^t(k-1)$  and  $v^t(k)$  are mutually independent white Gaussian noise with zero mean, and with covariance matrix  $Q(k-1)$  and  $R(k)$ , respectively.

The innovation mean (residual error) of measurement  $z_i(k)$  is given by

$$V_i^t(k) = z_i(k) - \hat{z}^t(k) \quad (3)$$

where

$$\hat{z}^t(k) = H^t(k)\bar{m}^t(k) \quad (4)$$

and the predicted state mean and covariance is defined as

$$\bar{m}^t(k) = A^t(k)m^t(k-1) \quad \text{and} \quad (5)$$

$$\bar{p}^t(k) = A^t(k)p^t(k-1)A^t(k)' + Q \quad (5)$$

Then, we can update state by



$$m^t(k) = \bar{m}^t(k) + K^t(k)V_{opt}(k) \quad (6)$$

where  $V_{opt}$  is the selected innovation mean from  $V_i^t(k)$  corresponding to the choosing measurement as a result of data association process,  $K^t(k)$  denotes gain matrix calculated by state error covariance  $p^t(k)$  and innovation covariance  $S^t(K)$ , their recursive equations can be represented as follows

$$p^t(k) = \bar{p}^t(k) - K^t(k)S^t(K)K^t(k)' \quad (7)$$

$$S^t(K) = H^t(k)\bar{p}^t(k)H^t(k)' + R(K) \quad (8)$$

$$K^t(k) = \bar{p}^t(k) - H^t(K)S^t(K)^{-1} \quad (9)$$

When multiple target tracking begins, we get for each target  $t$  measurements within correlation gate (gate size) as candidate measurements when  $z_i(k)$  satisfies condition

$$\left( z_i(k) - H^t(k)\bar{m}^t(k) \right)' S^t(k)^{-1} \left( z_i(k) - H^t(k)\bar{m}^t(k) \right) \leq \gamma \quad (10)$$

where  $\gamma$  denotes correlation gate. If there is only one measurement, this can be used for track update directly; otherwise if there is more than one measurement, we need to calculate the equivalent measurement.

## 2.2 Nearest Neighbor Kalman Filter

The NNKF is theoretically the most simple single-scan recursive tracking algorithm. The NNKF consists of a discrete-time Kalman filter (KF) together with a measurement selection rule. The NNKF takes the KF's state estimate  $\hat{x}(k-1 | k-1)$  and its error covariance  $P(k-1 | k-1)$  at time  $k-1$  and linearly predicts them to time  $k$ . The prediction is then used to determine a validation gate in the measurement space based on the measurement prediction  $\hat{z}^t(k | k-1)$  and its covariance  $S(k)$ . When more than one measurement  $z_i(k)$  fall inside the gate, the closest one to the prediction is used to update the filter. The metric used is the chi-squared distance:

$$D_i^t = \left( V_i^t \right)' S^t(k)^{-1} \left( V_i^t \right) \leq \gamma$$

$$= \left( z_i(k) - \hat{z}^t(k) \right)' S^t(k)^{-1} \left( z_i(k) - \hat{z}^t(k) \right) \leq \gamma \quad (11)$$

The update corrects the state prediction by a time-varying gain multiplying the difference between the prediction and the actual measurement. The error covariance is also updated (see [22] for further details). This filter is only mean-square optimal when there are no false alarms and a single target is present.

## 2.3 2-D Assignment Algorithm

When multiple targets are present, the nearest neighbor rule can be modified to take target multiplicity into account. Suppose there are T tracks and M validated measurements between them. The single-scan measurement-to-track association problem may be posed as a 2-D assignment problem [23] in which the assignment cost between measurements  $i$  and track  $t$  is taken as the negative logarithm of:

$$g_{it}^{2t} = \left| 2\pi S^t(k) \right|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left( V_i^t \right)' S^t(k)^{-1} \left( V_i^t \right) \right\} \quad (12)$$

The resulting assignment problem may be solved by the algorithms based on shortest augmenting paths [24]. The algorithm yields associations that enable tracks to be updated with their assigned measurement. Tracks not receiving a measurement are predicted but not updated.

## 3. Optimum Innovation Data Association

The NNKF suffers from tracking in dense clutter environment and its performance is degraded with many loss-tracks accordingly, a new suboptimal algorithm optimum innovation data association (OI-DA) is introduced to increase the tracking performance and to be able to track maneuvering targets in heavy clutter. The main idea based on detecting or distinguishing between the clutter measurements in the gate of the predicted target and the measurement originated from the moving target using two successive scan. The measurements at time  $k-1$  that lie in the gate of the predicted target position (predict to time  $k$ ) is processed by the following method with the measurements at time  $k$  that lie in the same gate to obtain the optimum innovation corresponding to distance metric between true target measurement and the predicted target

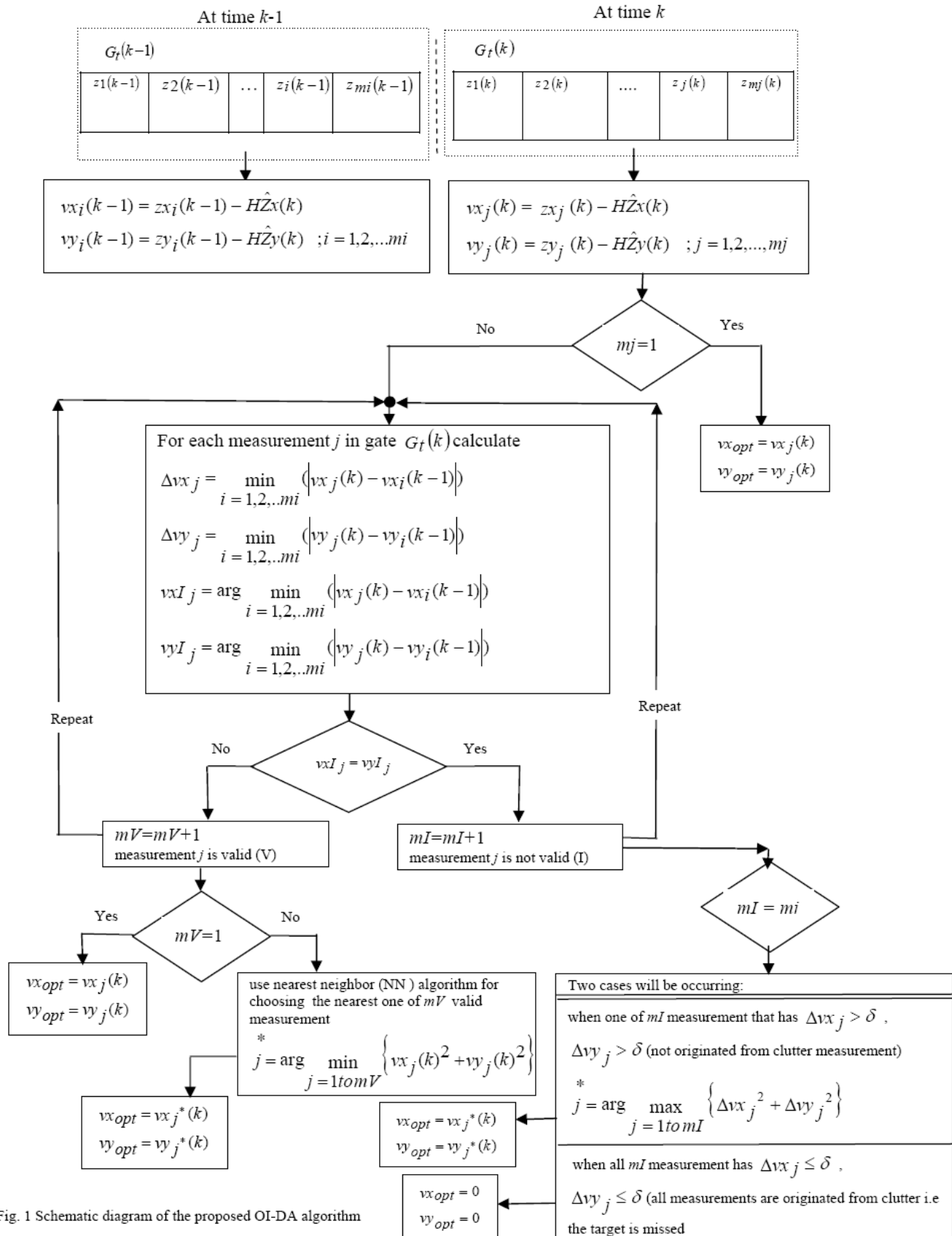


Fig. 1 Schematic diagram of the proposed OI-DA algorithm

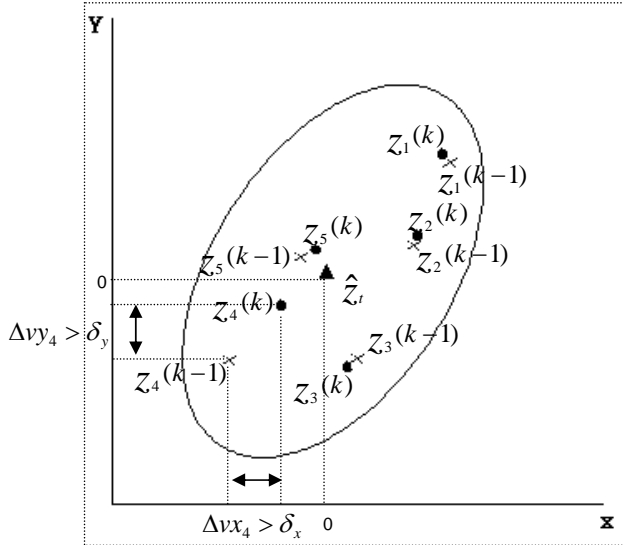


Fig. 2 The current and previous targets position of x,y coordinate in a gate

position. To obtain the optimum innovation we have three models that are processed individually, where the NN algorithm is used as one of them. In this section as shown in Fig 1, we introduce a new algorithm.

In the prediction step, Let  $Z(k-1) = \{z_1(k-1), z_2(k-1), \dots, z_{w_n}(k-1)\}$  be a set of points in the 2-D Euclidean space at time  $k-1$  where  $w_n$  is the number of points at time scan  $\Delta t$  and let  $\hat{z}^t(k)$  be a predicted position of the  $t^{th}$  tracked target at time  $k$ .

according to distance metric measure and gate size, let  $\bar{Z}^t(k-1) = \{z_1(k-1), \dots, z_i(k-1), \dots, z_{m_i}(k-1)\}$  be a set

of the candidate points detected in the  $t^{th}$  gate  $G_t(k-1)$  of predicted position  $\hat{z}^t(k)$  whose elements are a subset from the set  $Z(k-1)$  where  $i = 1$  to  $m_i$  ( number of detected points in gate  $G_t(k-1)$  at time  $k-1$ ) and  $\bar{Z}^t(k-1)$

be a set of all valid points  $z_i(k-1)$  that satisfy the distance measure condition

$$\left( z_i(k-1) - \hat{z}^t(k) \right)' S^t(k)^{-1} \left( z_i(k-1) - \hat{z}^t(k) \right) \leq \gamma$$

for each target  $t$  where  $\gamma$  is threshold value that determines the gate size and  $l = 1$  to  $w_n$ ,  $i = 1$  to  $m_i$ , i. e for each target  $t$ ,  $i$  is initialized by 1 and is increased by  $i = i + 1$  after each valid point is detected up to last  $m_i$  detected points.

In the updating step, let  $Z(k) = \{z_1(k), z_2(k), \dots, z_{w_c}(k)\}$  be a set of points in the 2-D Euclidean space at time  $k$  where

$w_c$  is the number of points at time scan  $\Delta t$ . The candidate points detected in the same gate  $G_t(k)$  as  $G_t(k-1)$  of the  $t^{th}$  predicted position  $\hat{z}^t(k)$  be a subset  $\bar{Z}^t(k) = \{z_1(k), \dots, z_j(k), \dots, z_{m_j}(k)\}$  from the set  $Z(k)$

where  $j = 1$  to  $m_j$  (number of detected points in  $t^{th}$  gate at time  $k$ ) and  $\bar{Z}^t(k)$  be a set of all valid points  $z_j(k)$  that satisfy the distance measure condition

$$\left( z_j(k) - \hat{z}^t(k) \right)' S^t(k)^{-1} \left( z_j(k) - \hat{z}^t(k) \right) \leq \gamma$$

for each target  $t$  where  $l = 1$  to  $w_c$ ,  $j = 1$  to  $m_j$  for  $j = j + 1$  after each valid point is detected. To distinguish between the detected measurements in  $G_t(k)$  that originated from the target or originated from clutter (false target), the nearest

of each measurement of  $x$  and  $y$  component in  $G_t(k)$  to its corresponding measurement in  $G_t(k-1)$  is calculated and to observe the distance measure between each measurement in  $G_t(k)$  and its nearest value. Then we consider that the measurement in  $G_t(k)$  is originated from clutter in case its nearest measure not exceed a threshold value which represent fixed or false moving target (clutter). This is based on calculation of the innovation mean for all detected points  $z_i(k-1)$ ,  $z_j(k)$  of  $x$  and  $y$

component as follow;

$$vx_i(k-1) = zx_i(k-1) - H\hat{Z}x(k) \quad (13)$$

$$vy_i(k-1) = zy_i(k-1) - H\hat{Z}y(k) \quad ; i = 1, 2, \dots, m_i$$

$$vx_j(k) = zx_j(k) - H\hat{Z}x(k) \quad (14)$$

$$vy_j(k) = zy_j(k) - H\hat{Z}y(k) \quad ; j = 1, 2, \dots, m_j$$

Each point  $j$  in  $G_t(k)$  has nearest point  $i$  in  $G_t(k-1)$  by calculating the minimum absolute difference value  $(\Delta vx_j, \Delta vy_j)$  and its index  $(vxI_j, vyI_j)$  between the calculated innovation means for all point  $i$  at each point  $j$  as follow;

$$\Delta vx_j = \min_{i=1, 2, \dots, m_i} \left( |vx_j(k) - vx_i(k-1)| \right)$$

$$\Delta vy_j = \min_{i=1, 2, \dots, m_i} \left( |vy_j(k) - vy_i(k-1)| \right)$$

$$vxI_j = \arg \min_{i=1, 2, \dots, m_i} \left( |vx_j(k) - vx_i(k-1)| \right) \quad (15)$$

$$vyI_j = \arg \min_{i=1, 2, \dots, m_i} \left( |vy_j(k) - vy_i(k-1)| \right) \quad (16)$$

Depending on the clutter point has very small change compared to the change in target point of  $x$  and  $y$  component at two successive scan in each gate its center is the prediction target position. For simplicity, if we assume as shown in Fig.(2), the gate includes measurements  $\{z1,z2,z3,z4,z5\}$  at time  $k$  and time  $k-1$  in  $x,y$  coordinate, it is clear that  $z1(k),z2(k),z3(k),z5(k)$  are measurements originated from clutter while  $z4(k)$  is a measurement originated from the target, we found that the considering of clutter point has high probability when index  $vxI_j$  is the same as or (equal to)  $vyI_j$  while the considering of target point has high probability when index  $vxI_j$  is different or (not equal to)  $vyI_j$ , according to the above consideration we detect how many points  $mI$  represent a clutter point (i.e the corresponding measurements  $j$  are not valid and are avoided from data association process) and how many point  $mV$  represent a target point (i.e corresponding measurements  $j$  are valid and one of them has the optimum index that is found by data association process). The data association process take in consideration the optimum innovation mean  $(vx_{opt},vy_{opt})$  directly in case that the number of detected points  $mV$  is one, which is the normal case when the target exist and the remaining points represent a clutter (invalid points)

$$\begin{aligned} vx_{opt} &= vx_j(k) \\ vy_{opt} &= vy_j(k) \end{aligned} \quad (17)$$

Another case that data association process take in consideration the optimum innovation mean  $(vx_{opt},vy_{opt})$  directly when existing target with no clutter without entering in calculation model of innovation mean process. i.e. the calculated number of detected point  $mj$  is one in  $G_t(k)$ .

$$\begin{aligned} vx_{opt} &= vx_j(k) \\ vy_{opt} &= vy_j(k) \end{aligned}, \text{ where } j=1 \quad (18)$$

Two special cases may be occurring according to the scenario in the following application assignment:-

The first case, gate contain more than one moving target and  $mV>1$  as a result of data association process. The optimum innovation mean  $(vx_{opt},vy_{opt})$  is calculated by NNDA as follow;

$$j = \arg \min_{j=1 \text{ to } mV} \left\{ vx_j(k)^2 + vy_j(k)^2 \right\} \quad (19)$$

$$\begin{aligned} vx_{opt} &= vx_{j^*}(k) \\ vy_{opt} &= vy_{j^*}(k) \end{aligned} \quad (20)$$

\*

Where  $j$  is the index of selected measurement from  $mV$  valid point that has the minimum distance from the predicted target position.

The second case, all measurements in the gate are calculated to be invalid as result of data association process i.e  $mV=0$ ,  $mI = mj$ . in this case we have two consideration:-

- The target may be exist and moves small distance when decreasing its velocity due to maneuvering and takes invalid consideration as the remaining false target in the gate but the change in distance is still higher than the threshold value that detect the target as clutter i.e  $\Delta vx_j > \delta$ ,  $\Delta vy_j > \delta$ . The optimum innovation mean

$(vx_{opt},vy_{opt})$  is calculated by selecting the measurement that has the maximum change in distance under condition  $\Delta vx_j > \delta$ ,  $\Delta vy_j > \delta$  as follow,

$$j = \arg \max_{j=1 \text{ to } mI} \left\{ \Delta vx_j^2 + \Delta vy_j^2 \right\} \quad (21)$$

$$vx_{opt} = vx_{j^*}(k) \quad (22)$$

$$vy_{opt} = vy_{j^*}(k)$$

- The target not detected in the gate (missed) and all measurements are considered to be false target. In this case, the updated target is assigned to the predicted target position and no innovation mean value is required i.e

$$\begin{aligned} vx_{opt} &= 0 \\ vy_{opt} &= 0 \end{aligned} \quad (23)$$

Finally, we obtain the optimum innovation mean that is related to the true selected target with decreasing the probability of error and is used in updating target to the correct position. Reducing the number of valid points in the  $t^{th}$  gate by detecting the false measurement to be invalid (i.e not include in the data association process), this increase the probability for choosing the true measurement originated from the target and improve the data association process.

#### 4. Implementation of Optimum Innovation Data Association (OI-DA) using the kalman filter.

We propose an algorithm which depends on the history of observation for one scan and uses innovation mean calculation with a fixed threshold to obtain the optimum innovation mean that is related to the association pairing between the choosing measurement and track (predicted target) and is used in update state estimation of the target.

In conventional data association approaches with a fixed threshold, all observations lying inside the reconstructed gate are considered in association. The gate may have a large number of observations due to heavy clutter, this leading to; increasing in association process since the probability of error to associate target-originated measurements may be increased. In our proposed algorithm detecting moving target indicator (MTI) filter is used to provide the possibility to decrease the number of observations in the gate by dividing the state of observations into valid represent moving targets and invalid represent the fixed or false targets that only the valid are considered in association. The proposed OI-DA using Kalman filter is represented in algorithm 1.

Algorithm 1 OI-DA using Kalman filter

1. for  $t = 1$  to  $T$  do
2. Do prediction step,  
 $x^t(k | k-1) \sim p(x^t(k) | Z_{1:k-1}) = N(x^t(k), \bar{m}^t(k), \bar{p}^t(k))$   
 where  
 $\bar{m}^t(k) = A^t(k)m^t(k-1)$   
 $\bar{p}^t(k) = A^t(k)p^t(k-1)A^{t'}(k) + Q$
3. Calculate optimum innovation mean  $V_{opt}(k)$  by OI-DA algorithm described in algorithm 2
4. Do update step  
 $m^t(k) = \bar{m}^t(k) + K^t(k)V_{opt}(k)$   
 $p^t(k) = \bar{p}^t(k) - K^t(k)S^t(k)K^{t'}(k)$   
 $S^t(k) = H^t(k)\bar{p}^t(k)H^{t'}(k) + R(k)$   
 $K^t(k) = \bar{p}^t(k) - H^t(k)S^t(k)^{-1}$
5. end for

Algorithm 2 Calculate  $V_{opt}(k)$  by OI-DA

1. Find validated region for measurements at time  $k-1$ :  
 $\bar{Z}^t(k-1) = \{z_i(k-1)\}, i = 1, \dots, mi$   
 By accepting only those measurements that lie inside the gate  $t$ :  
 $\bar{Z}^t(k-1) = \left\{ Z : \left( z_i(k-1) - H^t(k)\bar{m}^t(k) \right) S^t(k)^{-1} \right.$   
 $\left. \left( z_i(k-1) - H^t(k)\bar{m}^t(k) \right) \leq \gamma \right\}$
2. Find validated region for measurements at time  $k$ :  
 $\bar{Z}^t(k) = \{z_j(k)\}, j = 1, \dots, mj$

By accepting only those measurements that lie inside the gate  $t$

$$\bar{Z}^t(k) = \left\{ Z : \left( z_j(k) - H^t(k)\bar{m}^t(k) \right) S^t(k)^{-1} \right.$$

$$\left. \left( z_j(k) - H^t(k)\bar{m}^t(k) \right) \leq \gamma \right\}$$

where  $s^t(k) = H^t(k)\bar{p}^t(k)H^{t'}(k) + R$

3. Calculate innovation mean for all measurement lie inside the gate  $t$  at time  $k-1$  and  $k$  respectively

$$vx_i(k-1) = zx_i(k-1) - H\hat{Z}x(k)$$

$$vy_i(k-1) = zy_i(k-1) - H\hat{Z}y(k) \quad ; i = 1, 2, \dots, mi$$

$$vx_j(k) = zx_j(k) - H\hat{Z}x(k)$$

$$vy_j(k) = zy_j(k) - H\hat{Z}y(k) \quad ; j = 1, 2, \dots, mj$$

4. if  $m_j > 1$  calculate the index and change of the nearest measurement  $i$  in the gate  $t$  at time  $k-1$  to each measurement  $j$  in the gate  $t$  at time  $k$  for  $x$  and  $y$  component.

$$\Delta vx_j = \min_{i=1,2,\dots,mi} \left| vx_j(k) - vx_i(k-1) \right|$$

$$\Delta vy_j = \min_{i=1,2,\dots,mi} \left| vy_j(k) - vy_i(k-1) \right|$$

$$vxI_j = \arg \min_{i=1,2,\dots,mi} \left| vx_j(k) - vx_i(k-1) \right|$$

$$vyI_j = \arg \min_{i=1,2,\dots,mi} \left| vy_j(k) - vy_i(k-1) \right|$$

5. Calculate invalid  $mI$  measurement (false target) in case  $vxI_j = vyI_j$  and  $mV$  measurement (true moving target) in case  $vxI_j \neq vyI_j$

- Calculate directly the optimum innovation

$$v_{opt} = (vx_{opt}, vy_{opt})' \text{ in case } (mV = 1, j = \text{index}(mV))$$

or  $(mj = 1, j = 1)$

$$vx_{opt} = vx_j(k)$$

$$vy_{opt} = vy_j(k)$$

- Choose NN of  $mV$  valid measurement to be the optimum innovation  $v_{opt} = (vx_{opt}, vy_{opt})'$  in case

$(mV > 1, j = \text{index}(mV))$

$$* j = \arg \min_{j=1 \text{ to } mV} \left\{ vx_j(k)^2 + vy_j(k)^2 \right\}$$

$$vx_{opt} = vx_{j^*}(k)$$

$$vy_{opt} = vy_{j^*}(k)$$

- Choose the measurement to be the optimum innovation  $v_{opt} = (vx_{opt}, vy_{opt})'$  that has the maximum change in distance under condition



$\Delta vx_j > \delta$  ,  $\Delta vy_j > \delta$  in case  $mV = 0$ ,  $mI = mj$ ,

$j = \text{index}(mI)$  as follow,

$$j = \arg \max_{j=1 \text{ to } mI} \left\{ \Delta vx_j^2 + \Delta vy_j^2 \right\}$$

$$vx_{opt} = vx_{j^*}(k)$$

$$vy_{opt} = vy_{j^*}(k)$$

- Otherwise the above condition, the optimum will be set as

$$vx_{opt} = 0$$

$$vy_{opt} = 0$$

6- end

## 5. Simulation Results

Simulation results have been carried out to monitor the performance of the proposed OI-DA algorithm compared to the conventional NNKF and JPDA filter. To highlight the performance of the proposed algorithm, we used a synthetic dataset to track three maneuvering targets which are continues from the first frame to the last frame in varying clutter density. The initial mean  $m_0^t = (x, y, \dot{x}, \dot{y})'$  for the initial distribution  $p(x^t(0))$  is set to  $m_0^1 = [17.7, 9.16, 0, 0]$ ,  $m_0^2 = [13.3, 8.8, 0, 0]$ ,  $m_0^3 = [14.4, 11.7, 0, 0]$ , and covariance  $p_0^t = \text{diag}([400, 400, 100, 100])$ ,  $t = 1, 2, 3$ . The row and column sizes of the volume ( $V = \mathcal{S}_W \times \mathcal{S}_H$ ). We initiate the other parameters as:  $N = 20 \times 20$ , the sampling time  $\Delta t = 4$  sec,

$T = 4 \times 32 = 128$  sec,  $P_D = 0.99$ , in addition, we also set the matrices of (1),(2) as

$$A = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, Q = G G^T,$$

$$R = \begin{bmatrix} 40 & 0 \\ 0 & 40 \end{bmatrix}, G = \begin{bmatrix} \frac{\Delta t^2}{2} & 0 \\ 0 & \frac{\Delta t^2}{2} \\ \Delta t & 0 \\ 0 & \Delta t \end{bmatrix}$$

Given a fixed threshold ( $\gamma = 10^{-4}$ ), we showed that at high signal to noise ratio with low clutter density ( $\lambda = 0.0005 m^{-2}$ ), the three algorithms appear to perform as expected. Fig. 3(a),(b),(c) shows the estimated target tracks using the NNKF, JPDAF and the OI-DA filters respectively at low clutter density. The figures show that the three filters were effectively able to track the targets at high SNR. At low signal to noise ratios the corrupted target track in a uniform clutter with high varying clutter density ( $\lambda = 0.001 m^{-2}$  for medium clutter and  $\lambda = 0.01 m^{-2}$  for dense clutter) is shown in Fig.4, where the NNKF and JPDA filters were not be able to track the targets. Fig. 5,6 show the estimated target tracks using the NNKF, the JPDAF, and the proposed OI-DA filters at the two different SNR as mentioned above where In this figures, the colored solid line represents the underlying truth targets of the trajectory (each target with different color) while the colored + symbol represents trajectory of the tracked targets. The figures shows that only the OI-DA as shown in Fig. 5,6 (c) is able to track the targets at two different heavy clutter density. The explanation of this behavior is due to the fact that, at low SNR the target-originated measurement may fall outside the validation gate when choosing the wrong valid measurements during data association process and as a result, the estimated target states will be clutter- originated. The OI-DA has the advantage to increase the probability of choosing the correct candidate measurement. We also compared error root mean square value (RMSE) for the different three approaches each with three targets at our three cases in different clutter as shown in Fig. 7. Our proposed algorithm has lower error, RMSE values than JPDAF over frame numbers and approximately the same as NNKF.

## 6. Conclusions

From the results obtained in the simulations for multi-target tracking, it can be seen that at low clutter density (high SNR), all the tracking algorithm (NNKF, JPDAF and OI-DA) are able to track the targets. However, at

heavy varying clutter density (low SNR), NNKF and JPDA algorithm fail to track the targets, where the proposed OI-DA algorithm has the capability to maintain the tracked targets. From the valid based measurement regions, The OI-DA algorithm distinguishes between the fixed or false targets to be considered as invalid targets and the moving true targets to be valid during data association process. The OI-DA algorithm overcome the NNKF problem of loss tracking the targets in dense clutter environment and has the advantage of low computational cost over JPDAF. By using this new approach, we can obtain smaller validated measurement regions with improving the performance of data association Process which have been shown to give targets the ability to continue tracking in dense clutter.

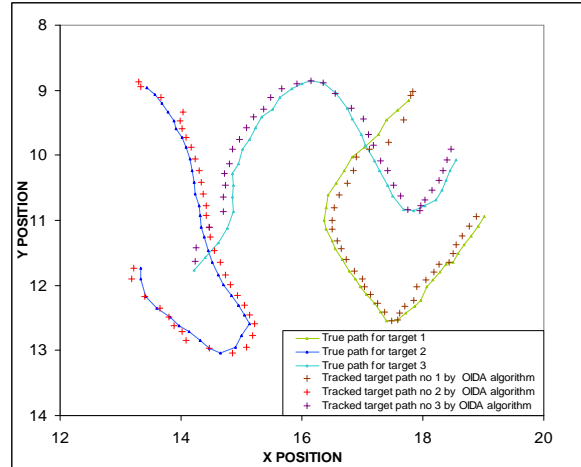


Fig. 3. X- and Y- trajectory show the state of successful tracking to maneuvering multi-targets (3 target with + symbol for tracked target position and solid line for true target path) move in low clutter using 3 approaches algorithm (a) NNKF (b) JPDAF (c) OI-DA.

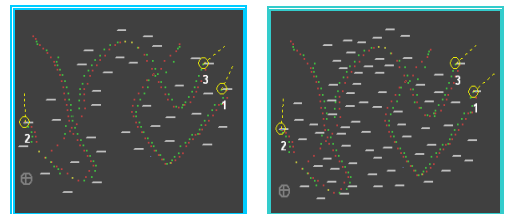
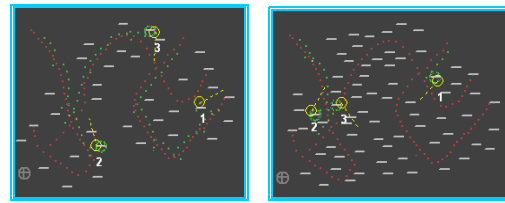
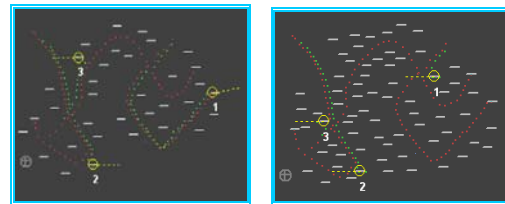
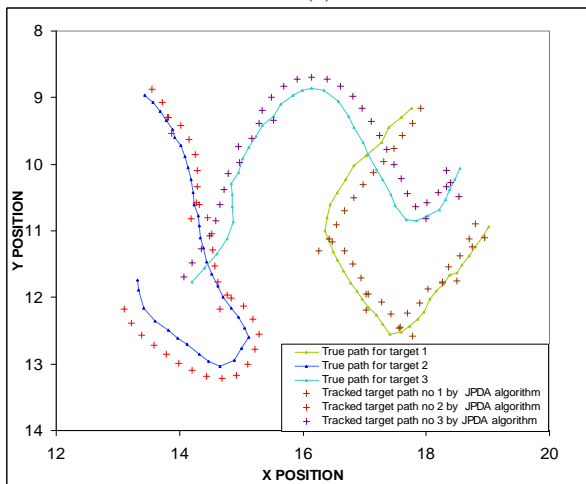
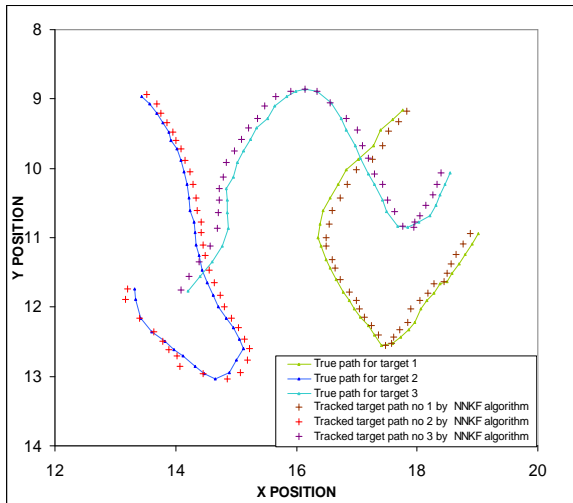
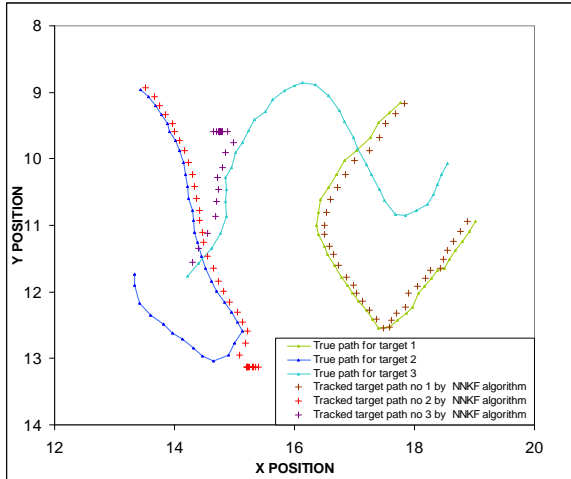
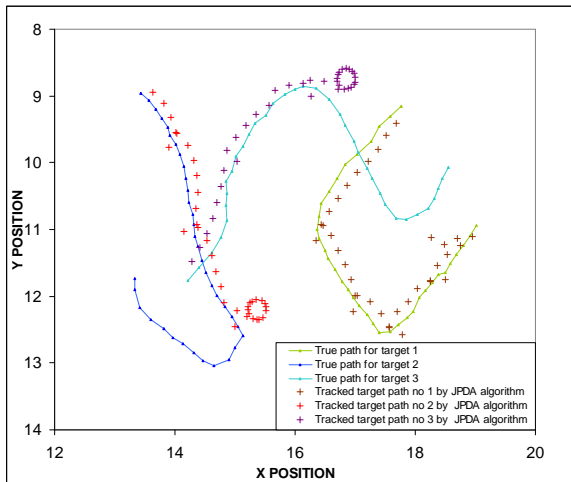


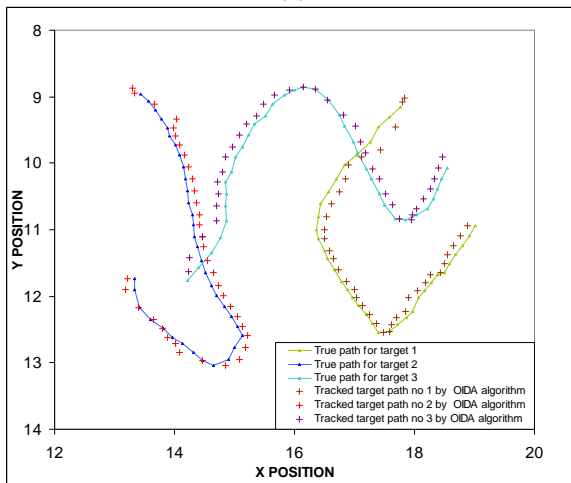
Fig. 4. The state of tracking 3 targets move in different clutter density using 3 approaches algorithm NNKF as in (a),(b), JPDAF as in (c),(d) and OI-DA as in (e),(f). Images (a),(c),(e) show tracking in medium clutter and images (b),(d),(f) show tracking in dense clutter



(a)

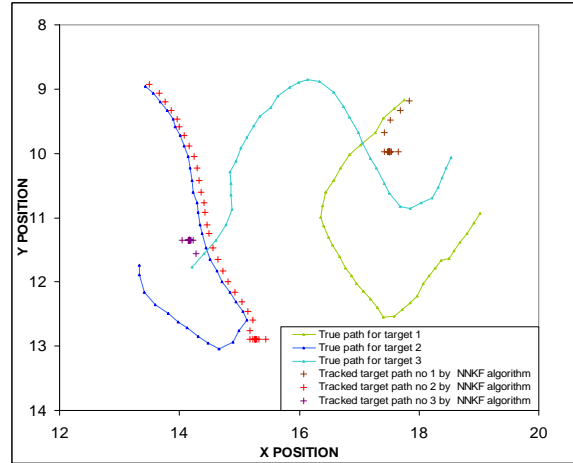


(b)

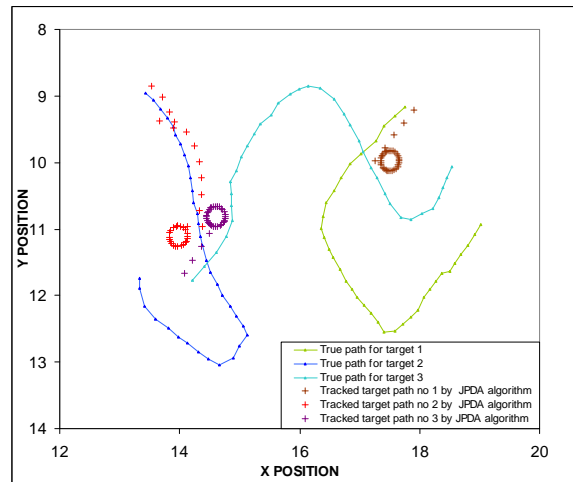


(c)

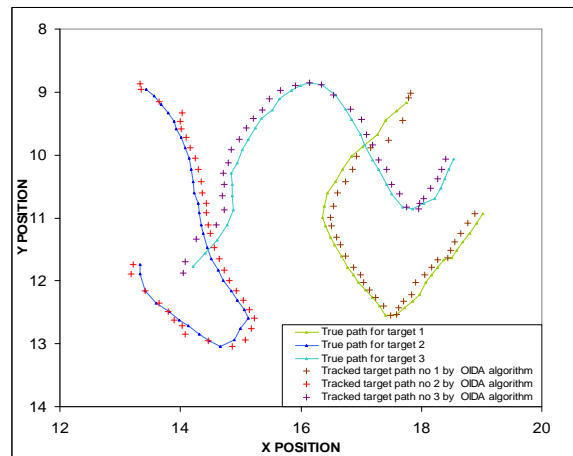
Fig. 5 X- and Y- trajectory show the state of tracking 3 targets in medium clutter (+ symbol refer to tracked target position and solid line to true target path) using 3 approaches algorithm (a) NNKF and (b) JPDAF loss track while (c) OI-DA maintains tracks



(a)

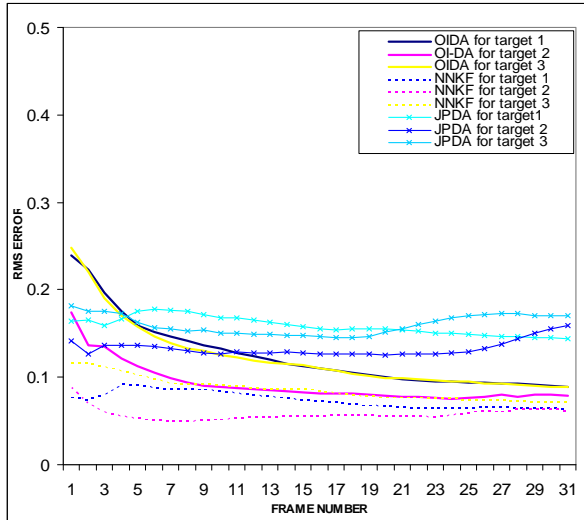


(b)

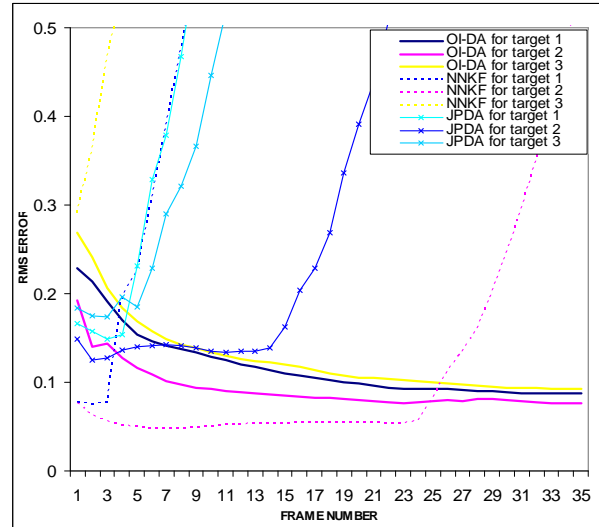


(c)

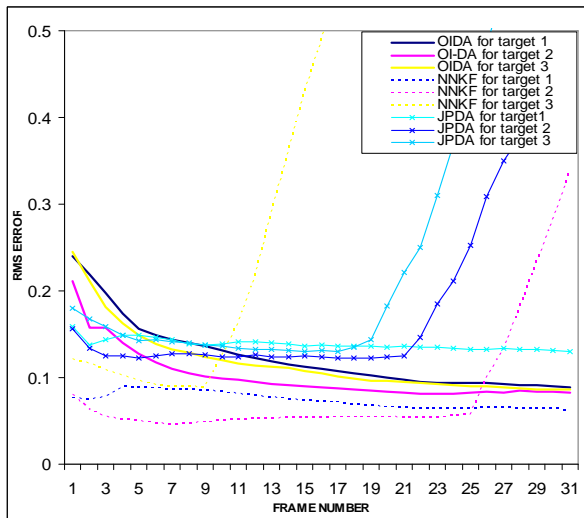
Fig. 6 X- and Y- trajectory show the state of tracking 3 targets in dense clutter (+ symbol and solid line refer to tracked target position and true target path respectively) using 3 approaches algorithm (a) NNKF and (b) JPDAF loss track while (c) OI-DA maintains tracks



(a)



(c)



(b)

Fig. 7 The root mean square error[RMSE] for each target (3 targets) separately over frame number (each frame take 4 sec / one scan) for the 3 approaches algorithm as (a) with low clutter ,(b) with medium clutter and (c) with dense clutter . From (b), (c) the RMSE is maintained minimum for the proposed OI-DA and less sensitivity to dense clutter.

## References

- [1] X. Wang, S. Challa, and R. Evans, "Gating techniques for maneuvering target tracking in clutter", IEEE Transactions on Aerospace and Electronic Systems, Vol.38, No.3, July 2002, PP. 1087-1097.
- [2] R. A. Singer, R. G. Sea, K. B. Housewright, "Derivation and evaluation of improved tracking filters for use in dense multi-target environments", IEEE Trans on Information Theory, Vol. 20, No. 4, 1974, PP.423-432.
- [3] Bar-Shalom Y., Li X. R., "Multitarget-Multisensor Tracking: Principles and Techniques.", Storrs, CT: YBS Publishing, 1995.
- [4] VPS Naidu, G Girija, J R Raol "Data association and fusing algorithms for tracking in presence of measurement loss" IE(I) journal-AS Vol.86, May 2005 , PP17-28.
- [5] Hui Chen, Chen Li "Data association approach for two dimensional tracking base on bearing-only measurements in clutter environment", Journal of Software Vol. 5, No. 3, 2010, PP. 336-343.
- [6] Y. Bar-Shalom, E. Tse, "Tracking in a cluttered environment with probabilistic data association", Automatica, Vol. 11, 1975, PP.451-460.
- [7] Yaakov Bar-Shalom, Fred Daum, and Jim Huang. "The Probabilistic Data Association Filter", IEEE Control Systems Magazine Vol. 29, No. 6, PP. 82-100, December 2009.
- [8] K. C. Chang, Y. Bar-Shalom, "Joint probabilistic data association for multi-target tracking with possibly unresolved measurements and maneuvers", IEEE Trans on Automatic Control, Vol. 29, 1984, PP.585-594
- [9] D. B. Reid, "An algorithm for tracking multiple targets", IEEE Trans on Automatic Control, Vol. 24, No. 6, 1979, PP.843-854,

- [10] G. W. Pulford "Taxonomy of multiple target tracking methods" IEE Proc.-Radar Sonar Navig. , Vol. 152, No.5, 2005, PP.291-304
- [11] E.M.Saad, El.Bardawiny , H.I.ALI and N.M.Shawky "New Data Association Technique for Target Tracking in Dense Clutter Environment using Filtered Gate Structure" Signal processing: An International Journal (SPIJ) Vol. 4, Issue 6 ,2011, PP. 337-350
- [12] E.M.Saad, El.Bardawiny , H.I.ALI and N.M.Shawky" Filtered Gate Structure Applied to Joint Probabilistic Data Association Algorithm for Multi-Target Tracking in Dense Clutter Environment", IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 2, 2011, PP.161-170
- [13] Blackman, Samuel, Robert Popoli, "Design and Analysis of Modern Tracking Systems", Boston: Artech House, 1999.
- [14] Y. Bar-Shalom, X. R. Li, T. Kirubarajan, "Estimation with applications to tracking and navigation", New York: John Wiley & Sons, 2001.
- [15] H. Zhou, Z. Jin, D. Wang, "Maneuver targets tracking", Beijing: Defense Industry Press, 1991.
- [16] Y. He, J. J. Xiu, J. Zhang, X. Guan, "Radar data processing and application", Beijing: Electronic Industry Press, 2006.
- [17] M. I. Skolnik. "Radar handbook", McGraw-Hill, Inc
- [18] Simon Haykin, "Classification of radar clutter in an air traffic control environment" proceedings of the IEEE, Vol. 79, No 6 ,1991, PP. 742-772
- [19] Gokhan Soysal and Murat Efe "Performance comparison of tracking algorithms for a ground based radar" Commun.Fac.Sci. Univ.Ank. series A2-A3, Vol. 51(1) (2007) PP 1-16
- [20] Simon Haykin. "Radar signal processing", IEEE ASSP Magazine, April 1985.
- [21] Grewal, M.S. and Andrews, A.p, "Kalman filtering, theory and practice using MATLAB", Wiley interscience, 2001.
- [22] Y. Bar-Shalom and T. E. Fortmann, *Tracking and Data Association*, Academic Press, 1988.
- [23] S. Blackman, *Multiple Target Tracking with Radar Applications*, Artech House, MA, 1986.
- [24] R. Jonker, A. Volgenant, "A Shortest Augmenting Path Algorithm for Dense and Sparse Linear Assignment Problems", Computing, v. 38, 1987, 325-340.



# An Efficient Quality of Service Based Routing Protocol for Mobile Ad Hoc Networks

T K Godder<sup>1</sup>, M. M Hossain<sup>2</sup>, M M Rahman<sup>1</sup> and Md. Sipon Miah<sup>1</sup>

<sup>1</sup> Dept. of Information & Communication Engineering, Islamic University, Kushtia, 7003, Bangladesh

<sup>2</sup> Dept. of Applied Physics & Electronic Engineering, Rajshahi University, Rajshahi University, 6205, Bangladesh

## Abstract

Ad-hoc network is set up with multiple wireless devices without any infrastructure. Its employment is favored in many environments. Quality of Service (QoS) is one of the main issues for any network and due to bandwidth constraint and dynamic topology of mobile ad hoc networks, supporting Quality of Service (QoS) is extremely a challenging task. It is modeled as a multi-layer problem and is considered in both Medium Access Control (MAC) and routing layers for ad hoc networks. Ad-hoc On-demand Distance Vector (AODV) routing protocol is one of the most used and popular reactive routing protocols in ad-hoc networks. This paper proposed a new protocol 'QoS based AODV' (QAODV) which is a modified version of AODV.

**Keywords:** QoS, Ad Hoc Network, Routing Protocol, AODV.

## 1. Introduction

A Mobile Ad Hoc Networks (MANET) is a collection of mobile nodes that can communicate with each other using multi-hop wireless links without utilizing any fixed based-station infrastructure and centralized management. Each mobile node in the network acts as both a host generating flows or being destination of flows and a router forwarding flows directed to other nodes. With the popularity of ad hoc networks, many routing protocols have been designed for route discovery and route maintenance. They are mostly designed for best effort transmission without any guarantee of quality of transmissions. Some of the most famous routing protocols are Dynamic Source Routing (DSR), Ad hoc On Demand Distance Vector (AODV), Optimized Link State Routing protocol (OLSR), and Zone Routing Protocol (ZRP). In MAC layer, one of the most popular solutions is IEEE 802.11. At the same time, Quality of Service (QoS) models in ad hoc networks become more and more required because more and more real time and multimedia applications are implemented on the network. In MAC layer, IEEE 802.11e is a very popular issue discussed to set the priority to users. In routing layer, QoS are guaranteed in terms of data rate, delay, and jitter and so

on. By considering QoS in terms of data rate and delay will help to ensure the quality of the transmission of real time media. For real time media transmission, if not enough data rate is obtained on the network, only part of the traffic will be transmitted on time. There would be no meaning to receiving the left part at a later time because real time media is sensitive to delay. Data that arrive late can be useless. As a result, it is essential for real time transmission to have a QoS aware routing protocol to ensure QoS of transmissions. In addition, network optimization can also be improved by setting requirements to transmissions. That is to say, prohibit the transmission of data which will be useless when it arrive the destination to the network. From the routing protocol point of view, it should be interpreted as that route which cannot satisfy the QoS requirement should not be considered as the suitable route in order to save the data rate on the network. In this paper, we describe a QoS-aware modification of the AODV reactive routing protocol called QoS Aware AODV (Q-AODV). This serves as our base QoS routing protocol.

## 2. Proposed Topology

In this section I would like to show the difference between the QAODV and the AODV routing protocols during transmission with the following simple topology. There are four nodes in this network, and the initial topology is a grid as shown in Figure: 1. The scenario is designed as in Table 1. According to the scenario, at the beginning of the transmission of nodes, the two pairs are not interference with each other. At 10s, Node 2 moves towards the direction of Node 0 with a speed of 10 m/s. The distance between Node 0 and Node 2 becomes smaller and smaller, and at time 15 s, these two nodes begin to be in each others carrier sensing range, which means that these two nodes begin to share the same channel. The maximum bandwidth of the channel is around 3.64 Mbps. In AODV, where there is no QoS requirement, when Node 2 is in the

interference range of Node 0, traffics are kept on and some packets are lost during the transmission, whereas, in QAODV, the QoS is ensured. When the promised data rate cannot be satisfied any more, traffic of Node 2 is stopped at once. From this case, we could see that the QAODV achieved the function of ensuring the QoS not only at the route discovery stage, but also during the transmission. Once the QoS is not satisfied, the traffic is stopped [1].

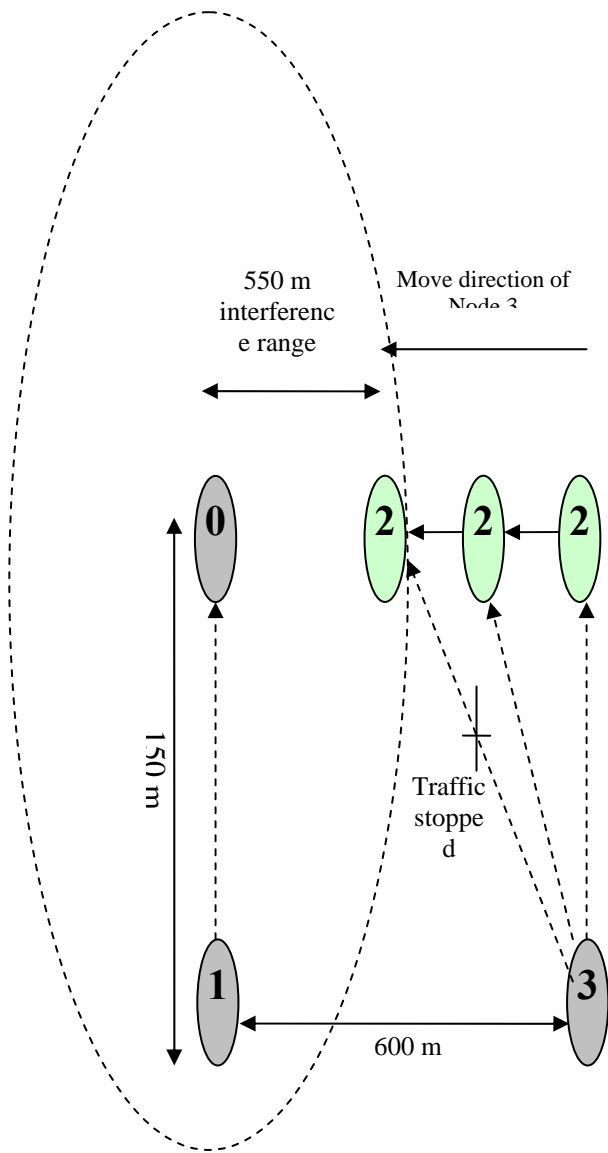


Fig. 1 A simple topology of four nodes

Table 1: Scenario descriptions for proposed topology

Node position	Node 0 (50, 250)	Node 1 (50, 100)	Node 2 (650, 250)	Node 3 (650, 100)
Traffic	Traffic direction	Durati on	Require d data rate	Traffic Type
	Node1 - >Node 0	6s - 18 s	1.8 Mbps	CBR
	Node2 - >Node 3	6 s - 18 s	2 Mbps	CBR
Node Movem ent	Node ID	Time that the node begins to move	Movem ent Speed	Movem ent Directio n (move toward a point)
	Node 2	10 s	10 m/s	(550, 250)

In the topology there were 20 nodes and the simulation environment was as described in Table1. The area size is 670 m \* 670 m, and 20 nodes are in this area. 50 s is added at the beginning of each simulation to stabilize the mobility model. Every simulation runs 500 s in total. Each data point in the results represents an average of ten runs with same traffic models but different randomly generated mobility scenarios. For fair comparisons, same mobility and traffic scenarios are used in both the AODV and the QAODV routing protocols . The screenshot of NAM (Network Animator) at 0 second is given in figure 2.

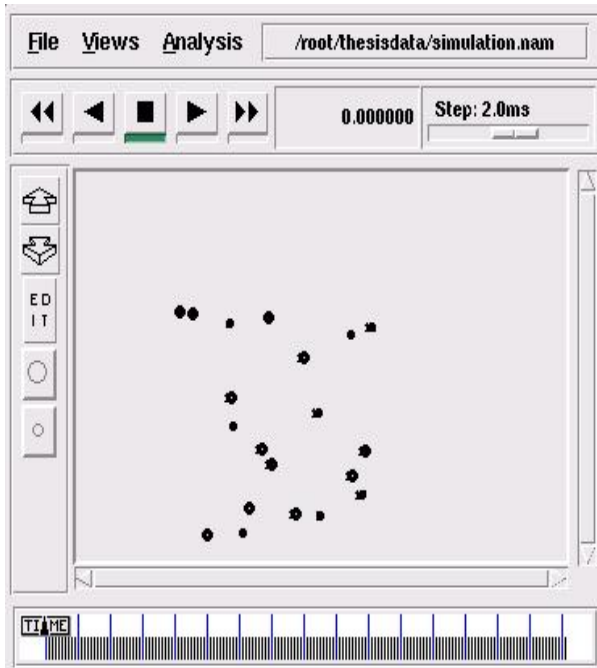


Fig. 2 NAM screenshot of the topology at 0 second

### 3. Simulation Traffic Pattern

The Random Waypoint model provided by NS2 is used as the mobility model. The traffic type in the application layer is CBR with packet size of 512 bytes and in transport layer User Datagram Protocol (UDP) is used. The traffic pattern that used in the simulation is shown in Table 2. It is the same as what the Reference [2] uses.

Table 2. Simulation traffic pattern

Traffic flow	Source and destination node	Start time (s)	End time (s)
Session 1	3 -> 4	53	174
Session 2	7 -> 8	144	280
Session 3	4 -> 5	290	315
Session 4	5 -> 7	305	475
Session 5	5 -> 6	445	483

Setting the traffic flow in such a manner aims at greater interference impact when sessions overlap. The source node and the destination node of each traffic flow are chosen by using function *cbrgen.tcl* randomly.

## 4. Simulation Results and Analysis

For comparing various routing protocols using UDP transport layer protocol, we have chosen three performance metrics Average End to End delay, Packet Delivery Ratio, Normalized Routing Load which are commonly used in the literature to evaluate the performance of the AODV and the QAODV routing protocols.

### 4.1 Data Rate

In this set of simulations, a group of data rates ranging from 50 kbps to 1800 kbps is applied. The mobility scenario is with a pause time of 30 seconds and the maximum node speed is 10 m/s. Three parameters defined above are calculated. The results are shown in the following figures (figure. 3, figure.4, figure.5).

#### 4.1.1 Average end to end delay

From figure.3, it can be seen that AODV routing protocol performs better than QAODV routing protocol when data rate is low (below 600 kbps). The QAODV routing protocol got higher average end to end delay at the low data rate than the AODV because intermediate nodes are not allowed to perform local route repairs in case of link failures with the QAODV routing protocol, thus, there is higher route recovery latency which results in higher end-to-end delay compared with the AODV routing protocol at low data rate.

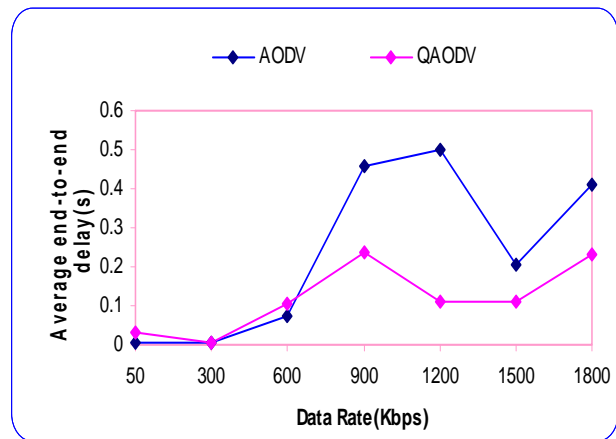


Fig. 3 Average End to End delays with different data rates

Another reason could be that, with the QAODV routing protocol, the number of transmitted routing packets is larger than the number of routing packets transmitted in the AODV routing protocol. In the QAODV routing protocol, all nodes use Hello messages to exchange

information with their neighbors. Routing packets including Hello messages which have higher priority always transmitted firstly and data packets are queued nodes. With the AODV routing protocol, when the traffic is low in the network, no matter which route the traffic flow chose, the route chosen can provide enough data rate at most of the time. As a result, the end to end delay with the AODV routing protocol is not high and can be lower than the QAODV routing protocol at low data rate. If we can take more time for simulation for each data rate comparatively accurate results can be found. For these above reasons, end to end delay in QAODV is higher than the AODV at low data rate. The average end to end delay of the QAODV is always below 240ms, whereas, the end to end delay of the AODV increases badly when the data rate of each traffic flow increases from 600 kbps to 1200 kbps. It shows that networks with the QAODV routing protocol can provide lower end to end delay for traffic flows than the AODV since the QAODV always choose to find a route with satisfying data rate. During the transmission, the QoS of the traffic is monitored in the QAODV routing protocol. Once the QoS is not satisfied as it promised, the traffic stopped. All in all, with the QAODV routing protocol, the average end to end delay is low even the load on the network increases to very high which is not true for the AODV routing protocol. This performance is very significant for real time traffic transmissions.

#### 4.1.2 Normalized Routing Load

In figure.4, the routing overload in AODV and QAODV decreases with the increase of the data rate. In QAODV with the increase of data rate, total number of packets sent increases. For this reason routing overload is relatively high in QAODV at the low data rate. In AODV, routing overload is always low because routing packets are only

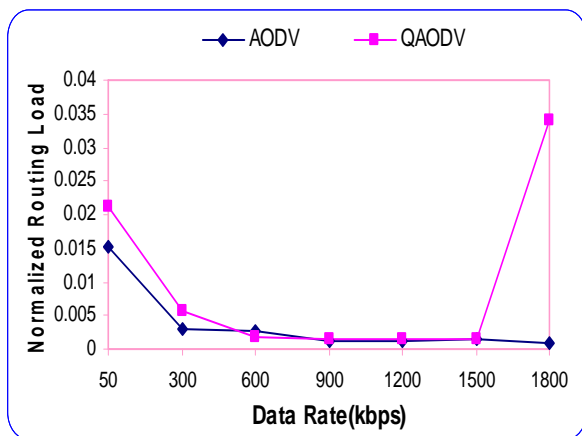


Fig. 4 Normalized routing load at different data rate

sent during the routing searching and maintenance periods without exchanging Hello messages. The Hello messages are needed in the QAODV routing protocol in order to exchange the precisely consumed data rate information of nodes who are sharing the same channel. It is hard to explain why the routing overload badly increase when data rate increases from 1500 kbps to 1800 kbps .

#### 4.1.3 Packet Delivery Ratio

From figure.5 we see that, either we use the QAODV routing protocol or the AODV routing protocol, the packet delivery ratio decreases with the increase of the data rate of traffic flows.

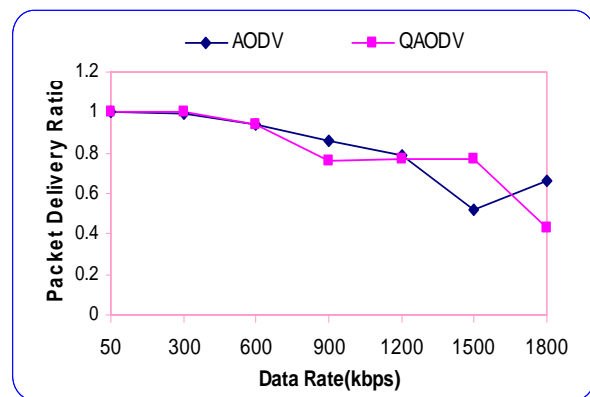


Fig. 5 Packet delivery ratio with different data rates

That is because the increasing data rate of flows increases traffic in the network. When the maximum throughput of nodes cannot satisfy the on-going traffic, queues at nodes begin to be full; the packets in the end of queues of nodes are dropped both at source nodes and at intermediate nodes.

The packet delivery ratio with the QAODV always lower than the AODV because the source node takes more time to find a suitable route in QAODV and during this period of time, the source which keeps on sending packets from the application layer of the node, it cause drops of packets at the end of the queue when the queue is full. Also, the traffic session can be paused anytime when the local available data rate of nodes in the path is not satisfied during the transmission in the QAODV routing protocol. There are strict requirements in terms of data rate for traffic flow with access admission control. When data rate increases from 1500 kbps to 1800 kbps, only paths with hop count 1 or 2 can be admitted. As a result, there is more decrease in PDR with the QAODV than in AODV when the data rate increases from 1500 kbps to 1800 kbps. It is hard to explain why the PDR increase in AODV when data

rate increases from 1500 kbps to 1800 kbps .

For the above reason, the packet delivery ratio with the QAODV routing protocol is lower than the one with the AODV routing protocol is that QAODV routing protocol has more restrictions to the route for transmission. Actually, the packets which are not delivered and dropped at the source node because of the delay for searching for a more suitable route in the QAODV routing protocol should be dropped. The reason is that if these packets are sent, and the route chosen is not satisfying the requirements, packets have more probability to be dropped at the intermediate node or packets may arrive at the destination node late because of the long duration of wait at the intermediate node. In other words, the QAODV routing protocol also helps to prohibit the packets, which have more probability to be dropped during the transmission or that arrived at the destination node late, to be transmitted on the network. It helps to save the data rate as well.

## 4.2 Maximum Node Moving Speed

In the following simulations, the data rate is fixed at 1200 kbps. The maximum node moving speed is increased to see the behaviors of the AODV and the QAODV in a fairly high mobility mode. Maximum node moving speed is changing in the range 1 m/s to 20 m/s. The results are shown in terms of average end to end delay, packet delivery ratio and normalized routing load shown in figure:6, figure:7, and figure:8.

### 4.2.1 Average end to end delay

As shown in figure:6, with the increase of the maximum moving speed, the average end to end delay does not increase much in QAODV as compared with the AODV routing protocol, it means that, this protocol is quite suitable for scenarios with different moving speeds.

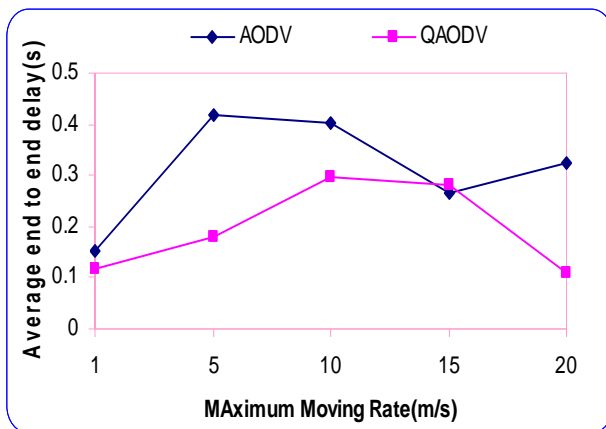


Fig. 6 Average end to end delay with different Max. moving speeds

In comparison, with the AODV routing protocol, the end to end delay varies a lot with the increase of the maximum moving speed . It can be obviously seen that, the end to end delay in QAODV is always much lower than the one in the AODV routing protocol. The low end to end delay of packets ensures the on time transmissions required by real time traffic transmissions.

To sum up, the QAODV routing protocol does decrease end to end delay significantly when the data rate of traffic flows is high.

### 4.2.3 Normalized Routing Load

The routing overload of AODV and QAODV almost zero at minimum speed. This is because once a route discovery process is completed; there is no need to perform the discovery process again. As shown in fig:7 the routing overload increases in AODV and QAODV with the increase of maximum moving speed. In higher mobility networks, a node which is on the route for transmitting traffic flow has higher possibility to move out of the transmission range of the upstream or the downstream nodes. The upstream nodes are nodes that transmit the packets to the considered moving node and the downstream nodes are those that receive packets from the considered moving node.

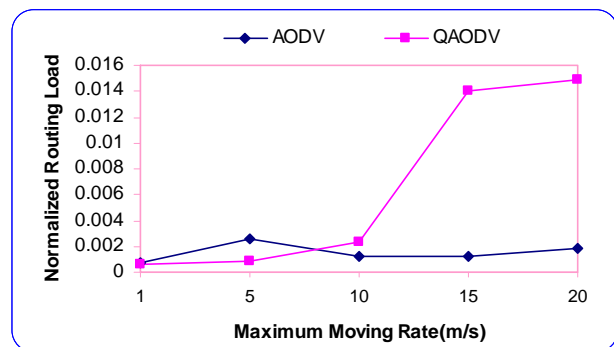


Fig. 7 Normalized routing load with different Max. moving speeds

In order to alert source nodes that there is a lost of one of the intermediate nodes on the route and to find a new route, more and more route discovery and route maintenance packets are sent with the increase of the maximum moving speed of nodes. Normalized routing load which is the number of routing packets divided by the number of successfully delivered packets, in general, increases with the maximum moving speed of nodes. The routing load in the QAODV routing protocol is always much higher than the one in the AODV routing protocol. Thus, we could see that, the QAODV routing protocol

improves the performance at the expense of sending more routing packets on the network. These packets are used to exchange the network information to help assure QoS.

#### 4.2.4 Packet Delivery Ratio

In figure. 8 with low max moving speed the packet delivery ratio in QAODV is higher than the AODV but with the increase of mobility speed the performance is lower than AODV. When the maximum moving speed is up to 20 m/s, almost half of the packets are dropped in QAODV. The reason that why more packets are dropped in QAODV and how they are dropped has been explained in the previous part of this section.

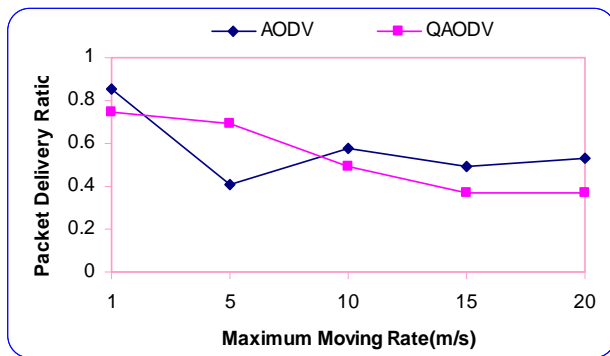


Fig. 8 Packet delivery ratio with different Max. moving speeds.

## 5. Conclusion

In this research, we described the importance of QoS routing in Mobile Ad-Hoc networks, the challenges we met, and the approach we took. We discussed in detail our idea of adding support for QoS into the AODV protocol. After observing the simulation and analyzing the data, it is found that packets could get less end to end delay with a QoS based routing protocol when the traffic on the network is high. This low end to end delay is meaningful for real time transmissions. When the traffic is relatively high on the network, not all the routes that are found by the AODV routing protocol have enough free data rate for sending packets ensuring the low end to end delay of each packet. As a result, the QAODV protocol works well and shows its effects when the traffic on the network is relatively high. People who work on the area of ad hoc networks with the aim of improving the QoS for ad hoc networks can get benefit from this QAODV protocol.

## References

[1] Shane Bracher, "A MECHANISM FOR ENHANCING

INTERNET CONNECTIVITY FOR MOBILE AD HOC NETWORKS", Proceedings of the Second Australian Undergraduate Students' Computing Conference, 2004.

[2] Ronan de Renesse, Mona Ghassemian, Vasilis Friderikos, A. Hamid Aghvami, "Adaptive Admission Control for Ad Hoc and Sensor Networks Providing Quality of Service" Technical Report, Center for Telecommunications Research, King's College London, UK, May 2005.

[3] H. Badis and K. Al Agha, "Quality of Service for Ad hoc Optimized Link State Routing Protocol (QOLSR)", IETF-63 Meeting, Internet Engineering Task Force, draftbadismanet-qolsr-02.txt, Vancouver, Canada, November 2005. Draft IETF.

[4] NS manual, available at: <http://www.isi.edu/nsnam/ns/ns-documentation.html>.

[5] Mario Joa Ng, "ROUTING PROTOCOL AND MEDIUM ACCESS PROTOCOL FOR MOBILE AD HOC NETWORKS", Ph.D. Thesis (Electrical Engineering), Polytechnic University, Hong Kong, January-1999.

[6] R. Ramanathan and M. Steenstrup, "Hierarchically organized multihop mobile wireless networks for quality-of-service support", ACM/Baltzer Mobile Networks and Applications, 3(1):101-119, 1998.

[7] Z. J. Haas and S. Tabrizi, "On some challenges and design choices in ad hoc communications", Proceedings of IEEE MILCOM'98, 1998.

[8] S. Murthy and J. J. Garcia-Luna-Aceves, "An efficient routing protocol for wireless networks", ACM Mobile Networks and App. J. Special Issue on Routing in Mobile Communication Networks, 1(2):183-197, 1996.

[9] C. E. Perkins, E. M. Royer, and S. R. Das, "Multicast operation of the ad hoc on-demand distance vector routing", Proceedings of Mobicom'99, pages 207-218, 1999.

[10] D. B. Johnson and D. A. Maltz, "The dynamic source routing protocol for mobile ad hoc networks", In Tomasz Imielinski and Hank Korth, editors, Mobile Computing, chapter 5, pages 153-181. Kluwer Academic Publishers, 1999.

[11] V. D. Park and M.S. Corson, "A Highly Adaptive Distributed Routing Algorithm for Mobile Wireless Networks," Proceedings of INFOCOM, pp. 1405-1413, 1997.

[12] Z. Haas, "A new routing protocol for the reconfigurable wireless networks", In Proc. of the IEEE Int. Conf. on Universal Personal Communications, 1997.

[13] P. Bose, P. Morin, I. Stojmenovic and J. Urrutia, "Routing with guaranteed delivery in ad hoc wireless networks", ACM DIALM 1999, 48- 5; ACM/Kluwer Wireless Networks, 7, 6, 609-616, November-2001.

[14] Palaniappan Annamalai, "Comparative Performance Study of Standardized Ad-Hoc Routing Protocols and OSPF-MCDS", Virginia Polytechnic Institute and State University, October-2005.

[15] L Xue, M S Leeson and R J Green, "Internet Connection Protocol for Ad Hoc Wireless Networks", Communications & Signal Processing Group, School of Engineering, University of Warwick, Coventry CV4 7AL-2004.

[16] Yuan Sun Elizabeth M. Belding-Royer, "Internet Connectivity for Ad hoc Mobile Networks",



Communications System Laboratory Nokia Research Center-2003.

- [17] Anne-Marie Kernnarrec, Frederic Le Mouel, FranCj:oise Andre, "Improvement of the QoS via an Adaptive and Dynamic Distribution of Application in Mobile Environment", Proceedings of the 19th IEEE Symposium on Reliable Distributed Susters (SRDS'00) 1060-9857/00 \$10.00 © 2000 IEEE, January-2000.
- [18] Gonzalo Camarillo, "Quality of Service routing in mobile wireless networks", Advanced Signalling Research Laboratory Ericsson, FIN-02420 Jorvas, Finland-2005.
- [19] Michael Gerharz, Christian Vogt, Christian de Waal, "Current Approaches towards Improved Quality-of-Service Provision in Mobile Ad-hoc Networks", Technical article, Computer Science Department IV Communications Systems, Rheinische Friedrich Wilhelm University at Bonn, Germany, March-2003.
- [20] C. E. Perkins, "Ad hoc Networking", Addison Wesley, 2001.
- [21] Gu, D; Zhang, J., "QoS Enhancement in IEEE802.11 Wireless Local Area Networks", IEEE Communications Magazine, ISSN: 0163-6804, Vol. 41, Issue 6, pp. 120-124, June 2003
- [22] L. Hanzo (II.) and R. Tafazolli, "Quality of Service Routing and Admission Control for Mobile Ad-hoc Networks with a Contention-based MAC Layer", Centre for Communication Systems Research (CCSR), University of Surrey, UK.- 2005.
- [23] Bracha Hod, "Cooperative and Reliable Packet-Forwarding On Top of AODV", Master of Science Thesis, School of Engineering and Computer Science The Hebrew University of Jerusalem Israel December 8, 2005.
- [24] Luiz Carlos Pessoa Albini, "Reliable Routing in Wireless Ad Hoc Networks: The Virtual Routing Protocol", Ph.D. Thesis, Via Buonarroti 2, 56127 Pisa, Italy, July-2006.
- [25] Rajarshi Gupta, "Quality of Service in Ad-Hoc Networks", PhD thesis, UNIVERSITY OF CALIFORNIA, BERKELEY, Spring-2005.

## Authors

**Tapan Kumar Godder** received the Bachelor's, Master's and M.Phil degree in Applied Physics & Electronics from Rajshahi University, Rajshahi. In 1994, 1995 and 2007, respectively. He is currently Associate Professor in the department of ICE, Islamic University, Kushtia-7003, Bangladesh. He has twenty one published papers in international and national journals. His areas of interest include internetworking, AI & mobile communication.



**M M Hassain** is professor in the department of Applied Physics and Electronics Engineering, Rajshahi University, Rajshahi 6205, Bangladesh. He is currently honorable Vice-Chancellor in Pabna Science and Technology University, Pabna,



Bangladesh.

**M. Mahbubur Rahman** received the Bachelor's and Master's Degree in Physics, Rajshahi University, in 1983, 1994 and PhD degree in Computer Science & Engineering in 1997. He is currently Professor in the department of ICE, Islamic University, Kushtia-7003, Bangladesh. He has twenty four published papers in international and national journals. His areas of interest include internetworking, AI & mobile communication.



**Md. Sipon Miah** received the Bachelor's and Master's Degree in the Department of Information and Communication Engineering from Islamic University, Kushtia, in 2006 and 2007, respectively. He is currently Lecturer in the department of ICE, Islamic University, Kushtia-7003, Bangladesh. Since 2003, he has been a Research Scientist at the Communication Research Laboratory, Department of ICE, Islamic University, Kushtia, where he belongs to the spread-spectrum research group. He is pursuing research in the area of internetworking in wireless communication. He has seven published paper in international and one national journals in the same areas. His areas of interest. include Wireless Communications, optical fiber communication, Spread Spectrum and mobile communication.



## SEWOS: Bringing Semantics into Web operating System

A. M. Riad<sup>1</sup>, Hamdy K. Elminir<sup>2</sup>, Mohamed Abu ElSoud<sup>3</sup>, Sahar. F. Sabbeh<sup>4</sup>

<sup>1</sup>Information system department, Faculty of computers and information sciences. Mansoura university, Egypt

<sup>2</sup>Department of Communication Misr Academy for Engineering & technology.

<sup>3</sup>Computer Science department, Faculty of computers and information sciences. Mansoura University, Egypt

<sup>4</sup>Alzarka Higher institute for administration & computer sciences, Damietta, Egypt

**ABSTRACT:** The revolution in web world led to increasing users' needs, demands and expectations. By the time, those needs developed starting from ordinary static pages, moving on to fully dynamic ones and reaching the need for services and applications to be available on the web!.. Those demands changed the perspective of our web today to what's said to be a cloud of computing that aims mainly to provide applications as services for web user. As time goes by, applications were just not enough; users needed their applications and data available anytime, anywhere. For these reasons, traditional operating system functionality was needed to be provided as a service that integrates several applications together with user's data. In this paper we present the detailed description, implementation and evaluation of SEWOS [1]- a semantically enhanced web operating system- that provides the feel, look and mimic traditional desktop applications using desktop metaphor.

**Keywords:** Web Operating System, Semantic, Ontology, Service Oriented Architecture.

### 1. INTRODUCTION

The World Wide Web has become a major delivery platform for a variety of complex and sophisticated applications in several domains. In this context, researchers investigated the ability to extend traditional web-based applications' functionality to enable users to interact with applications in much the same way as they do with desktop applications. Web operating systems were developed to provide users with an environment that pretty much resembles traditional desktop environment through web browser. They represent an advance in web utilities as they aim to provide better operational environments by moving users' working environment within web site including managing his/her files, installing his applications. Web operating system can be defined as a virtual desktop on the web, accessible via a browser as an interface designed to look like traditional operating system with multiple integrated built-in applications that allow user to easily manage and organize his data from any location[2]. Web operating system provides users with traditional operating system applications as services available for user to access transparently without any prior knowledge about where service is available, the cost or constraints [3]. In web operating system, applications, data files, configurations, settings and

access privileges reside remotely over network as services accessed by web browser which is used for input and display purposes [4].

As previously stated, web operating system – though its novelty - has drawn attention and many attempts have been made. WOS [3-9], the first known web-based operating system that provided a platform that enabled user to benefit from computational potential of the web. WOS provided users with plenty of tools through using a virtual desktop using the notion of distributed computing by replicating its services between multiple interacting nodes to manipulate user requests. WOS consists of three major components, graphical user interface, resource control unit which processes user request and finally a remote resource control unit which manages requests passed from other nodes.

The interest in web operating systems and their applications on academic communities resulted in VNet which was developed at the University of Houston and considered an access point to campus resources. VNet included variety of services that support students such as Desktop, admin management, contact management, file management services, calendar and scheduling services, report generation services, ... etc [10].

Based on the earlier work of WOS WEBRES was developed. WEBRES investigated the aspects of resource sharing that wasn't addresses in WOS and presented the notion of resource set which makes resources persistent rather than bounded to a specific user[11].

G.H.O.S.T (<http://g.ho.st/vc.html>), EyeOS "[www.eyeos.com](http://www.eyeos.com)" and DesktopTwo "[www.desktoptwo.com](http://www.desktoptwo.com)" are examples of systems that were built based on the trends of web operating systems. They mimic the look, feel and functionality of the desktop environment of an operating system. Moreover, they present variety of applications such as: File management, Address book, Calendar and text editing applications.

Implementing such application requires considering user's requirements in all phases as the final evaluation requires user participation and intervention. This paper is organized as follows; the next section presents SEWOS general architecture. In section 3, implementation of SEWOS and applications is provided. In section 4 presents the evaluation of the proposed system. Our conclusion and future work is presented in section 5.

## 2. THE PROPOSED ARCHITECTURE

Web operating systems as previously mentioned has the features and functionality of traditional desktop operating system. However, Web operating systems typically transfer applications to web server where user can manage his resources through virtual desktop using web browser. At the start of our research we had three main interests which we tried to satisfy.

- 1- Moving from fully personalized familiar desktop on PC to a virtual remote desktop, is a hard task, as users will accept nothing less than traditional desktop which they have been accustomed to. Thus, user data, preferences as well as sessions must be maintained ensuring that user will always has a personal experience that resembles his fully personalized traditional pc environment.
- 2- Semantic web technology plays a significant role in today's web as well as desktop systems [18-19]. That's why we thought that it was only a matter of time before semantic web techniques thrust in the research of web operating system.
- 3- A service-oriented architecture (SOA) is seen as the next evolutionary step in building web-based applications as it provides a set of principles of governing concepts used during phases of systems development. As in n-tier architectures SOA separates presentation/applications,

services and data into layers preventing dependency between layers.

In our work, we tried to merge the semantic web with web operating system utilizing the notion of SOA to support our architecture.

### 2.1 SEWOS ARCHITECTURE

SEWOS is SOA-based architecture that shows the underlying semantic file system of our semantic web operating system. SEWOS consists mainly of three layers, application layer, service layer and data layer as depicted in Figure 1. SEWOS architecture Application layer contains both user interface (portal) and application manager which in turn includes set of applications: file manager, word processing, spread sheets, web search and to-do list.

The second layer is service layer which includes transaction manager as well as personalization manager. Transaction manager controls user requests and works in correlation with both application layer and data layer in order to provide a virtual desktop. Personalization manager is responsible for generating a personalized desktop making use of user log, preferences and profile. Resource locator is used to locate where resources reside. Our architecture uses the notion of hybrid systems as it maintains a centralized resource location whereas resources themselves are decentralized. And finally, data layer contains back end databases that stores user profile, log file as well as user resources that are typical user files annotated using ontology. The next section embraces SEWOS implementation process.

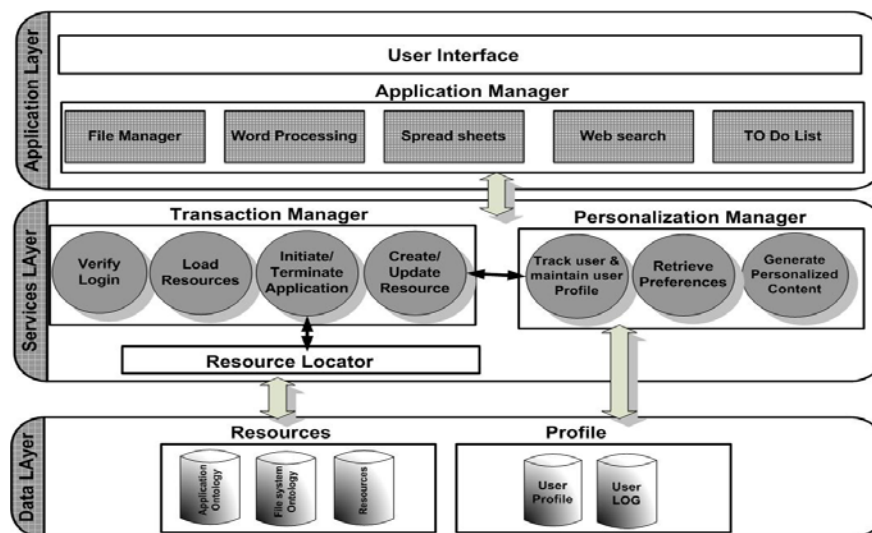


Fig.1 SEWOS Architecture

## 3. SYSTEM IMPLEMENTATION

SEWOS was developed based on SOA techniques. First, user must go through a registration process or an ordinary login for registered users. Afterwards, user will be able

to view his personalized desktop, access and manage his own resources and applications. SEWOS makes use of memorization as a personalization function, displaying a welcome message and a fills user's personalized start

menu with his recent file list, his events, his favorite resources and applications. Besides his start menu, user can start any application directly using application icon on his desktop. Moreover, user can start and deal with multiple applications at the same time. Options to manage workspace preferences are also available and accessible through personalized desktop.

The implementation of SEWOS home page and personalized desktop is shown in Fig 2.

**System's home page in Fig.8.a contains:**

- 1- **Welcome message/Log out:** system identifies user and displays a welcome message as an application to aforementioned salutation personalization function. The system also gives user the ability to log off at any time during navigation.
- 2- **Personalized work space:** this includes user's personalized background, calendar and clock. User can choose to display clock, calendar or not and he can choose his own background using preferences dialog.

3- **Personalized start menu:** User's start menus includes four tabs as follows:

- 1- **Recent tab:** This tab contains a list of user's personalized book-marking displaying a set of resources that were accessed during user's last visit.
  - 2- **Events tab:** This tab contains a list of user's events that are associated with today's date.
  - 3- **Favorite Files tab:** this tab contains a list of ranked files that are favorably accessed by user during that time of the day.
  - 4- **Favorite Applications tab:** Contains a list of SEWOS applications that are accessed by user during this time of the day.
- 4- **User Calendar and analog clock:** those two tools are added to user work space and can be hidden/ shown based on user preferences.

In the next section a detailed description of SEWOS's embedded applications, interface description...etc.

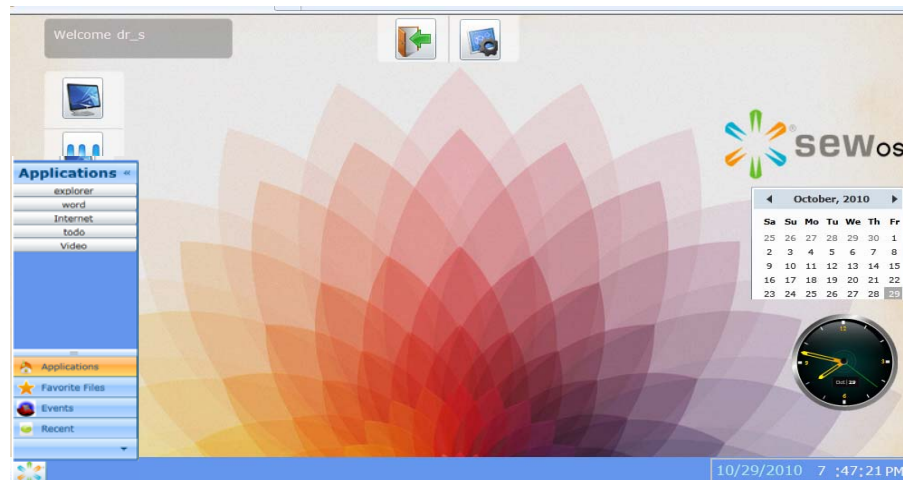


Fig.2. SWOS personalized Desktop

### 3.1 SEWOS FILE MANAGER

SEWOS file manager is a tool developed mainly to manage user's resources, this tool deals with resources and explores interrelationships for better processing. File manager interface and recycle bin are shown in fig.3 the main components are:

- 1- **Navigational tree:** This tree loads user's folders and arranges them in a hierarchal form to facilitate user navigating.
- 2- **Item viewer:** This Viewer is used to display user's folders as well as files. Used to access folders and enables user to select files for further processing.
- 3- **Recommendations:** System provides user with three recommended lists upon any

resource's selection in the viewer pane. These lists are as follows:

- a) **Accessed together list:** this list displays a list of ranked resources that are frequently accessed together with the selected resource during same sessions.
- b) **Same Type/same creation date list:** This list displays a list of related files based on their type and related folders based on their creation date.
- c) **Related content List:** this list displays a set of resources ranked according the degree of content relevancy between each resource and the selected resource.



**4- Function Buttons:** File manager has capabilities to create new folder, upload/download and delete resource.

As previously stated, this manager's functionality is incomplete unless there exists a way for user to restore his deleted files. This includes having a personalized recycle bin which we consider as a main part of SEWOS File system. This is described in the next section.

### 3.1.1 SEWOS RECYCLE BIN

SEWOS Recycle bin completes the functionality of the underlying file system by acting as intermediate storage space for user's resources before they can be permanently deleted from the system. Recycle bin includes options either to restore deleted resource or to delete it permanently from system.

### 3.2 SEWOS TEXT EDITOR

SEWOS Text editor enables creating, viewing, editing, formatting, annotating, printing and saving text files. The application interface can be shown in Fig.4, this contains:

#### 1) Clipboard section

This section includes buttons that provides the basic copy, cut and paste functions.

#### 2) Font section

This section includes buttons that provides the main formatting options. This includes changing fonts, font size, color and alignment of the selected text.

#### 3) Insert section

This section includes the basic options to insert pictures, tables and hyperlinks within text.

#### 4) File operations section

This section includes buttons that enables:

- Creating new document.
- Opening an existing document with extensions (.txt, .sav and .docx).
- Saving user's documents to user's space with an extension (.sav).
- Print preview of user document.
- Printing user's document to user's printer.
- Displaying XML code behind document authoring.

### 3.3 SEWOS WEB BROWSER AND SEARCH APPLICATION

Navigating the web is one of the main activities of almost every computer users, that's why SEWOS includes this application. Application's interface pretty much resembles the basic interface of web browser. With an address bar to write required URL and Go button to navigate directly to it. This application includes as well an interface to our developed personalized semantic search engine (PSSE) using a search button. Web browser and search application are both shown in Fig.5 (a, b).

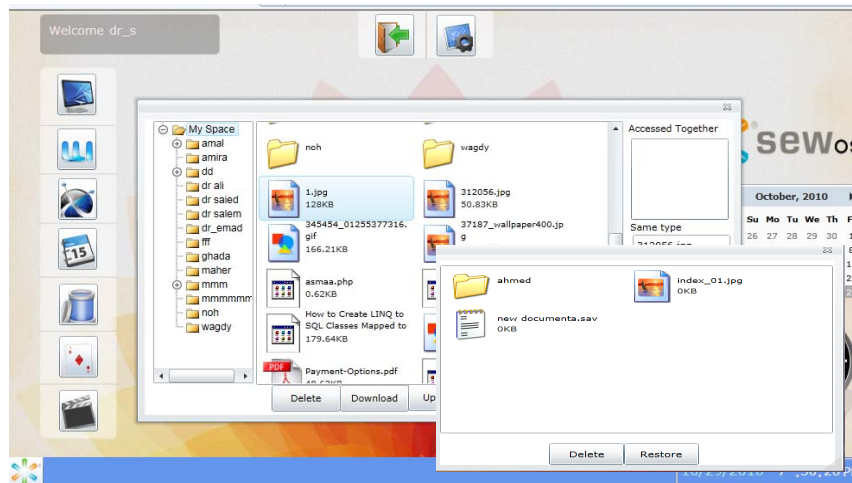


Fig.3 SEWOS File Manager and Recycle Bin

### 3.4 SEWOS CALENDAR

User's schedules and events are required to form user's every day to-do list, SEWOS calendar allows user to add his events and schedules for later retrieval. Important event are retrieved when user first logs into the system while related and similar events during

the current week can be displayed through Calendar application. SEWOS calendar includes displaying, adding and deleting applications as shown in Fig.6.

### 3.5 SEWOS VIDEO PLAYER

For their significance, multimedia files constitute huge part of today's web and user's resources. This tool is intended to provide user a way to open and view his multimedia files including both audio and video. Application's

interface as depicted in Fig.8 includes an open button, voice control section, play/pause button and two recommendation lists. Those lists retrieve resources related by both type and content



Fig.4 SEWOS Text Editor

### 3.6 SEWOS GAMING

Gaming and entertainments gain importance to user during his breaks and leisure times. SEWOS has a

built-in gaming application for the sake of user's entertainment. This application is shown in Fig.8.

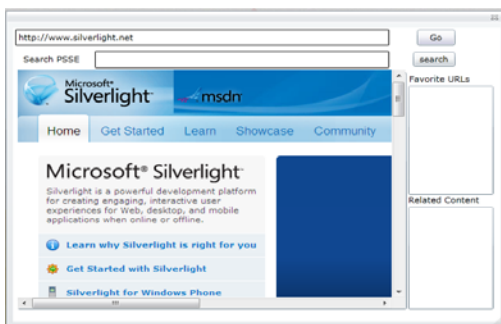


Fig.5(a) SEWOS Web Browser

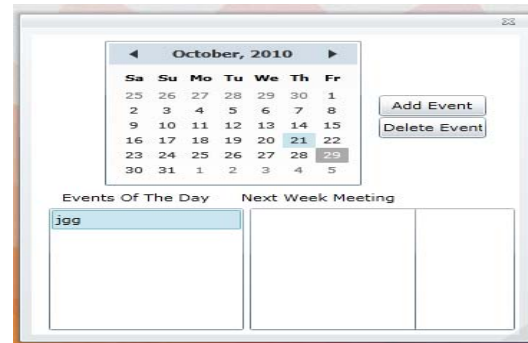


Fig.6 SEWOS Calendar

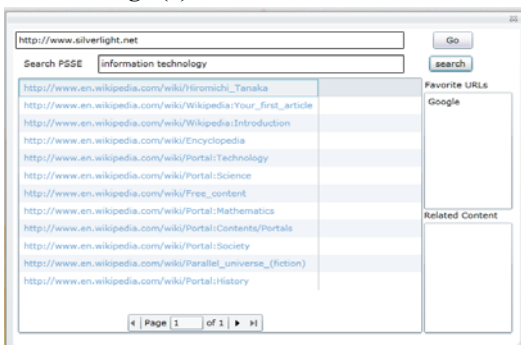


Fig.5(b) SEWOS Web Search (PSSE)

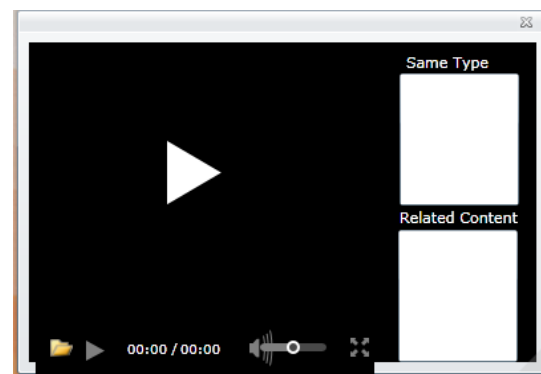


Fig.7 SEWOS Video Player



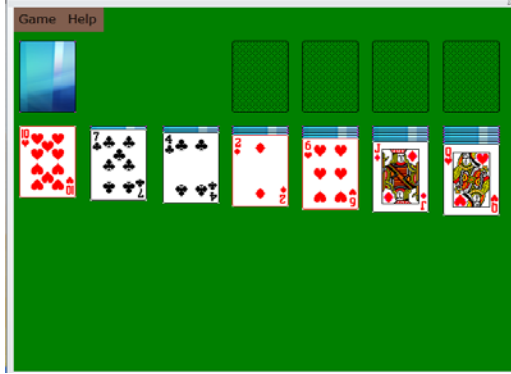


Fig.8 SEWOS Gaming

#### 4. SEWOS EVALUATION

Many aspects of usability can best be studied by simply asking users. This is especially true for issues related to users' subjective satisfaction and possible anxieties. Since the system is highly dependent on user participation, the design team from the beginning has taken steps in collecting and analyzing user feedback. Evaluation of the system has been from the beginning an integral part of our Participatory Design implementation. For this sake, a questionnaire was provided for collecting feedback about the general usability of the system as well as user satisfaction about each of the embedded applications. Secondly, standard evaluation measures for information retrieval techniques were used to evaluate the performance of our proposed personalized semantic search engine (PSSE).

##### 4.1 QUESTIONNAIRE

Twenty five experienced users responded to the questionnaire assessing the overall usability of the system. Questionnaire consists of forty four usability questions to which the respondent was to evaluate based on a five point likert-scale, ranging from 1 to - 5 (represent from 20% to 100% satisfaction). Questionnaire was divided into seven main categories that represent assessment for each of the individual applications embedded in the proposed system, in addition to a section for users' suggestions and comments. The data responses to the questionnaires were entered in a spreadsheet, analyzed and descriptive statistical analyses were performed. After careful investigation of data, we can assure that the overall satisfaction of the participants with system was high. Frequencies were analyzed to show that 76% of users were 100% satisfied with SEWOS functionality; whereas 16% were 80% satisfied and only 8% were 60% satisfied about the system. For SEWOS file manager, it gained 100% satisfaction of 64% of the participants, 80% by 20% of the participants and 16% were 60% satisfied. For the rest of the statistics, Table 1 includes the percentage of

participants for SEWOS and applications based on the available scales.

	100%	80%	60%	40%	20%
SEWOS Functionality	76%	16%	8%	0%	0%
SEWOS File Manager	64%	20%	16%	0%	0%
SEWOS Text Editor	72%	20%	8%	0%	0%
SEWOS web browser	76%	16%	8%	0%	0%
SEWOS Calendar	88%	12%	0%	0%	0%
SEWOS Video	72%	16%	12%	0%	0%
SEWOS gaming	56%	24%	4%	0%	0%

Table 1: Percentage of users votes with respect to each scale

Comments reflected a desire for adding more applications that both help desktop and web activities, improvement of the capabilities of the system by adding context menus and more visual aids.

A graphical representation of the overall ratings for all categories is provided in Figure 9. The former evaluation for our system depended on measuring user satisfaction of SEWOS and applications usability.

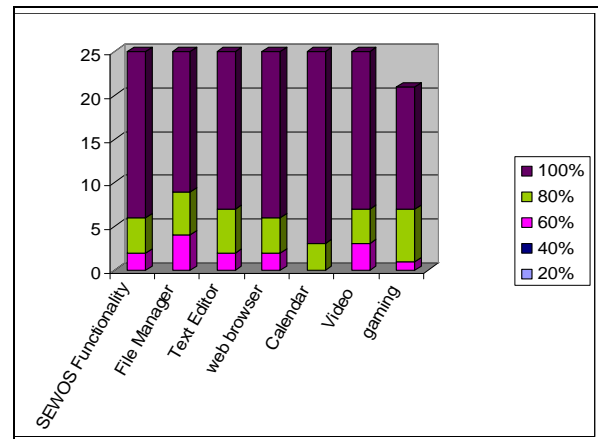


Fig.9: Plot of users ratings for SEWOS and Applications

#### 5. CONCLUSION AND FUTURE WORK

Gaining users' satisfaction was our main goal and motivation when developing SEWOS. Now, after statistical analysis of participants' ratings, we can say that impressions from the evaluation of our data were in favor of our proposed system. Our future work includes providing large-scale evaluation of SEWOS as well as investigating users' feedback and provides more applications that better suits SEWOS users' need. Moreover, we'll try to provide integration between SEWOS and services available on the World Wide Web.

## References

- [1] A.M.Riad, Hamdy k. Elminir, Mohamed Abu ElSoud, Sahar F. Sabbeh. " Sewos: a framework for semantic web operating system". International Journal of Electrical & Computer Sciences IJECS-IJENS Vol:10 Issue: 1 No.1. 2010.
- [2] G. Lawton, "Moving the OS to the Web," Computer, vol. 41, no. 3, pp. 16-19, 2008.
- [3] P. G. Kropf, J. Plaice, H. Unger. "Towards a Web Operating System (WOS)". In Proceedings of webnet'1997.
- [4] N. A. Kofahi, A. Al-Taani, "Web Operating system and Computing On The Web", Information Technology Journal Vol. 4 Issue 4, P. 360-366, 2005.
- [5] P. G. Korpff, "Overview of The WOS Project", Advanced Simulation Technologies Conference (ASTC1999). San Diego, California, USA, pp.~350--356, 1999.
- [6] A. Mufti, K. Salah, "Web Operating System", <https://eprints.kfupm.edu.sa/794/1/webos.pdf>, 2001. (Retrieved 5-2009)
- [7] A. Vahdat, T. Anderson, M. Dahlin, E. Belani, D. Culler, P. Eastham, C. Yoshikawa. "webos: Operating system services for Wide Area Applications", In Proceedings of the Seventh IEEE Symposium on High Performance Distributed Systems, 2002.
- [8] A. Vahdat, T. Anderson, M. Dahlin, E. Belani, D. Culler, P. Eastham, C. Yoshikawa. " webos: Software Support for Scalable Web Services". Proc. Of the 7th IEEE International Symposium on High Performance Distributed Computing, Aug. 1998, pp. 52-63.
- [9] N. Abdennadher, G. Babin, P. Kropf. " A WOS<sup>TM</sup> - Based Solution For High Performance Computing", IEEE Computer Society Press, pp. 568-573. May 2001.
- [10] S. B. Franceschi, L. Le, D. Velez. "Web-Based Technologies: Reaching Their Ultimate Potential On Restricted Budgets", Proceedings of the 32nd annual ACM SIGUCCS conference on User services, p.336-339, 2004.
- [11] O. Krone, S. Schubiger. "WEBRES: Towards a Web Operating System", Kommunikation in Verteilten Systems. P.418-429, 1999.

in automatic control system in 1996. He obtained his PhD degree from the Czech Technical University in Prague in 2001. Currently he is an associate professor and the head of communication department – masr academy for engineering, Mansoura, Egypt.



Sahar F. Sabbeh was born in Damietta, Egypt in 1982. She received the B.Sc. in Information systems from Mansoura University, Egypt in 2003 and completed her master degree in Information systems 2008. Currently she is an assistant lecturer in Alzarka Higher Institute For Computer And Administration Sciences, Damietta, Egypt.



A.M. Riad - Head of Information Systems department, Faculty of Computers and Information Systems, Mansoura University. Graduated in Mansoura University from electrical engineering department in 1982.

Obtained Master degree in 1988, and Doctoral degree in 1992. Main research points currently are intelligent information systems and e-Learning.



Hamdy K. Elminir was born in EI-Mahala, Egypt in 1968. He received the B.Sc. in Engineering from Monofia University, in 1991 and completed his master degree

# Segmenting and Hiding Data Randomly Based on Index Channel

Emad T. Khalaf<sup>1</sup> and Norrozila Sulaiman<sup>2</sup>

<sup>1,2</sup> Faculty of Computer Systems & Software Engineering, University Malaysia Pahang,  
Kuantan, 26300, Malaysia

## Abstract

Information hiding is a technique of hiding secret using redundant cover data such as images, audios, movies, documents, etc. In this paper, a new technique of hiding secret data using LSB insertion is proposed, by using the RGB channels of the cover image for hiding segmented data. One of the three channels became the index to the two other channels. Firstly, the secret data are segmented into Even segment and Odd segment. Then, four bits of each segment is hidden separately inside the two channels depending on the numbers of "1"s inside the index channel. If the numbers were Even, then four bits of Even segment will be hidden. However, if they were Odd then four bits of Odd segment will be hidden. The opposite process retrieve the secret data from image by reading the bits of the index channel and check the numbers of "1"s to extract the Even segment and Odd segment. Finally, recombining the two segments to extract the secret data. Experimental results show that the proposed method can provide high data security with acceptable stego-images.

**Keywords:** *Steganography, Data hiding, Data segmenting, Index channel*

## 1. Introduction

Information security requirement became more important, especially after the spread of Internet applications [1]. However, Owners of sensitive documents and files must protect themselves from unwanted spying, copying, theft and false representation. This problem has been solved by using a technique named with the Greek word "steganography" it is mean hiding information [2]. Steganography is the art and science of hiding information. The data-hiding system design challenge is to develop a scheme that can embed as many message bits as possible while preserving three properties: imperceptibility, robustness, and security [4]. In addition, proposing an effective method for image hiding is an important topic in recent years [5],[6]. There have been many techniques for hiding information or messages in images in such a way that the alterations made to the image are perceptually indiscernible. Common approaches include [7]:

(i) Least significant bit insertion (LSB)

(ii) Masking and filtering  
(iii) Transform techniques

Information hiding is an emerging research area, which encompasses applications such as copyright protection for digital media, watermarking, fingerprinting, and steganography [8]. All these applications of information hiding are quite diverse [8] and many encoding methods was proposed, a reversible image hiding scheme based on histogram shifting for medical images was proposed in [5]. An image-in-image hiding scheme, based on dirty-paper coding, that is robust to JPEG and additive white Gaussian noise (AWGN) attacks was proposed in [9]. Chen et al. [10] used a vector quantification method, but the method required a set of look-up tables. Moreover, the decoded images were little distorted from original images. Wang et al. [11] proposed a least significant bit technique to hide information. The technique could improve the visual quality of cover images, but the reconstruction processes were very complicated calculations. Chang et al. [12] proposed two kinds of hiding techniques and the hiding techniques secured better visual quality. However, the information capacity of these hiding techniques was low. Yang and Lin [13] used a basal-bit orientation method to hide images, and the method had large hiding capability and good visual quality of the secret image. In this paper we proposed a new method of segmenting and hiding the secret data in bmp color image by segmenting these data into two segment, i.e. Even segment and Odd segment. Then those two segments of characters will be hidden separately and randomly inside the cover image. By using random pixels to insertion secret data with modifying those data, this could avoid the detection by comparison of modified image with original image [3]. Two channels were used for hiding data in 24-bit BMP image and the third channel was used as index channel for the hidden data.

## 2. Steganography Techniques

Steganography is the art of embedding information in such a way that prevents the detection of hidden messages. It means hiding secret messages in graphics, pictures, movie, or sound. Steganography comes from the Greek word steganos, which means 'covered', and -graphy, which means 'writing'. Covered writing has been manifested way back during the ancient Greek times around 440 B.C. Some of old steganography examples are shaving the heads of slaves and tattoo messages on them. Once the hair had grown back, the message was effectively hidden until the receiver shaved the heads once again. Another technique was to conceal messages within a wax tablet, by removing the wax and placing the message on the wood underneath [14]. The most popular and frequently method of Steganography is the Least Significant Bit embedding (LSB). The level of precision in many image formats is far greater than that perceivable by average human vision. Therefore, an altered image with slight variations in its colors will be indistinguishable from the original by a human being, just by looking at it. If we using the least significant bits of the pixels' color data to store the hidden message, the image itself is seemed unaltered [15],[16] and changing the LSB's value will have no effect on the pixel's appearance to human eye. In our method, the 24-bit BMP image and least significant bit (LSB) insertion were used. The reason behind using BMP type is that it is more accurate in showing the image without any of compressed data and it is considered to be the most used format in hiding operation and analyzed.

## 3. The Proposed Method

A new steganography technique of uses the RGB images to hide the data in different channels was proposed. Two files are require to embedding a message into an image. The first is the message (the information to be hidden), a message may be plain-text, cipher-text, other images, or anything that can be embedded in a bit stream. The second file is the innocent-looking image that will hold the hidden information, called the cover image. Generally, Digital images are stored in computer systems as an array of points (pixels) where each pixel is consisting of three channels: (Red, Green, and Blue) and  $0 \leq R,G,B \leq 255$  [17]. In our method the data is hidden into two of the RGB

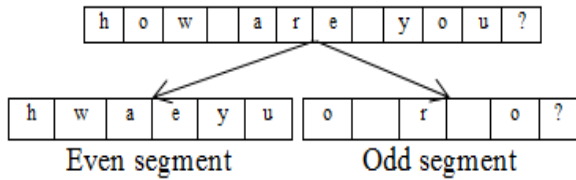
pixel channels based on the third channel. However, when using a 24 bit color image, two bits of two colors components can be used, so a total of 2 bits can be stored in each pixel

At the beginning, the secret data is split programmatically into array of characters and this array was segmented into two segments, Even segment and Odd segment

For example:

The secret data is: "how are you?"

Thus, the segmenting process will be:



Each segment is hiding separately in random two channels. One of the three channels was used as index channel to the next two channels by counting the number of "1"s in the index channel. If it is Even, then four bits of the Even segment data will be hidden inside the least significant bit of channel 1 and the least significant bit of channel 2. If it is Odd, then four bits of the Odd segment data will be hidden inside the least significant bit of channel 1 and the least significant bit of channel 2, The following example explain the idea.

Suppose that three adjacent pixels (nine bytes) with the following RGB encoding are used.

	<u>Index Ch.</u>	<u>Channel1</u>	<u>Channel2</u>	
<u>Pixel(1):</u>	<u>10010101</u>	<u>00001100</u>	<u>11001001</u>	<i>Even</i>
<u>Pixel(2):</u>	<u>11010111</u>	<u>00001110</u>	<u>11001011</u>	<i>Even</i>
<u>Pixel(3):</u>	<u>10011011</u>	<u>00010000</u>	<u>11001010</u>	<i>Odd</i>

Now, in pixel1 4bits from Even segment of data will be hidden because number of "1"s in index channel is Even. In addition, 4bits from Even segment of data will be hidden in pixel2. However, in pixel3, 4bits of Odd segment will be hidden because number of "1"s in index channel is Odd. In this example, the red color was used as index channel. To improve security, the index channel is not fixed. The indexes are chosen sequentially, the first index is Red, and the subsequent indexes are Green and

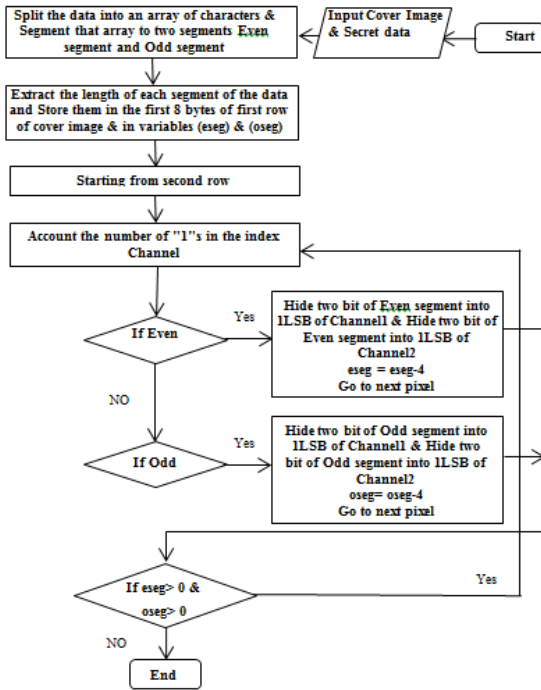
Blue respectively. The index LSB bits are naturally available at random, based on image profile and its properties. Table 1 shows the relationship between the index channel and the hidden data inside the other channels.

Table 1. Meaning of Index channel values

No of "1"s in the index Channel	Channel1	Channel2
<i>Even</i>	<i>2bit of Even segment</i>	<i>2bit of Even segment</i>
<i>Odd</i>	<i>2bit of Odd segment</i>	<i>2bit of Odd segment</i>

As shown in the table, if the index channel is Red, channel1 will be Green and channel 2 will be Blue and the sequence will be RGB. In the second pixel, the index is Green. Channel 1 and channel 2 will be Red and Blue respectively. Hence, the sequence is GRB. In the third pixel, the index is Blue. Therefore, channel 1 is Red and channel 2 is Green. The sequence is BRG. The processes method is as shown in Figure 1. First process is used to input the cover image and the secret data. Then, the data will be segmented and the length of each segment will be stored in the first 8 bytes of the beginning of the image. The hiding process starts from the second row and it depends on the numbers of "1"s in the index channel.





The recovery processes for the proposed method is shown in Figure 2. It is the exact reverse of the hiding process, starting with extracting the two segment of the data length from the first 8 bytes of the image.

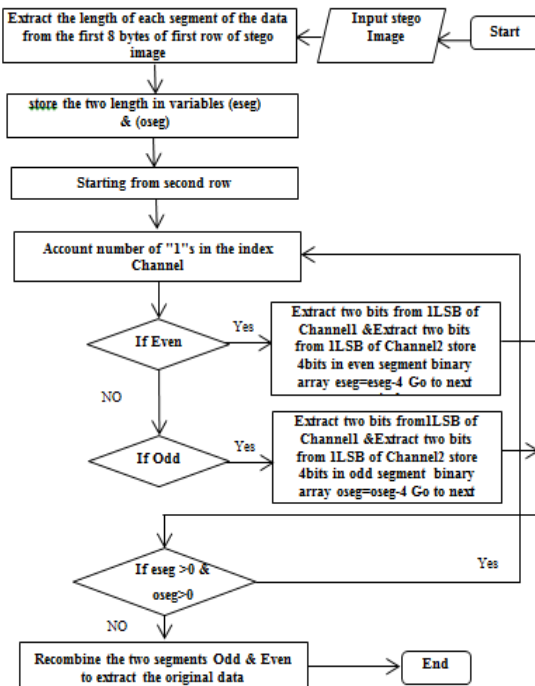


Fig. 2 Recovery and Recombine process flowchart

## 4. Experimental Results and Analysis

We have tested our algorithm for different sets of images as well as text messages. Histogram analysis has been implemented to the image before and after embedding data, to compare the original channels, before and after modifying channels. This can give a clear idea of the security and if change is minimal, then the method is considered secure. Figure 3a shows the original image of Mosul City and Figure 3b shows the stego image of Mosul City. Another image is an original image of a bird, as shown in Figure 4a and its stego image is as shown in Figure 4b. The Red, Green and Blue histograms for Mosul City image is as shown in Figure5, Figure6, Figure7 and Figure8. The modified images after applying the method did not show any identifiable visual difference.



Fig. 3a: Original image (Mosul City) length 500x330



Fig. 3b: Stego image (Mosul City) with text size 1420 characters



Fig. 4a: Original image (Bird) length 760x570



Fig. 4b: Stego image (Bird) length 760x570 with text length 2300

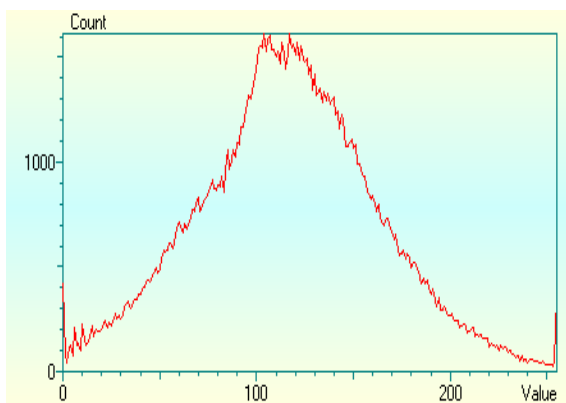


Fig. 5a: Original image (Mosul City) histogram of Red channel



Fig. 5b: Modified image (Mosul City) histogram of Red channel

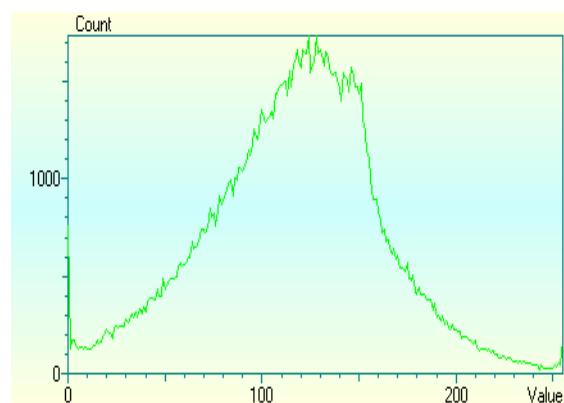


Fig. 6a: Original image (Mosul City) histogram of Green channel

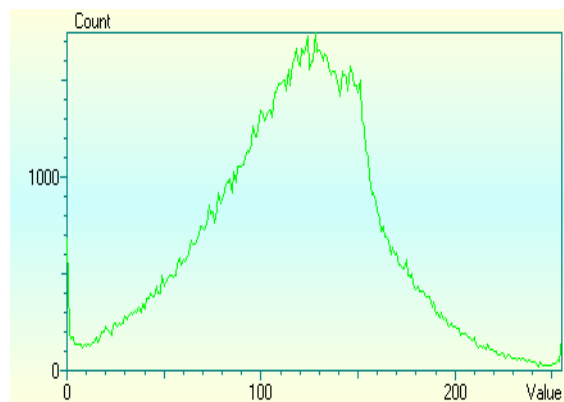


Fig. 6b: Modified image (Mosul City) histogram of Green channel

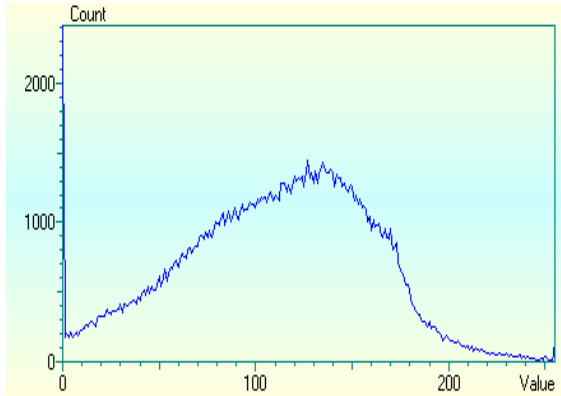


Fig. 7a: Original image (Mosul City) histogram of Blue channel

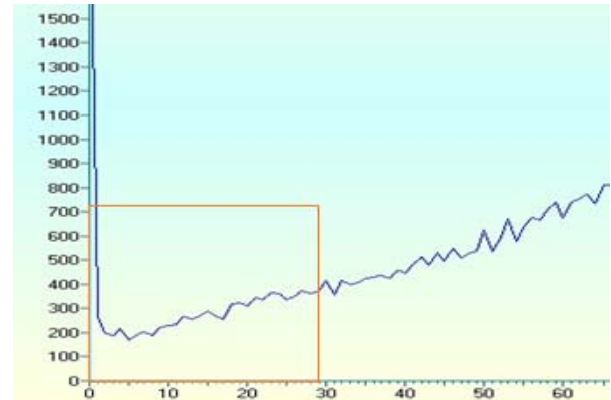


Fig. 8b: Histogram Zooming of Blue channel of Modified image (Mosul City)

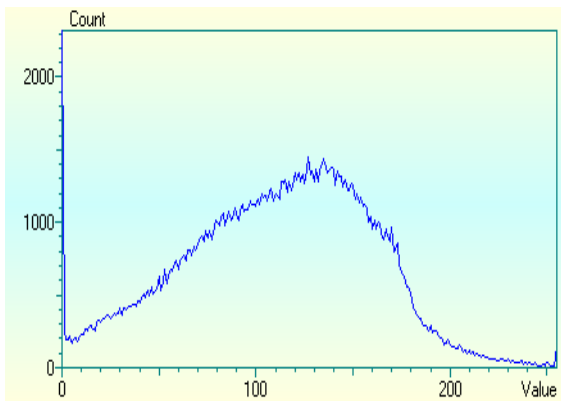


Fig. 7b: Modified image (Mosul City) histogram of Blue channel

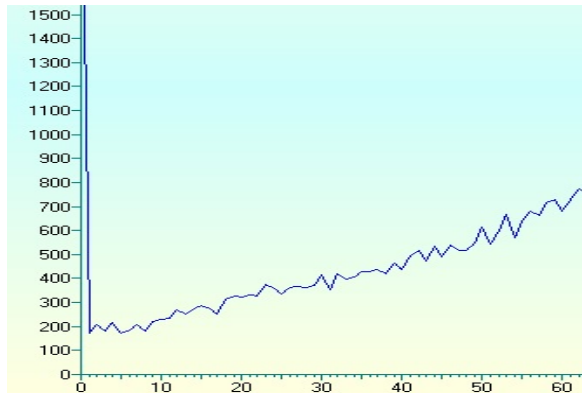


Fig. 8a: Histogram Zooming of Blue channel of original image (Mosul City)

By comparing the three RGB channels before and after hiding the data, the security aspect can be discussed. By observing the channels histograms before and after the modification, i.e. Figure 5a, Figure 5b, Figure 6a, Figure 6b, Figure 7a & Figure 7b, the change cannot be easily detected. However, if part of histogram is enlarged, some changes in the curve can be seen as shown in Figure 7a & Figure 7b. This creates some future work to be investigated including the reasons and implications of this issue. From many test runs, different distributions between the three channels were identified, which continued varying between the channels with no detected pattern. This undetectable pattern changing within RGB channels promise that the proposed technique may be considered random or pseudorandom based on the randomness of the index channel. Imperceptibility takes advantage of human psycho visual redundancy, which is very difficult to quantify for image steganography, existing metrics to measure imperceptibility include mean-square-error (MSE) and peak-signal-to-noise ratio (PSNR) [18] which is defined as:

$$MSE = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (f_{ij} - g_{ij})^2$$

$$PSNR = 10 \log_{10} \frac{L^2}{MSE}$$

where

$M, N$  are the row and column numbers of the cover image,  
 $f_{ij}$  is the pixel value from the cover image,  
 $g_{ij}$  is the pixel value from the stego-image, and

$L$  is the peak signal level of the cover image (for  $\&bit$  gray-scale images,  $L = 255$ ).

Table 2 shows the values of PSNR and MSE with different sizes of images. Referring to Table 2, the column labeled SHDRIC is our proposed

Table 2: (PSNR and MSE) of four sample images

Image name	MSE	SHDRIC PSNR	Simple LSB PSNR
Image 1	26.919	33.830	31.760
Image 2	8.801	38.686	35.554
Image 3	1.241	47.192	43.959
Image 4	0.499	51.146	51.083

From the table, it was noted that the increase in the text caused an increase in the MSE and decrease in the PSNR. However, we can see the improvement in PSNR values than in the simple LSB. So, it becomes difficult to discover the hidden text within the image.

## 5. Conclusion

The suitability of steganography as a tool to conceal highly sensitive data has been discussed using a new method of randomizing the secret data. The method is based on two level of security where the data will be segmented into even and odd segments, before hiding the two segments separately and randomly inside image. This suggests that an image containing encrypted data can be transmitted anywhere across the world, in a complete secured form. This method can use in any other application such as image watermarking. It can be concluded that randomizing and hiding the secret data can provide a double layer of protection.

## References

- [1] J. Anitha and S. Immanuel Alex Pandian" A Color Image Digital Watermarking Scheme Based on SOFM" International Journal of Computer Science Issues, Vol 7, Issue 5, Sept 2010, Pages 302-309.
- [2] Mayank Srivastava, Mohd. Qasim Rafiq and Rajesh Kumar Tiwari "A Robust and Secure Methodology for Network Communications" International Journal of Computer Science Issues, Vol. 7, Issue 5, September 2010, Pages 135-141
- [3] Geeta S. Navale ; Swati S. Joshi ; Aarad\_ hana A Deshmukh "M-Banking Security – a futuristic improved security approach", International Journal of Computer Science Issues, Vol 7, Issues 1, Jan 2010, Pag. 68-71
- [4] I. Cox, M. Miller, and J. Bloom, Digital Watermarking, Academic Press, 2002.
- [5] P. Tsai, etal. Reversible image hiding scheme using predictive coding and histogram shifting, Signal Processing, vol. 89, pp. 1129-1143, 2009.
- [6] H. Sajedi, M. Jamzad, Cover Selection Steganography method Based on Similarity of Image Blocks, in Proc. of Int. IEEE 8th Conference on Computer and Information Technology, 2008.
- [7] N.F. Johnson and S. Jajodia, Exploring Steganography: Seeing the Unseen, IEEE, pp. 26-34, 1998.
- [8] R A Isbell, Steganography: Hidden Menace or Hidden Savior, Steganography White Paper, 10 May 2002.
- [9] K. Solanki, N. Jacobsen, U. Madhow, B.S. Manjunath, and S. Chandrasekaran, Robust image-adaptive data hiding using erasure and error correction, IEEE Trans. Image Processing, vol. 13, pp.1627–1639, Dec. 2004.
- [10] T. S. Chen, C. C. Chang, and M. S. Hwang, A virtual image cryptosystem based on vector quantization, IEEE Trans. on Image Process, 7 (1998) 1485.
- [11] R. Z. Wang, C. F. Lin, and J. C. Lin, Image hiding by LSB substitution and genetic algorithm, Pattern Recogn., 34 (2001) 671.
- [12] C. C. Chang, J. C. Chung, and Y. P. Lau, Hiding data in multitone images for data communications, IEE Proc. of Vision Image Signal Process, 151 (2004) 137.
- [13] C. Y. Yang and J. C. Lin, Image hiding by base-oriented algorithm, Optical Eng-ineering, 45 (2006) Paper No. 117001
- [14] Peter Wayner, Disappearing Cryptography –Information Hiding: Steganography & Watermarking–Second Edition. San Fransisco, California, U.S.A.: Elsevier Science, 2002, ISBN 1558607692.
- [15] Neil F. Johnson and Sushil Jajodia, Exploring Steganography: Seeing the unseen IEEE transaction on Computer Practices. 1998.
- [16] Ross Anderson, Roger Needham, Adi Shamir, The Steganographic File System, 2nd Information Hiding Workshop, 1998.



- [17] Dung Dang, Wenbin Luo " Color image noise removal algorithm utilizing hybrid vector filtering" AEU-International Journal of Electronics and Communications, Vol 62, Issue 1, 2 Jan 2008, Pages 63-67
- [18] Qi, Hairong; Snyder, Wesley E. & Sander, William A., 2002; Blind Consistency-Based Steganography for Information Hiding in Digital Media. Multimedia and Expo, 2002. ICME '02. Proceedings. 2002 IEEE International Conference on Vol. 1, p.: 585- 588.



**Emad T. Khalaf**

Graduated in Computer Information Systems and Informatics Engineering and he worked as a Technical in Internet Services Company for more than nine years. He had experience as a trainer for various computer courses. His research interests include network technology and security. He is currently studying MSc degree in the area of computer networks security.

**Norrozila Sulaiman**



Graduated from Sheffield Hallam University with a BSc (Hons) in Computer Studies in 1994. She worked with Employment Service in UK as a network support assistant and she involved on a research on Novell Netware. After graduated, she worked as a research officer at Artificial Intelligence System and Development Laboratory and involved in joint collaboration projects between the government of Malaysia and

Japan for about 5 years. She completed her MSc degree in Information Technology and involved in a research on Wireless Application Protocol (WAP). She obtained her PhD degree in mobile communication and networks from Newcastle University in UK. Currently, she is a senior lecturer at Faculty of Computer System and Software Engineering, University Malaysia Pahang. Her main research interests include heterogeneous networks, mobile communication networks and information security.



# Data-Acquisition, Data Analysis and Prediction Model for Share Market

Harsh Shah<sup>1</sup>, Prof. Sukhada Bhingarkar<sup>2</sup>

<sup>1</sup> MAEER'S MIT College of Engineering,  
Pune-38, Maharashtra, India.

<sup>2</sup> MAEER'S MIT College of Engineering,  
Pune-38, Maharashtra, India.

## Abstract

Data-acquisition involves gathering signals from measurement sources and digitizing the signal for storage, analysis and presentation on a PC. Analysis and prediction is very necessary in today's market for the accurate utilization of funds at hand. For analysis, there has to be a proper system where in the required data is first acquired from the destination. This data then needs to be analysed using any analysis model. Currently there are many analysis models available in the market. These models are based on the past behaviour of the stocks. However, it is seen that there is no model which predicts the future behaviour of the stocks. For this reason, a model is developed which not only analyses the stocks but also predicts its future behaviour based on the past conduct.

**Keywords:** *Data-acquisition, share-market analysis, share-market predictions.*

## 1. INTRODUCTION

Data-Acquisition systems are in great demand in the industry and consumer applications. Data-acquisition systems are defined as any instrument or computer that acquires data from sensors via amplifiers, multiplexers, and any necessary analog to digital converters or the internet. The system then returns data to a central location for further processing. An acquisition unit is designed to collect data in their simplest form from the internet.

Now-a-days Data-Acquisition systems are used more and more as these systems provide precise accuracy. Also, these systems remove the overhead of constant monitoring. A single person can monitor the entire system and also interact with the system if required. These systems enable the user to analysis the acquired data and also produce required predictions. Data-Acquisitions can be a very tedious task or even virtually impossible if these systems were not in place. These systems have allowed us to make more accurate, reliable and fool-proof data sharing, data analysis and data collection.

Share-Market Analysis is an important part of market analysis and indicates how well a firm is doing in the

market place compared to its competitors. Analysis helps the share broker to carefully study the behaviour of the stocks and utilize his funds in a more veracious way. Analysis of stocks takes into consideration the past behaviour of the particular stock and analysis shown to the user in the form of graphs. These graphs can be represented in a number of ways depending on the preferences of the users.

In this paper, the share market analysis and prediction model is proposed. This model is established using a reliable data-acquisition system which acquires data from the internet. This data is then analysed using the analysis module. After analysing the data the prediction module starts working. It does its calculations and the resulting predictions are recorded in table format and are reflected on the graphs.

## 2. RELATED WORK

There are data-acquisition and control devices that will be a substitute for a supervisor in a multisite job operation. A single person can monitor and even interact with the ongoing work from a single base station. An acquisition unit designed to collect data in their simplest form is detailed in [1]. Data collection via wireless internet-based measurement architecture for air quality monitoring is discussed in [2]. Some applications adding remote accessibility are detailed in [3] and [4], which are built to collect and send data through a modem to a server. Some applications have integrated systems for data-acquisitions. One such system is used in [5].

There are a number of analysis models that are available. These models provide analysis as desired by the user. One such model is discussed in [6]. This is stock market software, which supports multiple countries' stock market. (11 countries at this moment) It provides Real-Time stock info, Stock indicator editor, Stock indicator scanner, Portfolio management and Market chat chat

features. One more such model is shown in [7]. It provides a free web based stock price analysis module. The easy to use interface incorporates Fundamental Analysis to calculate: Fair Value stock price; comparative stock Value; profit Target sell price; Stop Loss sell price; Price Earnings Ratio (PE) for Fair Value and Buy prices; stock Return on Investment %; and provides access to Technical Analysis charts to evaluate stock movements and buy/sell signals.

In section 3, a software analysing and making predictions for share market is introduced. In section 4, an example that shows a working model of the discussed software is presented. This section also makes comparisons of the discussed software with the currently available software's. Section 5 presents the conclusion.

### 3. PROPOSED SOFTWARE

In the proposed software, the real-time data from the share market is taken from the internet. This data is then processed and analysed. After analysis, predictions for each stock are calculated using formulas. Thus the proposed software is divided into three main modules viz (3.1) Data-acquisition. (3.2) Data-analysis and (3.3) Prediction Model.

#### 3.1 Data-acquisition

The data for the stocks in the market is acquired from the internet. This data comes in the DBF file form. A snapshot of this file is shown in Fig1.

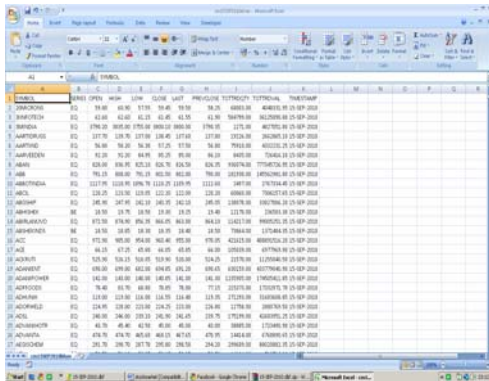


Fig. 1. DBF file.

The required data from this file i.e. the highest, lowest and the closing price of each stock for the particular day is extracted and forwarded to the analysis module. A sample of this file is shown in Fig 2. This process is done every day as each stock can have different values each day. This also helps in better analysis and more accurate predictions.

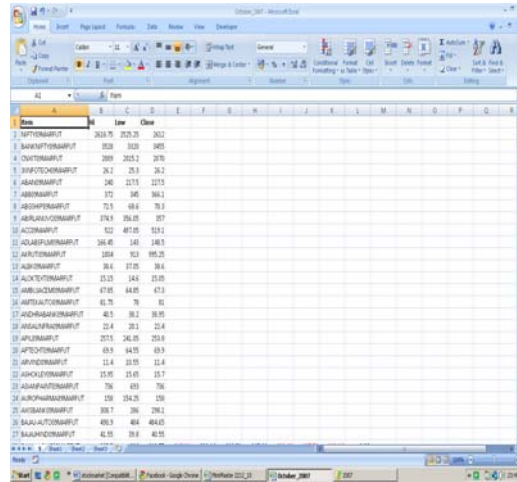


Fig. 2. Sample file showing highest lowest and close price of stocks for a particular day

#### 3.2 Data-analysis

This model starts its work once the data-acquisition process has finished. Data-analysis reports can be made and shown to the users in a number of ways. In the proposed software, reports for the weekly, monthly and yearly highest, lowest and average prices are shown to the user in an excel sheets. The user can also directly see the graphs of all these values. A sample report file is shown in Fig 3. A sample graph is also shown in Fig 4.

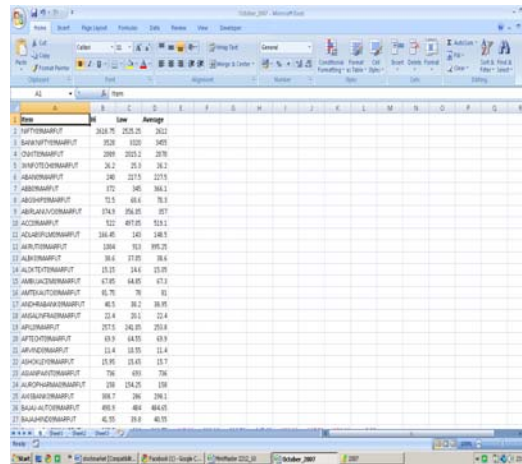


Fig. 3. Sample report file for highest, lowest and average values of stocks for a particular year.

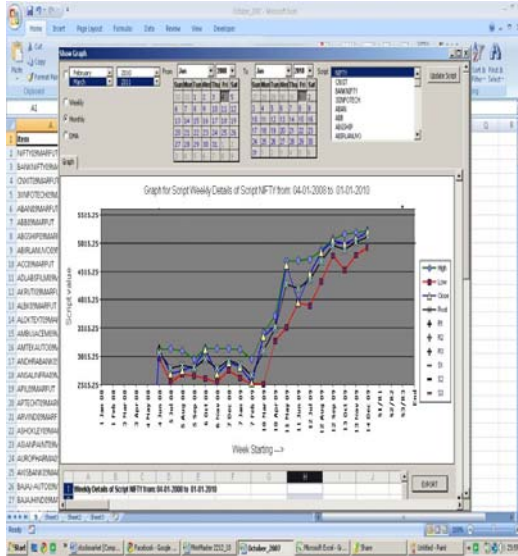


Fig 4. Sample graph for showing highest, lowest, average and average\_close price of a particular stock.

### 3.3 Prediction Model

This model works in accordance with the analysis model. The prediction model makes use of a set of formulae to estimate the future behaviour of the stocks. The inputs to these formulae are the values obtained during analysis of the particular stock. As the future behaviour of the stocks can be predicted only after analysis the past conduct of the stocks, prediction has to work hand in hand with analysis. After doing all the predictions this module generates a report as shown in Fig 5. The predictions model gives a distinct colour code to all its different types of predictions. These colour codes helps the user to identify whether the particular stock is a good stock to invest on, or whether the currently possessing stocks are predicted to abate.

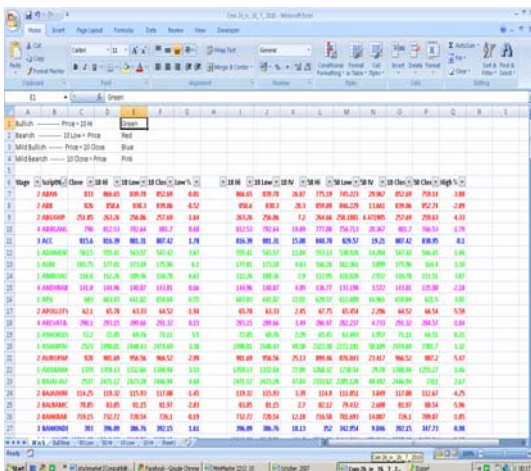


Fig. 5. Report generated by the prediction model

## 4. EXPERIMENTS & RESULTS

Data acquired from the internet, the acquired data analysed and the predictions calculated all these stages are shown in Fig 1, Fig 2, Fig 3, Fig 4 and Fig 5. This software is briefly explained in the following discussion.

Data is acquired through the internet directly from the website of BSE(Bombay Stock Exchange). These bulks of data are then sorted out and only the useful data is exported from the bulk and stored in excel sheets for the use of the software. The software imports this data into its database and starts analysing it. Once all the necessary analyses are done the prediction model takes charge. It uses the analysis results to make its predictions.

The user has the option of analysing the data according to his needs. There are many different analysis models included in the software. The snapshots of the software at different stages of working are shown below.

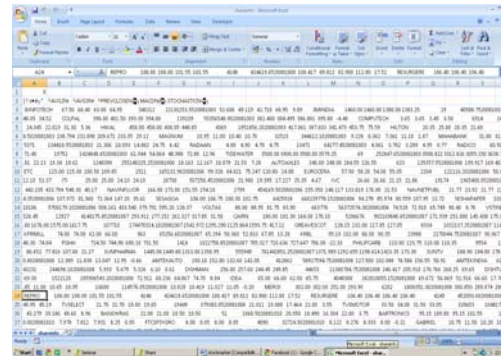


Fig. 6. Raw data acquired from the internet.

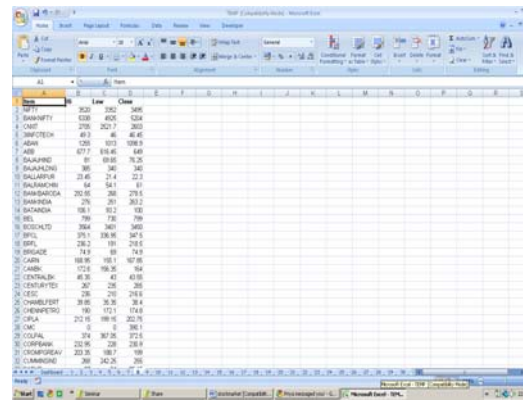


Fig. 7. Filtered data.



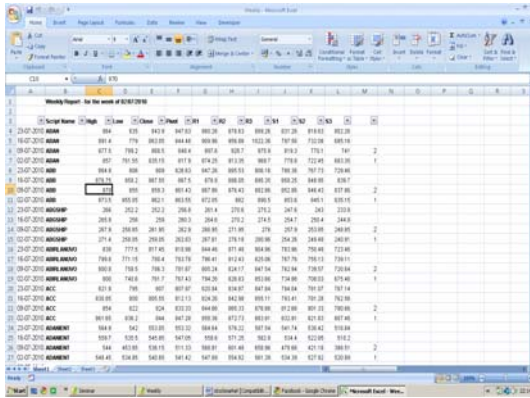


Fig. 8. Analysis Report – 1

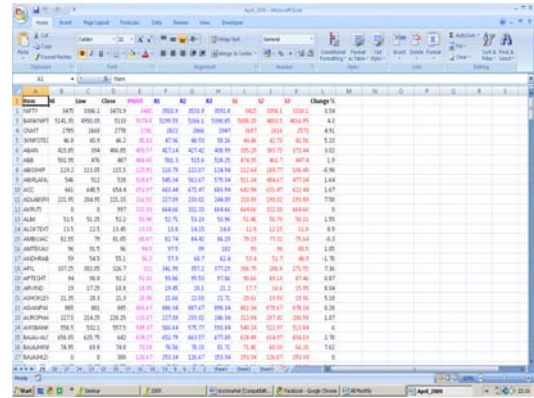


Fig. 11. Prediction Report – 1

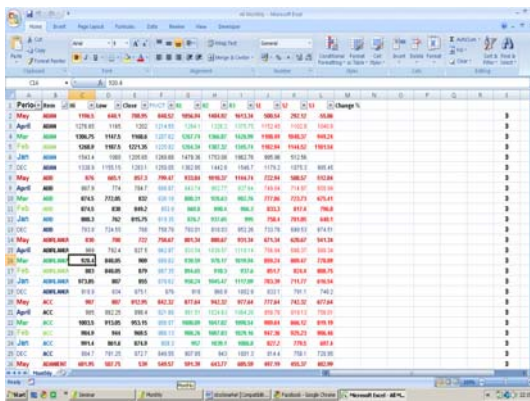


Fig. 9. Analysis Report - 2

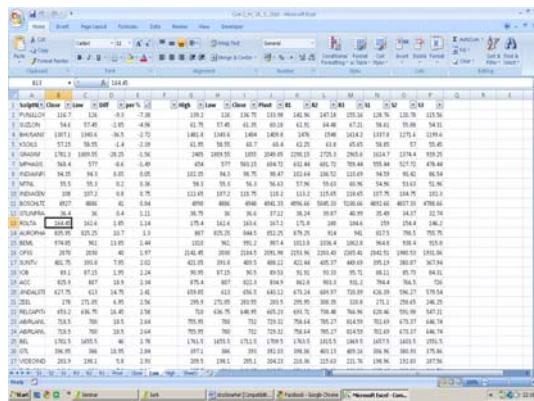


Fig. 12. Prediction Report – 2

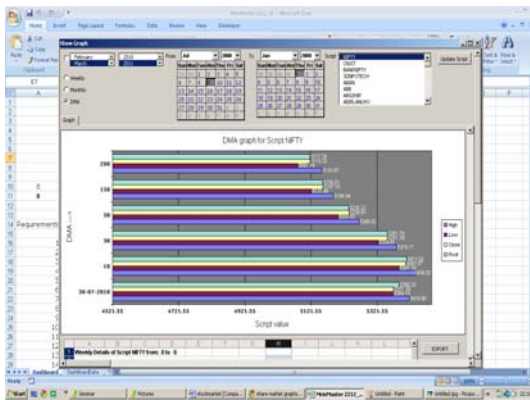


Fig. 10. Analysis graph.

Now let us see the comparison of the share market graphs with this software.



Fig. 13. Analysis graph - 1

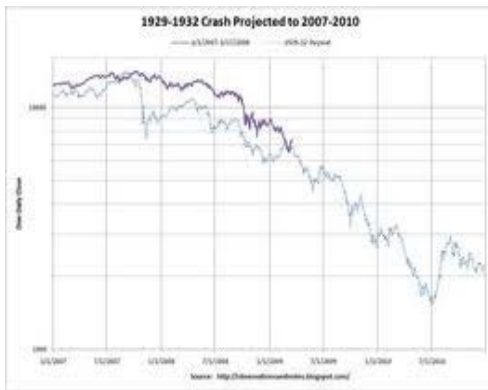


Fig. 14. Analysis graph - 1



Fig. 15. Prediction graph - 1

As it is seen, Fig 13 and Fig 14, it shows graphs for analysis of a particular stock. But these graphs only take into consideration the past behaviour of the stocks. They show no predictions. Whereas the graph shown in fig 15 shows the past behaviour as well as predict the future of the stock.

## 5. CONCLUSION

In this software, analysis and predictions are made that should find interest from the stock market investors. The software has a huge possibility of development with addition of more and more powerful analysis models. A prediction module cannot sustain on itself without analysis, the stronger analysis will help give more accurate predictions (up to 90% efficiency). The current predictions have tested to be 80% accurate.

Compared to other share market software's, this software has an advantage as it provides a prediction module. The software is designed to sustain both analyses and predictions at the same time.

Research for more improvement in the prediction model is still in the process. Also developments in increasing the speed of operation of the software are being made.

## REFERENCES

- [1] K. Jacker and J. Mckinney, "TkDAS—A data acquisition system usingRTLinux, COMEDI, and Tcl/Tk," in Proc. Third Real-Time Linux Workshop, 2001. [Online]. Available: The Real Time Linux Foundation: <http://www.realtimelinuxfoundation.org/events/rtlws-2001/papers.html>
- [2] A. Sang, H. Lin, and C. E. Y. Z. Goua, "Wireless Internet-based measurement architecture for air quality monitoring," in Proc. 21st IEEE IMTC, May 18–20, 2004, vol. 3, pp. 1901–1906.
- [3] W. Kattaneq, A. Schreiber, and M. Götze, "A flexible and cost-effective open system platform for smart wireless communication devices," in Proc. ISCE, 2002.
- [4] J. E. Marca, C. R. Rindt, and M. G. McNally, "The tracer data collection system: Implementation and operational experience," Inst. Transp. Studies, Univ. California, Irvine, CA, Uci-Its-As-Wp-02-2, 2002.
- [5] M. A. Al-Taei, O. B. Khader, and N. A. Al-Saber, "Remote monitoring of vehicle diagnostics and location using a smart box with Global Positioning System and General Packet Radio Service," in Proc. IEEE/ACS AICCSA, May 13–16, 2007, pp. 385–388.
- [6] JStock - Stock Market Software 1.02, [http://www.topshareware.com/download.aspx?id=67171&p=&url=http%3a%2f%2fdownloads.sourceforge.net%2fjstock%2fjstock-1.0.2-setup.exe%3fbig\\_mirror%3d0](http://www.topshareware.com/download.aspx?id=67171&p=&url=http%3a%2f%2fdownloads.sourceforge.net%2fjstock%2fjstock-1.0.2-setup.exe%3fbig_mirror%3d0)
- [7] stock price analysis 1, <http://www.topshareware.com/download.aspx?id=77845&p=&url=http%3a%2f%2fwww.stockpriceanalysis.com%2fspa.exe>



# Fast Handoff Implementation by using Curve Fitting Equation With Help of GPS

Debabrata Sarddar<sup>1</sup>, Shubhajeet chatterjee<sup>2</sup>, Ramesh Jana<sup>1</sup>, Shaik Sahil Babu<sup>1</sup>, Hari Narayan Khan, Utpal Biswas<sup>3</sup>, M.K. Naskar<sup>1</sup>.

1. Department of Electronics and Telecommunication Engg, Jadavpur University, Kolkata – 700032.

2. Department of Electronics and Communication Engg, Institute of Engg. & Managment college, saltlake, Kolkata-700091.

3. Department of Computer Science and Engg, University of Kalyani, Nadia, West Bengal, Pin- 741235.

## Abstract

Due to rapid growth in IEEE 802.11 based Wireless Local Area Networks (WLAN), handoff has become a burning issue. A mobile station (MS) requires handoff when it travels out of the coverage area of its current access point (AP) and tries to associate with another AP. But handoff delays provide a serious barrier for such services to be made available to mobile platforms. Throughout the last few years there has been plenty of research aimed towards reducing the handoff delay incurred in the various levels of wireless communication. In this paper we propose a method using the GPS(Global Positioning System) to determine the positions of the MS at different instants of time and then by fitting a trend equation to the motion of the MS to determine the potential AP(s) where the MS has maximum probability of travelling in the future. This will result in a reduction of number of APs to be scanned as well as handoff latency will be reduced to a great extent.

**Keywords:** IEEE 802.11, Handoff latency, GPS (Global Positioning System), Regression, Neighbor APs.

## 1. Introduction

IEEE 802.11 based wireless local area network (WLAN) are widely used in domestic and official purpose due to its flexibility of wireless access. However, WLANs are restricted in their diameters to campus, buildings or even a single room. Due to the limited coverage areas of different APs a MS has to experience handoff from one AP to another frequently.

### 1.1 Handoff

When a MS moves out of reach of its current AP it must be reconnected to a new AP to continue its operation. The search for a new AP and subsequent registration under it constitute the handoff process which takes enough time (called handoff latency) to interfere with proper functioning of many applications.

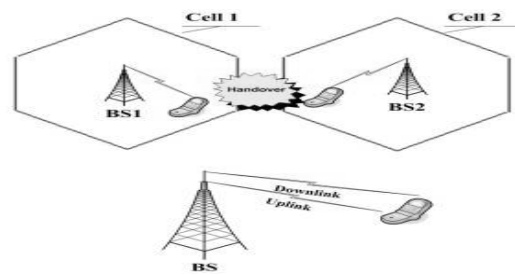


Figure 1. Handoff process

For successful implementation of seamless Voice over IP communications, the handoff latency should not exceed 50ms. It has been observed that in practical situations handoff takes approximately 200-300 ms to which scanning delay contributes almost 90%. This is not acceptable and thus the handoff latency should be minimized.

Three strategies have been proposed to detect the need for hand off[1]:

1) *mobile-controlled-handoff (MCHO)*: The mobile station (MS) continuously monitors the signals of the surrounding base stations (BS) and initiates the hand off process when some handoff criteria are met.

2) *network-controlled-handoff (NCHO)*: The surrounding BSs measure the signal from the MS and the network initiates the handoff process when some handoff criteria are met.

3) *mobile-assisted-handoff (MAHO)*: The network asks the MS to measure the signal from the surrounding BSs. The network makes the handoff decision based on reports from the MS.

Handoff can be of many types:

*Hard Handoff*: In this process radio link with old AP is broken before connection with new AP. This in turn results

in prolonged handoff latency which is known as link switching delay.

*Soft Handoff*: This mechanism is employed nowadays. Here connection with old AP is maintained until radio link with new AP is established. This results in reduced handoff time in comparison to hard handoff as shown in figure 2.

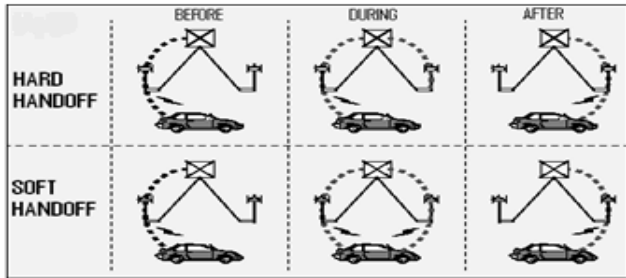


Figure 2. Hard & soft handoff

In NGWS(next generation wireless system),two types of handoff scenarios arise: horizontal handoff, vertical handoff[2][3].

- *Horizontal Handoff*: When the handoff occurs between two BSs of the same system it is termed as horizontal handoff. It can be further classified into two:
  - *Link layer handoff* : Horizontal handoff between two BSs that are under the same foreign agent(FA).
  - *Intra system handoff* : Horizontal handoff between two BSs that belong to two different FAs and both FAs belong to the same gateway foreign agent (GFA) and hence to the same system.
- *Vertical Handoff* : When the handoff occurs between two BSs that belong to two different GFAs and hence to two different systems it is termed as vertical handoff as shown in figure 3.

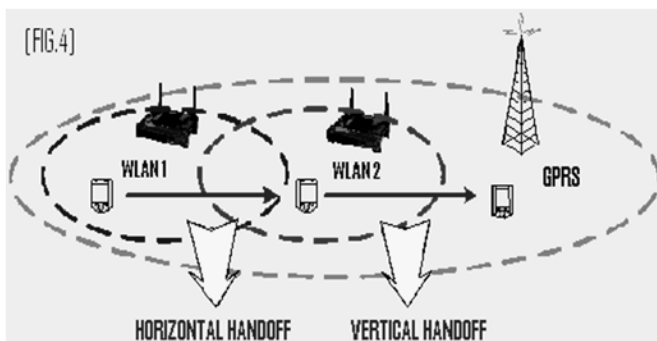


Figure 3. Horizontal & vertical handoff

## 1.2 Handoff Mechanism

The handoff process is composed of the following three stages:

*Scanning*: The scanning process constitutes the bulk (almost 90%) of the total handoff time [4]. As the MS starts moving away from the AP the *Signal-to-Noise-Ratio* (SNR) starts decreasing and this phenomenon triggers the initiation of handoff. The MS has to establish a radio link with a potential AP before the connectivity with the current AP is detached. This is accomplished by means of a MAC(Medium Access Control) layer function called *scanning*.

*Authentication*: After scanning, The MS sends authentication frames to inform the AP (selected by the scanning process) of its identity. The AP then responds by sending an authentication response frame indicating approval or rejection

*Re-association*: It is the process by which association transfer takes place from one AP to another. This process follows the authentication process depending on the authentication response sent by the AP.

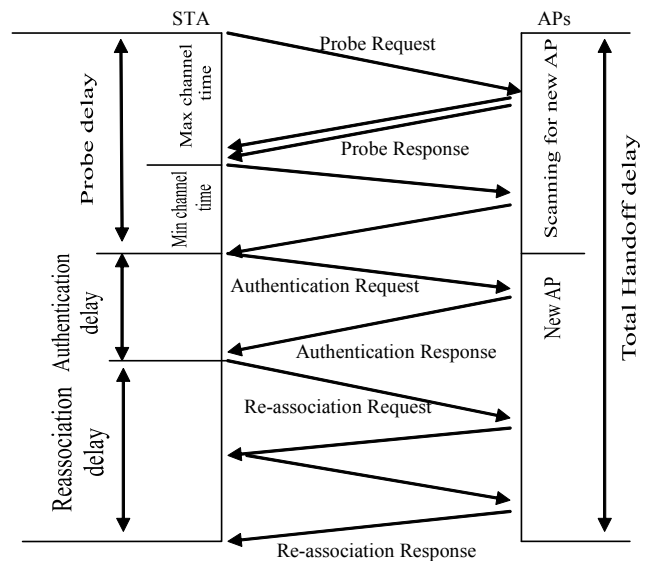


Figure 4. The handoff process

## 1.1 Global Positioning System

The *Global Positioning System* (GPS) is a space based global navigation satellite system which is used in map making, land surveying, navigation, geocaching and in

other fields. A GPS receiver is able to calculate its position by precisely timing the signals sent by the GPS satellites. The receiver uses the received messages from the satellites to determine the transit time of each message and calculates the distance to each satellite. These distances are then utilized to compute the position of the receiver. For normal operation a minimum of four satellites are necessary. Using the messages received from the satellites the receiver is able to calculate the times sent and the satellite positions corresponding to these points. Each MS is equipped with a GPS receiver which is used to determine the positions of the MS at different instants of time. This will provide knowledge about the MS's movement within 1 to 2 meter precision.

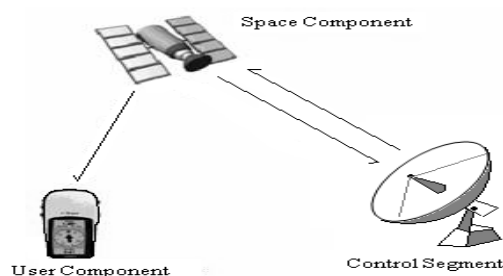


Figure5. Components of GPS

In section II we take you through the various works that have already been done to achieve successful handoff and in section III we introduce a new method using the statistical regression over the movement of MS by which we intend to reduce the handoff delay to the range of a few milliseconds. This is followed by performance evaluation of our proposed technique using simulations in section IV after which in section V we propose a few areas in which further improvement can be made.

## 2. Related works

A number of different schemes have been proposed to reduce handoff latency in IEEE 802.11 wireless LANs. Authors of [8] aimed at reducing the authentication process which contributes very little to the handoff time.

In [5] authors present a useful method using a neighbor graph and a non overlap graph. This concept was used to reduce total number of channels to be scanned and the waiting time on each channel. However the complexity of implementation of the algorithm was a major setback. In [6] a channel mask scheme was introduced where a selective scanning algorithm was proposed along with a caching mechanism. In [7] authors propose selective scanning algorithm using neighbor graphs. This method requires changes in the network infrastructure and use of

IAPP. Moreover, these processes involve channel scanning of all neighboring APs and do not consider the position or velocity of MS to select potential APs. Hence these methods are more power consuming and are less effective for reducing handoff.

## 3. Proposed Works

Here we propose a method depending upon statistical regression to minimize the handoff delay. We will select the potential APs where the MS has maximum probability of travelling when it moves out of the coverage area of its present AP. Thus we will minimize handoff delay by Scanning only the potential APs for available channels. . We implement our method with the help of GPS. We present the method in the following four sections:

### 3.1 Definition of Parameters

In an idealized model we approximate the overlapping circular cell areas by hexagonal cells that cover the entire service region with the AP being located at the centre of the hexagon. For the sake of simplicity we consider that a particular hexagonal cell is surrounded by six similar cells (7 cell cluster). Considering the entire cell area as a two dimensional plane, we define two mutually perpendicular coordinate axes namely the X and Y axes with the AP as the origin. Now let us consider the motion of a MS in a particular cell. The position namely the X and Y coordinates of a MS can be obtained via a GPS.

As shown in the figure.6 we divide the cell into two regions:

- (a) REGION 1(white region): This denotes the core region where the GPS is used to monitor the position of the MS.
- (b) REGION 2(grey region): This denotes the region where the signal strength received by the MS falls below a threshold value and the MS starts the scanning process to initiate handoff.

### 3.2 Handoff Initiation

As long as the MS is travelling in region 1 no handoff is required. Here we note the  $(x, y)$  coordinates of the MS via the GPS. Let the initial position(at time  $t=t_0$  sec) of the MS be denoted by  $(x_0, y_0)$  with respect to the origin. This process is repeated after a fixed time interval of T sec. When the MS leaves region 1 and enters region 2 the scanning of MS's position by the GPS is stopped. Handoff process has to be initiated. Here we propose a statistical method of regression to determine the potential APs where the MS has maximum chance of moving. During handoff it will scan only the channels of those APs.

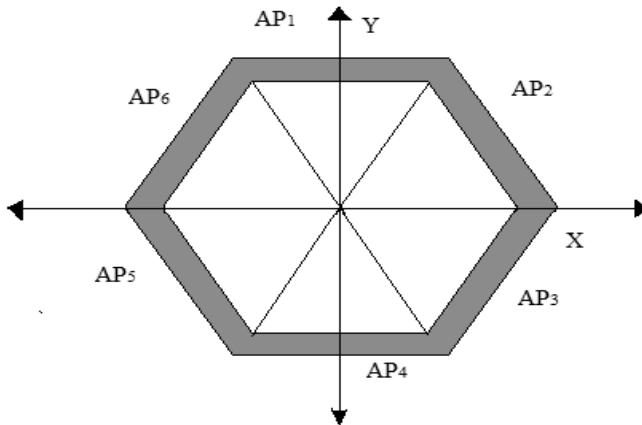


Figure 6. The hexagonal cell

Table 1. Angle division

UPPER LIMIT OF $\theta$	LOWER LIMIT OF $\theta$	AP TO BE SCANNED(VIDE FIG.6)
$0^\circ$	$60^\circ$	AP2
$60^\circ$	$120^\circ$	AP1
$120^\circ$	$180^\circ$	AP6
$180^\circ$	$240^\circ$	AP5
$240^\circ$	$300^\circ$	AP4
$300^\circ$	$360^\circ$	AP3

### 3.3 Selection of Potential AP by Regression Method

We employ the *method of least squares* to fit curves for the motion of the MS. Let the equation of motion of the MS along the Y-direction be denoted by  $y=a_0+a_1t$ . We are now faced with the problem of choosing the values of the variables  $a_0$  and  $a_1$ . The sum of the squares of the deviations of the data points with those obtained from the proposed curve is given by,

$$S = \sum (y_i - a_0 - a_1t_i)^2$$

The method of least squares states that S should be minimum with respect to  $a_0$  and  $a_1$ . This metric has many desirable characteristics:

*Errors of opposite sign are not cancelled.*

*It weighs large errors more than small errors.*

To minimize S we take the partial derivative of S with respect to  $a_0$  and  $a_1$  and set these to zero. Thus,

$$\partial S / \partial a_0 = \sum 2(y_i - a_0 - a_1t_i)(-1) = 0 \dots\dots\dots(1)$$

$$\partial S / \partial a_1 = \sum 2(y_i - a_0 - a_1t_i)(-t_i) = 0 \dots\dots\dots(2)$$

Solving the above equations we obtain estimates for  $a_0$  and  $a_1$ .

$$a_0 = [\sum y_i \sum t_i^2 - \sum t_i \sum y_i t_i] / [n \sum t_i^2 - (\sum t_i)^2] \dots\dots\dots(3)$$

$$a_1 = [n \sum y_i t_i - \sum y_i \sum t_i] / [n \sum t_i^2 - (\sum t_i)^2] \dots\dots\dots(4)$$

Similarly by the least squares method it is possible to fit a regression line for the equation of motion of the MS along the X direction. Let the regression equation along y axes be denoted by  $x=b_0 + b_1t$  where  $b_0$  and  $b_1$  are estimated by the above mentioned method.

Now, with the regression equations it will be possible to predict the position of the MS outside the present cell. Let the MS enters region 2 of the present cell at time  $t=t'$  sec. Before the handoff process starts the regression lines are computed. We evaluate the probabilistic (x, y) coordinates of the MS at time  $t=(t' + T')$  sec by putting  $t=(t' + T')$  sec in the two regression equations for the x and y coordinates. If the predicted MS's position still falls within the present AP then the probabilistic position is calculated at time  $t=(t' + 2T')$  sec and the process continues. Here  $T'$  is a fixed time interval after which probabilistic positions are computed and should be appropriately chosen depending on the cell size. The first position of the MS which falls outside the current cell area as indicated by the regression equations is denoted by  $(x', y')$ . The AP within which this point falls can be evaluated by knowing the angle this point makes with the coordinate axes.

It is to be noted that the time consumed to predict the MS's position outside the present cell area is quite small and can be neglected for practical purposes. The angle  $\theta$  that the point  $(x', y')$  makes with the X axis is given by  $\theta = \tan^{-1}(y'/x')$ . The value of  $\theta$  can be used to determine the potential AP which has to be scanned.

### 3.4 Error Estimation

Although we have opted for the best fit yet there will be some amount of error, however small, which has to be taken into consideration. For the error estimation we propose the following method.

Let us first concentrate on the regression line of x on t. Let  $\alpha$  denote the maximum magnitude of the deviation of the data points from the predicted values obtained from the equation  $x=a_0 + a_1t$ . Hence

$$\alpha = \max (|x_i - a_0 - a_1t_i|) \text{ for } i=0, 1, 2 \dots\dots\dots n$$

Similarly let  $\beta$  denote the maximum magnitude of the deviation of the data points from the predicted values obtained from the equation  $y=b_0 + b_1t$ . Hence

$$\beta = \max (|y_i - b_0 - b_1t_i|) \text{ for } i=0, 1, 2 \dots\dots\dots n$$

Thus we may assume the maximum variation that can take place in the y coordinate of a predicted value obtained from the equation  $y=a_0 + a_1t$  to be  $\alpha$ . Thus the y coordinate of the MS may vary between  $y'-\alpha$  and  $y'+\alpha$ . Similarly this variation takes the value of  $\beta$  for motion along x axis. Thus

considering error measures the MS has the probability of being located in the rectangle as shown in the following figure with maximum probability being located at the centre  $(x', y')$ . From the above discussion it is clear that the coordinates of the vertices of the rectangle ABCD are A  $(x' - \alpha, y' + \beta)$ , B  $(x' + \alpha, y' + \beta)$ , C  $(x' + \alpha, y' - \beta)$  and D  $(x' - \alpha, y' - \beta)$ . However, our concern is variation of  $\theta$ . Clearly considering such error measures minimum value of  $\theta$  will result when the MS is located at C and maximum value will result when the MS is located at vertex A.

Thus  $\theta_{\min} = \tan^{-1}(y' - \beta / x' + \alpha)$   
 and  $\theta_{\max} = \tan^{-1}(y' + \beta / x' - \alpha)$ .

Thus we effectively have a range of values of  $\theta$  with the most probable value being  $\tan^{-1}(y'/x')$ . For good fitting the values of  $\alpha$  and  $\beta$  will be quite small in comparison to  $x'$  and  $y'$  and hence in most cases the range of values of  $\theta$  will be small enough to yield 1 AP or at the most 2 APs for scanning purposes.

### 3.5 Scanning and Pre-authentication

All necessary information like MAC(Medium Access Control) addresses and operating channels of the neighbor APs are downloaded by the MS from the server data. Selective channel scanning with the help of unicast instead of broadcast efficiently reduces the handoff delay to a great extent. Moreover, the MS has to wait for only the 'round trip time' (rtt) for scanning each channel instead of the Min Channel Time or the Max Channel Time. When the MS responds to handoff, according to the proposed algorithm, it first looks for the potential AP and then scans the channels of that AP. As proposed in [9], the expected scanning delay using selective scanning is,

$$t = N \times \Gamma + \omega$$

where 't' is the scanning delay, 'N' is the number of channels scanned, ' $\Gamma$ ' is the round trip time and ' $\omega$ ' is the message processing time. ' $\Gamma$ ' is the summation of the time taken for the Probe Request to be sent to the selected AP's and for the Probe Response to be received, which has been estimated to be around 3-7 ms. By pre-authentication during the scanning phase the factor  $\omega$  would also be greatly reduced and would consist only of the re-association time. This mechanism can be implemented as proposed in [10].

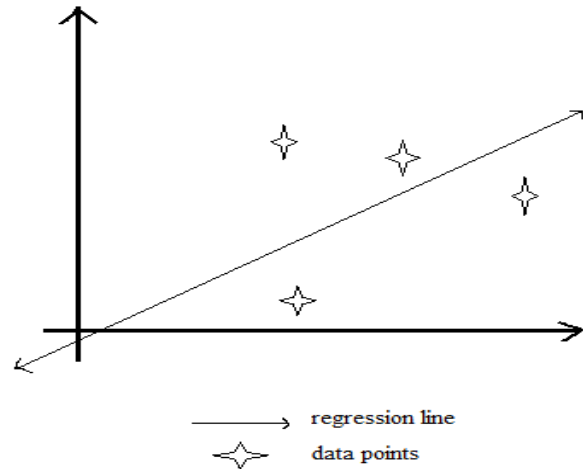


Figure.7A. Illustration of the error estimation

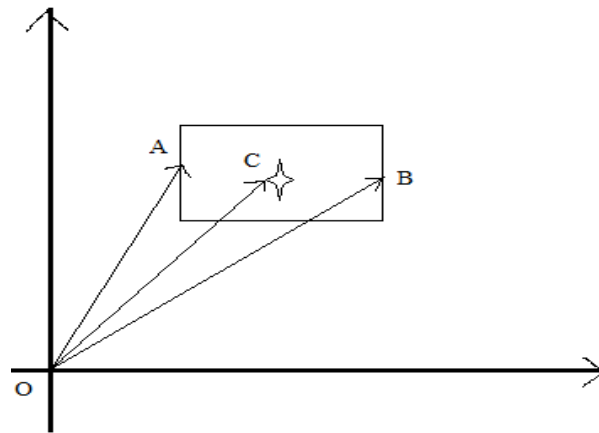


Figure.7B Illustration of the error estimation

## 4. Simulation Results

Simulations of a sample run of our experiment have been presented here. We consider the handover for a MS from the cell in which its call originates. The coverage region of the AP is taken as regular hexagon of length 100m approx. The handoff region starts at a radial distance of 90m from the AP. The mobility pattern of the MS has been presented in Fig.8 which was tracked by GPS at an interval of 5 s.



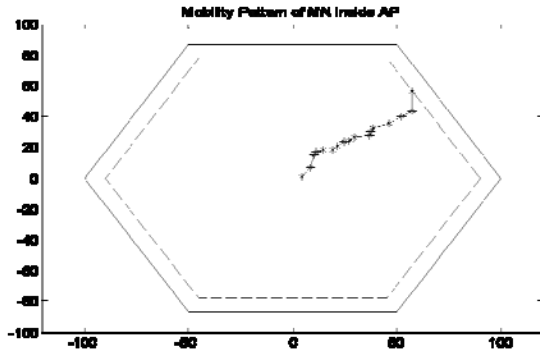


Figure.8

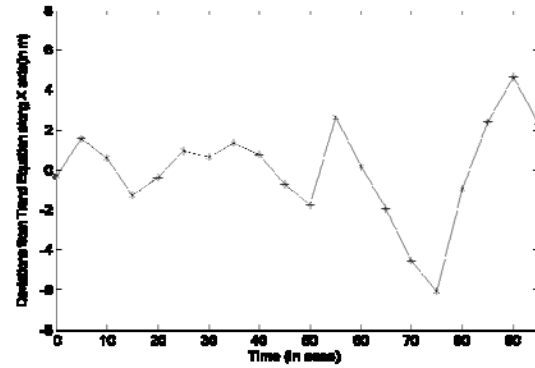


Figure.11

On entering the handoff region GPS was turned off. Regression lines for X and Y axes were computed. The equations in this case turned out to be  $x = 4.583159 + 0.532045 t$  and  $y = 7.017031 + 0.399405 t$

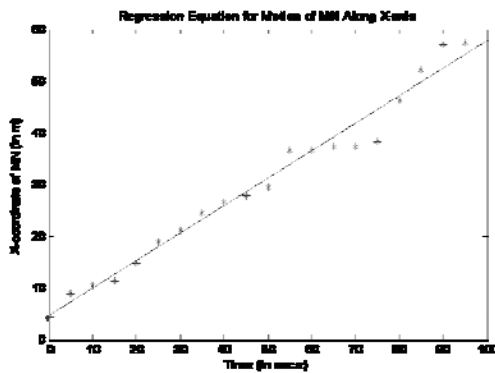


Figure.9

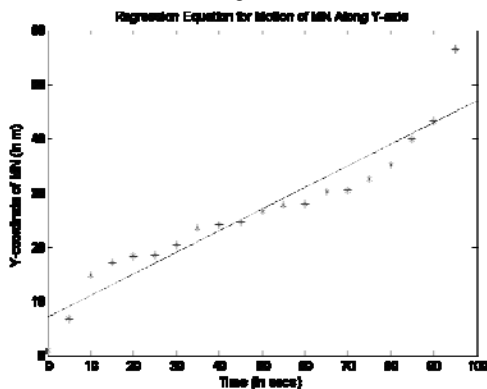


Figure 10

Hence the parameters computed were  $x' = 71.089$  m,  $y' = 56.943$  m and  $\theta = 38.695$  (degree)

The deviations of predicted from actual positions of MS obtained by GPS were plotted for both axis.

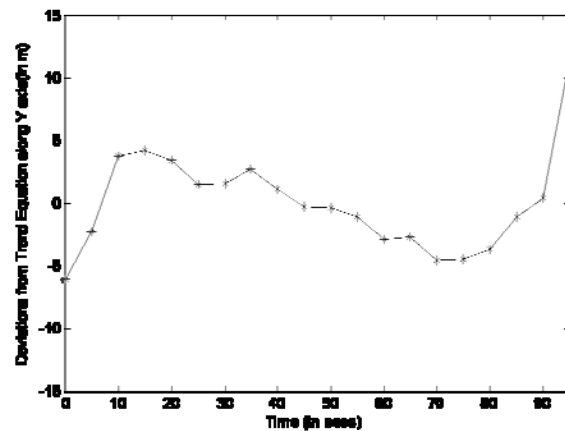


Figure.12

Hence,  $\alpha = 6.086$  m

and  $\beta = 11.579$  m

Thus,  $\theta_{\min} = \tan^{-1}(y' - \beta/x' + \alpha) = 30.447$ (degree)

and  $\theta_{\max} = \tan^{-1}(y' + \beta/x' - \alpha) = 46.510$ (degree)

This indicates that the MS is moving towards AP2 as the expected range of angle lies between 0 and 60 degrees.

We made 100 such sample runs by varying various parameters like mobility range and velocity of MS, cell coverage area, etc. In 89% of the cases one potential AP was selected while in 9% cases two potential APs were selected. The remaining 2% constituted cases where potential APs selected by the proposed algorithm resulted in association failure leading to an efficient full scanning of the channels of other APs (approx 30-40 ms). Taking the 'round trip time' (rtt) as 3 msec the average handoff latency measured was 6.563 msec which is a drastic improvement in comparison to earlier proposed methods. The graph of this simulation is plotted in Fig.12, which shows the various handoff delay times in the Y-axis in msec, for each experiment, which is shown in the X-axis.

The success of our simulation clearly depicts the applicability of our proposed algorithm.

## 5. Conclusion

Our proposed method aims at reducing handoff time by reducing the number of APs to be scanned which is accomplished by fitting a trend equation to the motion of the MS. This in turn reduces the number of channels to be scanned which brilliantly reduces the handoff delay as is clear from the simulation presented in the above section. In the proposed algorithm a linear trend equation has been fitted because it is the most common and trustworthy fit. However higher order polynomials may also be used for fitting and best fit may be chosen by comparing the norm of the residuals.

However the proposed algorithm may prove erroneous if the motion of the MS is too much random to be used for prediction purposes. Future works in this field may include research on more refined algorithms regarding curve fitting and prediction. Error estimation method may also be improved. It is worth mentioning here that although the proposed work has been presented considering honeycomb structures yet our algorithm would work in a similar manner for other cell structures and neighbor AP locations. Minor changes would be introduced depending on the network topology.

## References

- [1] Yi-Bing Lin Imrich Chalmatc, "Wireless and Mobile Network Architectures," pp. 17.
- [2] AKYILDIZ, I. F., XIE, J., and MOHANTY, S., "A survey on mobility management in next generation all-IP based wireless systems," *IEEE Wireless Communications*, vol. 11, no. 4, pp. 16-28, 2004.
- [3] STEMME, M. and KATZ, R. H., "Vertical handoffs in wireless overlay networks," *ACM/Springer Journal of Mobile Networks and Applications(MONET)*, vol. 3, no. 4, pp. 335-350, 1998.
- [4] J. Pesola and S. Pokanen, "Location-aided Handover in Heterogeneous Wireless Networks," in Proceedings of Mobile Location Workshop, May 2003.
- [5] M. Shin, A. Mishra, and W. Arbaugh, "Improving the Latency of 802.11 Hand-offs using Neighbor Graphs," in Proc. ACM MobiSys 2004, pp 70-83, June 2004.
- [6] S. Shin, A. Forte, A. Rawat, and H. Schulzrinne, "Reducing MAC Layer Handoff Latency in IEEE 802.11 Wireless LANs," in Proc. ACM MobiWac 2004, pp 19-26, October 2004.
- [7] H.-S. K. et. al. "Selecive channel scanning for fast handoff in wireless LAN using neighbor graph", International Technical Conference on Circuits/Systems, Computer and Communications. LNCS Springer, Vol 3260, pp 194-203, 2004.
- [8] S. Park and Y. Choi. "Fast inter-ap handoff using predictive-authentication scheme in a public wireless lan. Networks2002 (Joint ICN 2002 and ICWLHN 2002), August 2002.

- [9] Chien-Chao Tseng, K-H Chi, M-D Hsieh and H-H Chang," Location-based Fast Handoff for 802.11 Networks", IEEE Communication Letters, Vol-9, No 4, pp 304-306, April 2005.
- [10] Yogesh Ashok Powar and Varsha Apte, "Improving the IEEE 802.11 MAC Layer Handoff Latency to Support Multimedia Traffic", Wireless Communications and Networking Conference, 2009. WCNC 2009. IEEE Xplore, pp 1-6, April 2009



**Debabrata Sarddar** is currently pursuing his PhD at Jadavpur University. He completed his M.Tech in Computer Science & Engineering from DAVV, Indore in 2006, and his B.Tech in Computer Science & Engineering from Regional Engineering College, Durgapur in 2001. His research interest includes wireless and mobile system.



**Shubhajeet Chatterjee** is presently pursuing B.Tech Degree in Electronics and Communication Engg. at Institute of Engg. & Managment College, under West Bengal University Technology. His research interest includes wireless sensor networks and wireless communication systems.



**Ramesh Jana** is presently pursuing M.Tech (2nd year) in Electronics and Telecommunication Engg. at Jadavpur University. His research interest includes wireless sensor networks, fuzzy logic and wireless communication systems



**Hari Narayan Khan** is presently pursuing M.Tech (Final year) in Computer Technology at Jadavpur University. He completed his B.Tech in Electronics & Communication Engineering in 2006 from Institute of Technology & Marine Engineering under West Bengal University of Technology. His research interest includes wireless and mobile system.

### SHAIK SAHIL BABU



is pursuing Ph.D in the Department of Electronics and Telecommunication Engineering under the supervision of Prof. M.K. NASKAR at Jadavpur University, KOLKATA. He did his Bachelor of Engineering in Electronics and Telecommunication Engineering from Muffa Kham Jah College of Engineering and Technology, Osmania University, Hyderabad, and Master of Engineering in Computer Science and Engineering from Thapar Institute of Engineering and Technology, Patiala, in Collaboration with National Institute of Technical Teachers' Training and Research, Chandigarh.

**Utpal Biswas** received his B.E, M.E and PhD degrees in Computer Science and Engineering from Jadavpur University, India in 1993, 2001 and 2008 respectively. He served as a faculty member in NIT, Durgapur, India in the department of Computer Science and Engineering from 1994 to 2001. Currently, he is working as an associate professor in the department of Computer Science and Engineering, University of Kalyani, West Bengal, India. He is a co-author of about 35

research articles in different journals, book chapters and conferences. His research interests include optical communication, ad-hoc and mobile communication, semantic web services, E-governance etc.



**Mrinal Kanti Naskar** received his B.Tech. (Hons) and M.Tech degrees from E&ECE Department, IIT Kharagpur, India in 1987 and 1989 respectively and Ph.D. from Jadavpur University, India in 2006.. He served as a faculty member in NIT, Jamshedpur and NIT, Durgapur during 1991-1996 and 1996-1999 respectively. Currently, he is a professor in the Department of Electronics and Tele-Communication Engineering, Jadavpur University, Kolkata, India where he is in charge of the Advanced Digital and Embedded Systems Lab. His research interests include ad-hoc networks, optical networks, wireless sensor networks, wireless and mobile networks and embedded systems. He is an author/co-author of the several published/accepted articles in WDM optical networking field that include "Adaptive Dynamic Wavelength Routing for WDM Optical Networks" [WOCN,2006], "A Heuristic Solution to SADM minimization for Static Traffic Grooming in WDM uni-directional Ring Networks" [Photonic Network Communication, 2006], "Genetic Evolutionary Approach for Static Traffic Grooming to SONET over WDM Optical Networks" [Computer Communication, Elsevier, 2007], and "Genetic Evolutionary Algorithm for Optimal Allocation of Wavelength Converters in WDM Optical Networks" [Photonic Network Communications,2008].

# Visual Cryptography Scheme for Color Image Using Random Number with Enveloping by Digital Watermarking

Shyamalendu Kandar<sup>1</sup>, Arnab Maiti<sup>2</sup>, Bibhas Chandra Dhara<sup>3</sup>

<sup>1,2</sup> Computer Sc. & Engineering  
Haldia Institute of Technology  
Haldia, West Bengal, India

<sup>3</sup>Department of Information Technology  
Jadavpur University  
Kolkata, West Bengal, India

## Abstract

Visual Cryptography is a special type of encryption technique to obscure image-based secret information which can be decrypted by Human Visual System (HVS). This cryptographic system encrypts the secret image by dividing it into  $n$  number of shares and decryption is done by superimposing a certain number of shares ( $k$ ) or more. Simple visual cryptography is insecure because of the decryption process done by human visual system. The secret information can be retrieved by anyone if the person gets at least  $k$  number of shares. Watermarking is a technique to put a signature of the owner within the creation.

In this current work we have proposed Visual Cryptographic Scheme for color images where the divided shares are enveloped in other images using invisible digital watermarking. The shares are generated using Random Number.

**Keywords:** Visual Cryptography, Digital Watermarking, Random Number.

## 1. Introduction

Visual cryptography is a cryptographic technique where visual information (Image, text, etc) gets encrypted in such a way that the decryption can be performed by the human visual system without aid of computers [1].

Like other multimedia components, image is sensed by human. Pixel is the smallest unit constructing a digital image. Each pixel of a 32 bit digital color image are divided into four parts, namely Alpha, Red, Green and Blue; each with 8 bits. Alpha part represents degree of transparency.

A 32 bit sample pixel is represented in the following figure [2] [3].

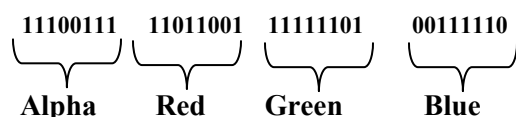


Fig 1: Structure of a 32 bit pixel

Human visual system acts as an OR function. Two transparent objects stacked together, produce transparent object. But changing any of them to non-transparent, final objects will be seen non-transparent. In  $k$ - $n$  secret sharing visual cryptography scheme an image is divided into  $n$  number of shares such that minimum  $k$  number of shares is sufficient to reconstruct the image. The division is done by Random Number generator [4].

This type of visual cryptography technique is insecure as the reconstruction is done by simple OR operation.

To add more security to this scheme we have proposed a technique called digital enveloping. This is nothing but an extended invisible digital watermarking technique. Using this technique, the divided shares produced by  $k$ - $n$  secret sharing visual cryptography are embedded into the envelope images by LSB replacement [5]. The color change of the envelope images are not sensed by human eye [6]. (More than 16.7 million i.e.  $2^{24}$  different colors are produced by RGB color model. But human eye can discriminate only a few of them.). This technique is known as invisible digital watermarking as human eye can not identify the change in the envelope image and the enveloped (Produced after LSB replacement) image [7].

In the decryption process  $k$  number of embedded envelope images are taken and LSB are retrieved from each of them followed by OR operation to generated the original image.

In this paper Section 2 describes the Overall process of Operation, Section 3 describes the process of  $k$ - $n$  secret sharing Visual Cryptography scheme on the image, Section 4 describes the enveloping process using invisible digital watermarking, Section 5 describes decryption process, Section 6 describes the experimental result, and Section 7 draws the conclusion.

## 2. Overall Process

**Step I:** The source image is divided into  $n$  number of shares using  $k$ - $n$  secret sharing visual cryptography scheme such that  $k$  number of shares is sufficient to reconstruct the encrypted image.

**Step II:** Each of the  $n$  shares generated in Step I is embedded into  $n$  number of different envelope images using LSB replacement.

**Step III:**  $k$  number of enveloped images generated in Step II are taken and LSB retrieving with OR operation, the original image is produced.

The process is described by Figure 2

## 3. $k$ - $n$ Secret Sharing Visual Cryptography Scheme

An image is taken as input. The number of shares the image would be divided ( $n$ ) and number of shares to reconstruct the image ( $k$ ) is also taken as input from user. The division is done by the following algorithm.

**Step I:** Take an image IMG as input and calculate its width ( $w$ ) and height ( $h$ ).

**Step II:** Take the number of shares ( $n$ ) and minimum number of shares ( $k$ ) to be taken to reconstruct the image where  $k$  must be less than or equal to  $n$ . Calculate  $RECONS = (n-k)+1$ .

**Step III:** Create a three dimensional array  $IMG\_SHARE[n][w*h][32]$  to store the pixels of  $n$  number of shares.  $k$ - $n$  secret sharing visual cryptographic division is done by the following process.

```

for i = 0 to (w*h-1)
{
    Scan each pixel value of IMG and convert it into 32 bit
    binary string let PIX_ST.
    for j = 0 to 31
    {
        if(PIX_ST.charAt(i)=1){
            call Random_Place (n, RECONS)
        }
        for k = 0 to (RECONS-1)
        {
            Set IMG_SHARE [RAND[k]][i][j] = 1
        }
    }
}
    
```

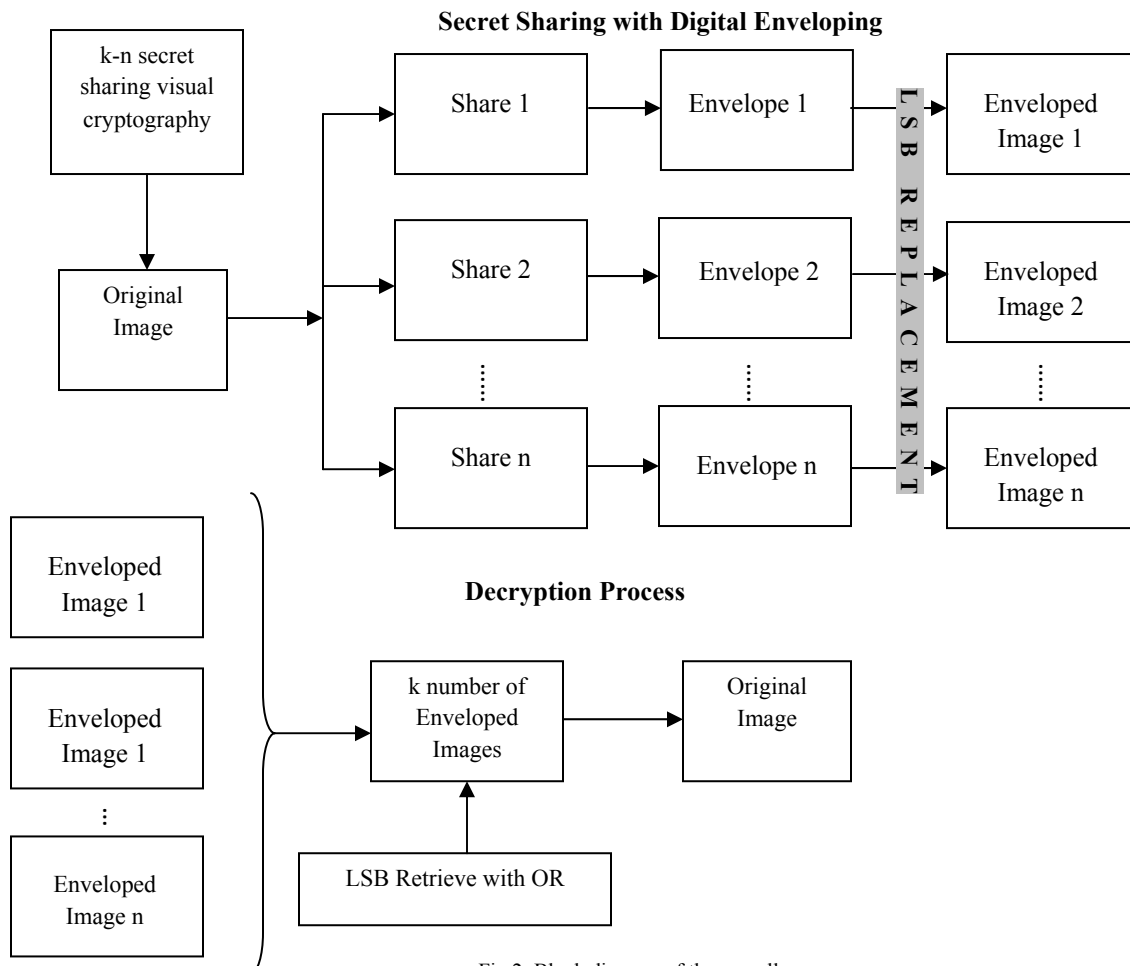


Fig 2: Block diagram of the overall process



**Step IV:** Create a one dimensional array `IMG_CONS[n]` to store constructed pixels of each `n` number of shares by the following process.

```
for k1 = 0 to (n-1)
{ for k2 = 0 to (w*h-1)
{ String value=""
for k3 = 0 to 31 {
value = value+IMG_SHARE [k1][k2][k3]
}
Construct alpha, red, green and blue part of each pixel
by taking consecutive 8 bit substring starting from 0.
Construct pixel from these part and store it into
IMG_CONS[k1] [4].
}
}
Generate image from IMG_CONS [k1]1 [8].
}
```

#### subroutine int Random\_Place(n, RECONS)

```
{ Create an array RAND[RECONS] to store the
generated random number.
for i = 0 to (recons-1)
{
Generate a random number within n, let rand_int. [9]
if (rand_int is not in RAND [RECONS])
RAND [i] = rand_int
}
}
return RAND [RECONS]
}
```

## 4. Enveloping Using Invisible Digital Watermarking

Using this step the divided shares of the original image are enveloped within other image. Least Significant Bit (LSB) replacement digital watermarking is used for this enveloping process. It is already discussed that a 32 bit digital image pixel is divided into four parts namely alpha, red, green and blue; each with 8 bits. Experiment shows that if the last two bits of each of these parts are changed; the changed color effect is not sensed by human eye[6]. This process is known as invisible digital watermarking [7]. For embedding 32 bits of a pixel of a divided share, 4 pixels of the envelope image is necessary. It means to envelope a share with resolution  $w \times h$ ; we need an envelope image with  $w \times h \times 4$  pixels. Here we have taken each envelope of size  $4w \times h$ . The following figure describes the replacement process. For replacing 8 bit alpha part, a pixel of the envelope is

needed. In the same way red, green and blue part are enveloped in three other pixels of the envelope image. The enveloping is done using the following algorithm

**Step I:** Take number of shares ( $n$ ) as input.  
for share = 0 to  $n-1$  follow Step II to Step IV.

**Step II:** Take the name of the share, let `SHARE_NO` (`NO` is from 0 to  $n-1$ ) and name of the envelope, let `ENVELOPE_NO` (`NO` is from 0 to  $n-1$ ) as input. Let the width and height of each share are  $w$  and  $h$ . The width of the envelope must be 4 times than that of `SHARE_NO`.

**Step III:** Create an array `ORG` of size  $w \times h \times 32$  to store the binary pixel values of the `SHARE_NO` using the loop for  $i = 0$  to  $(w \times h - 1)$

```
{ Scan each pixel value of the image and convert it into
32 bit
```

```
binary string let PIX
for j = 0 to 31
{ ORG [i*32+j] = PIX.charAt(j)
}
}
```

Create an array `ENV` of size  $4 \times w \times h \times 32$  to store the binary pixel values of the `ENVELOPE_NO` using the previous loop but from  $i = 0$  to  $4 \times w \times h \times 32 - 1$ .

**Step IV:** Take a marker  $M = -1$ . Using the following process the `SHARE_NO` is embedded within `ENVELOPE_NO`.

```
for i = 0 to  $4 \times w \times h - 1$ 
```

```
{
ENV [i*32+6] = ORG [i+M];
ENV [i*32+7] = ORG [i+M];
ENV [i*32+14] = ORG [i+M];
ENV [i*32+15] = ORG [i+M];
ENV [i*32+22] = ORG [i+M];
ENV [i*32+23] = ORG [i+M];
ENV [i*32+30] = ORG [i+M];
ENV [i*32+31] = ORG [i+M];
}
```

Construct alpha, red, green and blue part of each pixel by taking consecutive 8 bit substring starting from 0.

Construct pixel from these part and store it into a one dimensional array let `IMG_CONS` of size  $4 \times w \times h$  [4].

```
}
Generate image from IMG_CONS [ ]1.
```

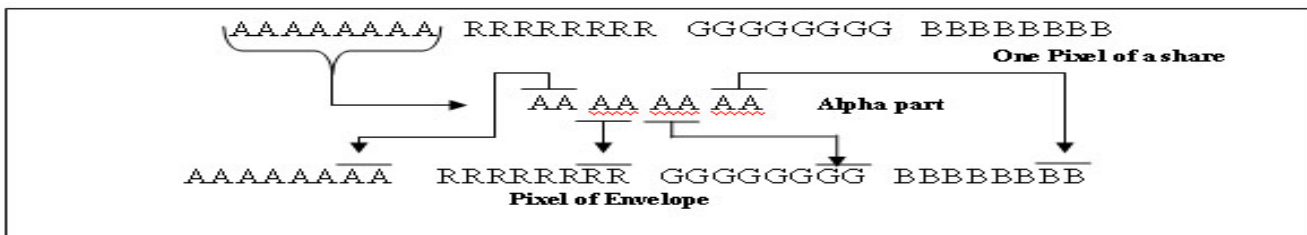


Fig 3: Enveloping Process

## 5. Decryption Process

In this step at least k numbers of enveloped images are taken as input. From each of these images for each pixel, the last two bits of alpha, red, green and blue are retrieved and OR operation is performed to generate the original image. It is already discussed that human visual system acts as an OR function. For computer generated process; OR function can be used for the case of stacking k number of enveloped images out of n.

The decryption process is performed by the following algorithm.

**Step I:** Input the number of enveloped images to be taken (k); height (h) and width (w) of each image.

**Step II:** Create a two dimensional array STORE[k][w\*h\*32] to store the pixel values of k number of enveloped images. Create a one dimensional array FINAL[(w/4)\*h\*32] to store the final pixel values of the image which will be produced by performing bitwise OR operation of the retrieved LSB of each enveloped images.

**Step III:**

```

for share_no = 0 to k-1
{
    Take the name of the enveloped image to be taken and
    store the pixel values in STORE [share_no][w*h*32]
    using the following loop.
    for i = 0 to (w*h-1)
    { Scan each pixel value of the Enveloped image and
    convert
        it into 32 bit binary string let PIX.
        for j = 0 to 31
        { STORE[share_no][i*32+j] = PIX.charAt(j)
        }
    }
}
    
```

**Step IV:** Take a marker M= -1. Using the following process the last two bits of alpha, red, green and blue of each pixel of each k number of enveloped images are OR ed to produce the pixels of the original image.

```

for i = 0 to w*h
{
    Consider 8 integer values from C0 to C7 and set all of
    them to 0.
    for SH_NO = 0 to k-1
    {
        c0 = c0 | STORE [SH_NO] [i*32+6]; // | is bitwise
    OR
        c1 = c1 | STORE [SH_NO] [i*32+7];
        c2 = c2 | STORE [SH_NO] [i*32+14];
        c3 = c3 | STORE [SH_NO] [i*32+15];
        c4 = c4 | STORE [SH_NO] [i*32+22];
    }
}
    
```

```

c5 = c5 | STORE [SH_NO] [i*32+23];
c6 = c6 | STORE [SH_NO] [i*32+30];
c7 = c7 | STORE [SH_NO] [i*32+31];
}
    
```

```

FINAL [++M] = c0;
FINAL [++M] = c1;
FINAL [++M] = c2;
FINAL [++M] = c3;
FINAL [++M] = c4;
FINAL [++M] = c5;
FINAL [++M] = c6;
FINAL [++M] = c7;
}
    
```

Create a one dimensional array IMG\_CONS[ ] of size (w/4)\*h to store constructed pixels.

Construct alpha, red, green and blue part of each pixel by taking consecutive 8 bit substring from FINAL[ ] starting from 0.

Construct pixel from these part and store it into IMG\_CONS[(w/4)\*h]

Generate image from IMG\_CONS[ ].

## 6. Experimental Result

**Division using Visual Cryptography:**

Source Image: Lena.png

Source image is



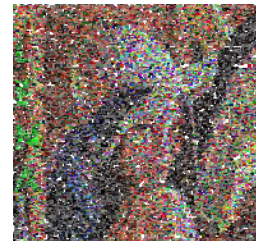
Fig 4: Source Image

Number of Shares: 4

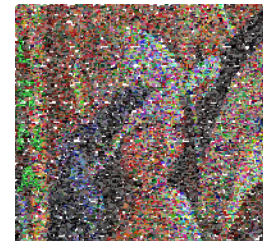
Numbers of shares to be taken: 3

Image shares produced after applying Visual Cryptography

are:



0img.png



1img.png



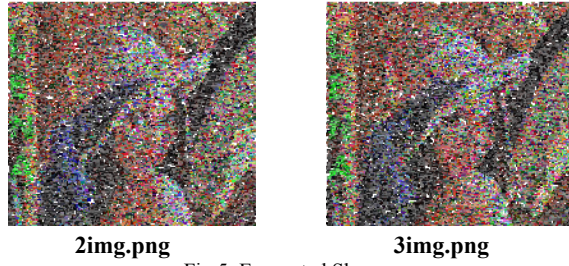


Fig 5: Encrypted Shares

**Enveloping using Watermarking:**



0img.png

+



Envelope0.png



Final0.png



1img.png

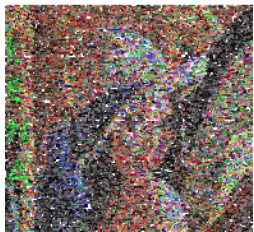
+



Envelope1.png



Final1.png



2img.png

+



Envelope2.png



Final2.png



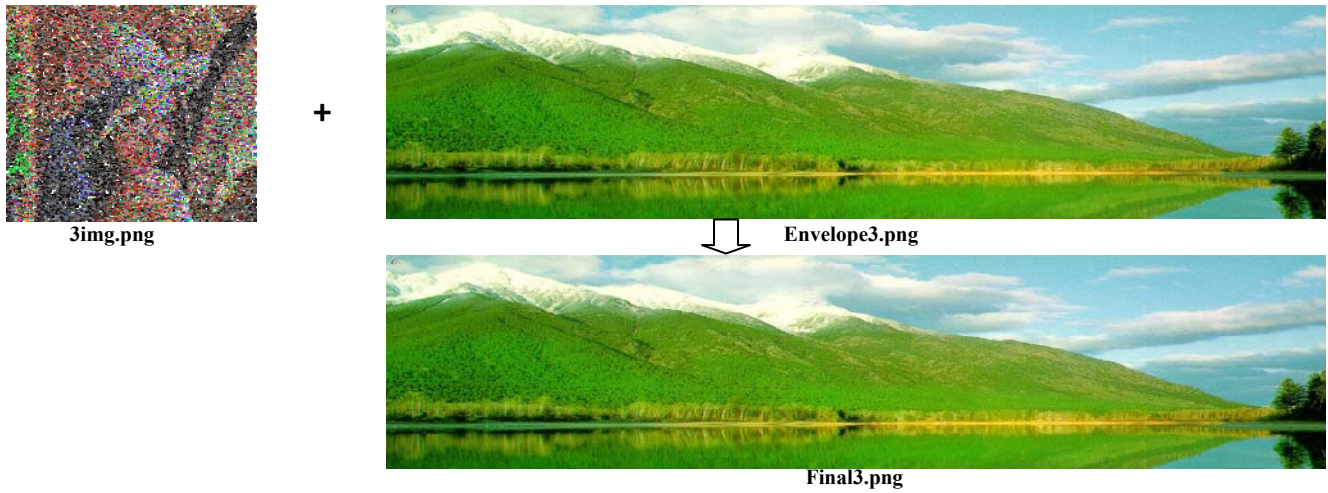


Fig 6: Enveloping shares using Digital Watermarking

**Decryption Process:**

Number of enveloped images taken: 3

Name of the images: Final0.png, Final2.png, Final3.png

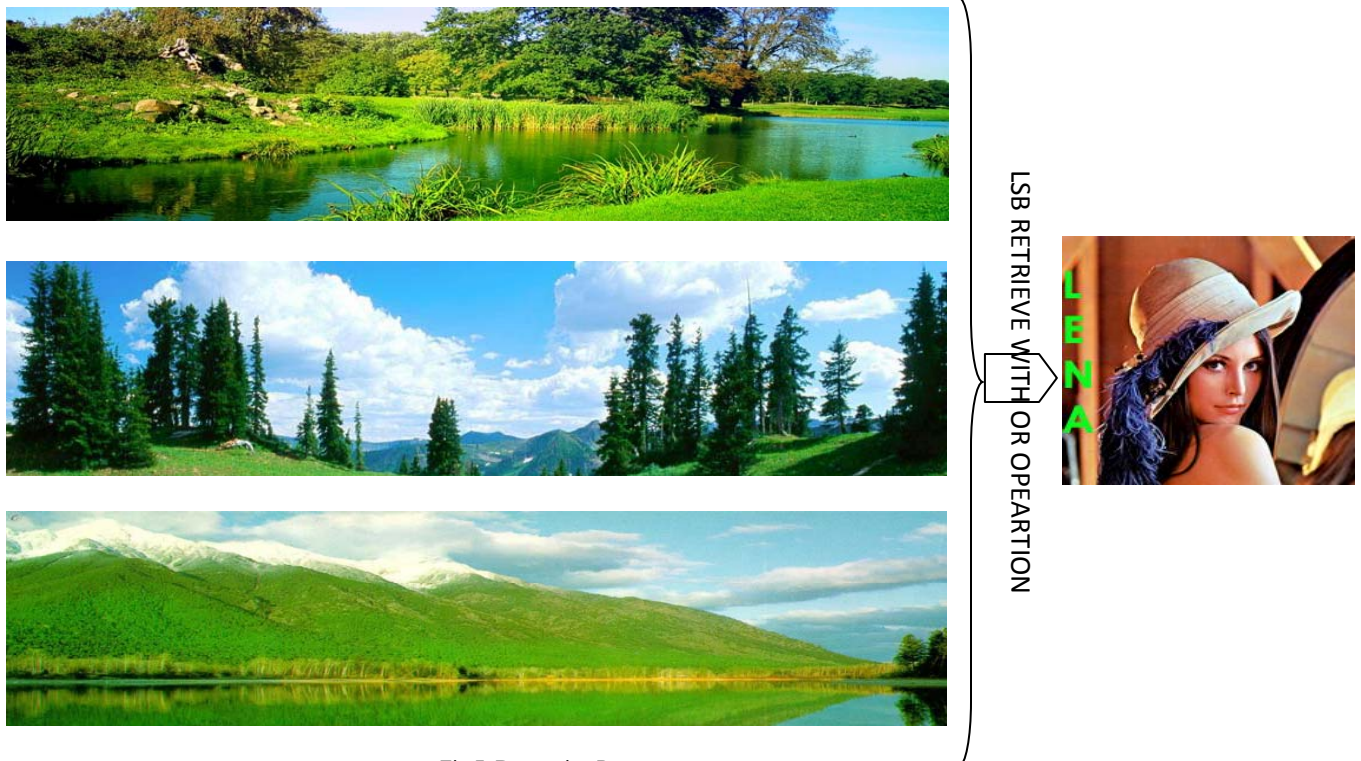


Fig 7: Decryption Process

## 7. Conclusion

Decryption part of visual cryptography is based on OR operation, so if a person gets sufficient k number of shares; the image can be easily decrypted. In this current work, with well known k-n secret sharing visual cryptography scheme an enveloping technique is proposed where the secret shares are enveloped within apparently innocent covers of digital pictures using LSB replacement digital watermarking. This adds security to visual cryptography technique from illicit attack as it befools the hackers' eye.

The division of an image into n number of shares is done by using random number generator, which is a new technique not available till date. This technique needs very less mathematical calculation compare with other existing techniques of visual cryptography on color images [10][11][12][13]. This technique only checks '1' at the bit position and divide that '1' into (n-k+1) shares using random numbers. A comparison is made with the proposed scheme with some other schemes to prove the novelty of the scheme.

Table 1: Margin specifications

Other Processes	Proposed Scheme
1. k-n secret sharing process is Complex[10][11][12].	1. k-n secret sharing process is simple as random number is used.
2. The shares are sent through different communication channels, which is a concern to security issue [10][11][12][13].	2. The shares are enveloped into apparently innocent cover of digital pictures and can be sent through same or different communication channels. Invisible digital watermarking befools the hacker.

## References:

[1] M. Naor and A. Shamir, "Visual cryptography," Advances in Cryptology-Eurocrypt'94, 1995, pp. 1-12.  
 [2] P. Ranjan, "Principles of Multimedia", Tata McGraw Hill, 2006.  
 [3] John F Koegel Buford, Multimedia Systems, Addison Wesley, 2000.  
 [4] Kandar Shyamalendu, Maiti Arnab, "K-N Secret Sharing Visual Cryptography Scheme For Color Image Using Random Number" International Journal of Engineering Science and Technology, Vol 3, No. 3, 2011, pp. 1851-1857.  
 [5] Naskar P., Chaudhuri A, Chaudhuri Atal, Image Secret Sharing using a Novel Secret Sharing Technique with Steganography, IEEE CASCOM, Jadavpur University, 2010, pp 62-65.  
 [6] Hartung F., Kuttter M., "Multimedia Watermarking Techniques", IEEE, 1999.

[7] S. Craver, N. Memon, B. L. Yeo, and M. M. Yeung, Resolving Rightful Ownerships with Invisible Watermarking Techniques: Limitations, Attacks and Implications. IEEE Journal on Selected Areas in Communications, Vol16, No.4 May 1998, pp.573-586.  
 [8] Schildt, H. The Complete Reference Java 2, Fifth Ed. TMH, Pp 799-839  
 [9] Krishmoorthy R, Prabhu S, Internet & Java Programming, New Age International, pp 234.  
 [10] F. Liu<sup>1</sup>, C.K. Wu<sup>1</sup>, X.J. Lin, Colour visual cryptography schemes, IET Information Security, July 2008.  
 [11] Kang InKoo et. al., Color Extended Visual Cryptography using Error Diffusion, IEEE 2010.  
 [12] SaiChandana B., Anuradha S., A New Visual Cryptography Scheme for Color Images, International Journal of Engineering Science and Technology, Vol 2 (6), 2010.  
 [13] Li Bai , A Reliable (k,n) Image Secret Sharing Scheme by, IEEE,2006.

## Appendix:

<sup>1</sup> Java Language implementation is

```
int c=0;
int a=(Integer.parseInt(value.substring(0,8),2))&0xff;
int r=(Integer.parseInt(value.substring(8,16),2))&0xff;
int g=(Integer.parseInt(value.substring(16,24),2))&0xff;
int b=(Integer.parseInt(value.substring(24,32),2))&0xff;
img_cons[c++]=(a << 24) | (r<< 16) | (g << 8) | b;
```



# Computation of Multiple Paths in MANETs Using Node Disjoint Method

M.Nagaratna<sup>1</sup>, P.V.S.Srinivas<sup>2</sup>, V.Kamakshi Prasad<sup>3</sup>, C.Raghavendra Rao<sup>4</sup>

<sup>1</sup> Assistant Professor, Department of Computer Science & Engg, JNTUH, Hyderabad, India

<sup>2</sup> Professor of Computer Science & Engg, Geethanjali College of Engineering & Technology, Hyderabad, India

<sup>3</sup> Professor of Computer Science & Engg, JNTUH, Hyderabad, India

<sup>4</sup> Professor of Computer Science & Engg, University of Hyderabad, Hyderabad, India

## Abstract

A Mobile Ad-hoc Network (MANET) is a kind of wireless ad-hoc network, and is also a self-configuring network, where in, mobile nodes are connected through wireless links. The topology of mobile ad-hoc networks is arbitrary and changes due to the consequent movement of the nodes. This causes frequent failures in the routing paths. This paper proposes the computation of multiple paths between a pair of Source and Destination, through which the data packets can be transmitted, and this improves the QoS parameters like reliability, Route Request Frequency and end-to-end delay.

**Keywords:** MANETs, QoS, Multipath, node disjoint, link disjoint.

## 1. Introduction

Mobile ad hoc networks (MANETs) comprise mobile nodes connected wirelessly to each other without the support of any fixed infrastructure or central administration. The nodes are self organized and can be deployed “on the fly” anywhere, any time to serve the need. Two nodes can communicate if they are within each other’s radio range, otherwise, intermediate nodes serve as routers if they are out of range, thereby it becomes multihop routing. These networks have several salient features like rapid deployment, robustness, flexibility, inherent mobility support, highly dynamic network topology, the limited battery power of mobile devices, limited capacity and asymmetric or unidirectional links. MANETs

can be deployed in emergency and rescue operations, disaster recovery conferences, etc [1].

The rest of this paper is organized as follows. Section 2 describes QoS Routing challenges in MANETs. Section 3 describes a brief need for Multipath Routing. Section 4 presents the Discovery of multiple paths. In Section 5, describes the computation of multiple paths in MANETs which is validated through MAT Lab. Section 6 provides an algorithm for finding multiple paths. Section 7 concludes this paper.

## 2. QoS Routing Challenges in MANETs

Because of the inherent properties of MANETs, establishing a stable path which can adhere to the QoS requirements is a big challenge. The stability issues of a data transmission system in a MANET can be studied under the following aspects [2], [7], [8].

1. Existence of mobile nodes (Mobility factor): Nodes in a MANET form the network only when they are in the communication range of each other. If they move out of range, link between two nodes is broken. At times, breakage of a single link can lead to the major network partitioning. Hence, mobility of the nodes is a major challenging issue

for a stable network. Also, breakdown of certain links results in routing decisions to be made again.

2. Limited battery / energy factor: Mobile nodes are battery driven. Therefore, the energy resources for such networks are limited. Also, the battery power of a mobile node depletes not only due to data transmission but also because of interference from the neighboring nodes. Thus, a node loses its energy at a specific rate even if it is not transferring any data packet. Hence the lifetime of a network largely depends on the energy levels of its nodes. Higher the energy level, higher is the link stability and hence, network lifetime. Also lower is the routing cost.

3. Multiple paths: To send data from a source to destination, a path has to be found before hand. If a single path is established, sending all the traffic on it will deplete all the nodes faster. Also, in case of path failure, alternate path acts as a backup path. Thus, establishing multiple paths aids not only in traffic engineering but also prevents faster network degradation.

4. Node-disjoint paths: Multiple paths between two nodes can be either link-disjoint or node disjoint. Multiple link-disjoint paths may have one node common among more than one path. Thus, traffic load on this node will be much higher than the other nodes of the paths. As a result, this node tends to die much earlier than the other nodes, leading to the paths to break down much earlier. Thus, the presence of node disjoint paths prolongs the network lifetime by reducing the energy depletion rate of a specific node [6].

### 3. Need for Multipath Routing

Either point to point or multipoint to multipoint data transmission is necessary for the applications of MANETs, which made the multicast technology as one of the emerging area by the researchers. However, in multicasting network congestion, network load imbalance and QoS degradation are easy to occur when the network load increases

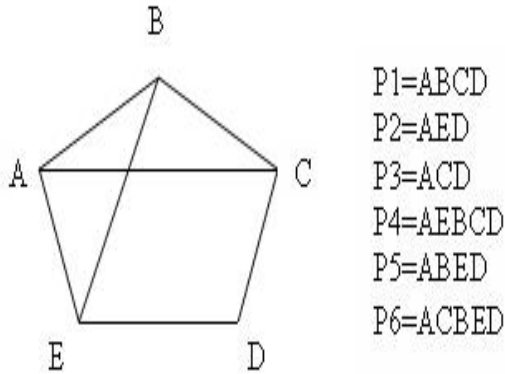
heavily. Multipath routing scheme has more advantages than unipath routing on the aspect of fault-tolerance, routing reliability and network load balance [3]. To improve the quality of MANET routing, multipath routing has attracted more and more research attentions.

### 4. Discovery of multiple paths

To discover multiple paths between a pair of source to destination the basic route discovery mechanisms used i.e DSR and AODV protocols. In fact, one of the major reasons for using multi path routing is to discover multiple paths by using either node disjoint or link disjoint methods. In the node disjoint method, nodes on the paths should not be common, where as, in the link disjoint method; links on the paths should not be common. Hence, the route discovery mechanisms of the existing routing protocols need to be modified to discover a maximum number of node- disjoint or link disjoint paths. Once all node disjoint or link disjoint paths have been discovered, there arises other issues like how to select a suitable path or a set of paths from all the discovered paths and what node should make this selection, namely the source or the destination [4, 5].

### 5. Computation of Multiple Paths in MANETs

A – Source node  
D – Destination node



For the above example graph we have constructed the path matrix by using the node disjoint paths. In the path matrix, number of paths is placed in rows, and number of vertices is placed in columns. For every path we identify the vertices, if the vertex is there in the path then we assign the value 1 for the corresponding vertex otherwise we assign the value 0.

Path Matrix:

	A	B	C	D	E	Weight
P1	1	1	1	1	0	4
P2	1	0	0	1	1	3
P3	1	0	1	1	0	3
P4	1	1	1	1	1	5
P5	1	1	0	1	1	4
P6	1	1	1	1	1	5

In the above path matrix, minimum weight is 3, here we have two minimum weight paths i.e., P2 and P3. By default we select P2 as first path.

To find Multiple Paths:

First we find the Hamming distance by using the path matrix. In hamming distance matrix the number of paths is taken as rows and columns i.e.,

symmetric matrix. To find the hamming distance we count the dissimilar values of P1 and the remaining individual paths. Similarly P2, P3, P4, P5 and P6.

Hamming distance matrix:

	P1	P2	P3	P4	P5	P6
P1	0	3	1	1	2	1
P2	3	0	2	2	1	2
P3	1	2	0	2	3	2
P4	1	2	2	0	1	0
P5	2	1	3	1	0	1
P6	1	2	2	0	1	0

From Node Disjoint Path matrix minimum weight path is selected as first path i.e., P2.

In P2 (row2) the maximum value is 3 that value is in column1 (P1).

Now we select second path P1.

For finding the next path is maximum sum of P1 and P2.

0	3	1	1	2	1
3	0	2	2	1	2
3	3	3	3	3	3

Here all the values are same. From this P1 and P2 are already selected. So the remaining paths are P3, P4, P5 and P6. By default we select third path P3.

For finding the next path is maximum sum of P1, P2 and P3.

0	3	1	1	2	1
3	0	2	2	1	2
1	2	0	2	3	2
<hr/>					
4	5	3	5	6	5
<hr/>					

Here the remaining paths are P4, P5 and P6. From this maximum value is 6. so that select P5. Now the fourth path is P5.

For finding the next path is maximum sum of P1, P2, P3 and P5.

0	3	1	1	2	1
3	0	2	2	1	2
1	2	0	2	3	2
2	1	3	1	0	1
<hr/>					
6	6	6	6	6	6
<hr/>					

Here the remaining paths are P4 and P6. Both are having the same value i.e., 6. By default we select fifth path is P4. The remaining path P6 is sixth path.

The sequences of multiple paths are **P2, P1, P3, P5, P4 and P6.**

6. Algorithm

1. From the given network select the source and destination node. Next find all possible paths from the source to destination node by using node disjoint method.
2. Next find the Node disjoint path matrix.
3. After finding the Node disjoint path matrix, for every path find the weight. Now select the minimum weight. This minimum weight path is first path of the given network
4. To find the multiple paths develop the hamming distance matrix. Read the dissimilar values of first path and the remaining individual paths
5. In the first path read the maximum value from the hamming distance matrix. Read the maximum value of the corresponding column number that is to be taken as second path.
6. Find the maximum sum of first path and second path.
7. If all the values are same then select any value and read the corresponding column number that is to be taken as third path.
8. Otherwise find the maximum value, and read the corresponding column number that is to be taken as third path.
9. repeat step number 6 until all the paths are computed.

## 7. Conclusions:

Since mobile nodes are potentially mobile in nature and infrequent path failures in MANETs inevitable, we propose an algorithm through which multiple paths can we computed there by control overhead can be drastically reduced.

## References:

- [1]. Osamah S. Badarneh and Michel Kadoch " Multicast Routing Protocols in Mobile Ad Hoc Networks: A Comparative Survey and Taxonomy" Hindawi Publishing Corporation EURASIP Journal on Wireless Communications and Networking Volume 2009, Article ID 764047, 42 pages
- [2]. Dr. Shuchita Upadhyaya and Charu Gandhi " Node Disjoint Multipath Routing Considering Link and Node Stability protocol: A characteristic Evaluation" International Journal of Computer Science Issues, Vol. 7, Issue 1, No. 2, January 2010.

[3]. Hong Tang, Fei Xue and Peng Huang,  
“MP-MAODV: a MAODV-Based Multipath Routing  
Algorithm” IFIP International Conference on Network  
and Parallel Computing 2008 IEEE DOI  
10.1109/NPC.2008.23 296

[4]. Mohammed Tarique a,\_, KemalE.Tepe b,  
SasanAdibi c, ShervinErfani “Survey of multipath  
routing protocols for mobile ad hoc networks” Journal of  
Network and Computer Applications 32 (2009) 1125–  
1143

[5]. Y. Ganjali and A. Keshavarzian, "Load balancing in  
ad hoc networks:  
Single-path routing vs. multi-path routing", in  
Proceedings of IEEE. INFOCOM, 2004.

[6]. S. Lee and M. Gerla, "*Split multipath routing with  
maximally disjoint paths in ad hoc networks*",  
Proceedings of the IEEE ICC, pp. 3201-3205, June 2001.

[7]. H. Hassanein and A. Zhou, Routing with Load  
Balancing in Wireless AdHoc Networks, in Proc. ACM  
MSWIM, Rome, Italy, July (2001).

[8]. M. Abolhasan, T. Wysocki, E. Dutkiewicz, “A  
review of routing protocols for mobile ad hoc networks,  
Ad Hoc Networks,” Vol. 2, No. 1, pp.1-22, 2004.



# WLAN Security: Active Attack of WLAN Secure Network (Identity theft)

Anil Kumar Singh<sup>1</sup>, Bharat Mishra<sup>2</sup>

<sup>1</sup>Jagran Institute of Management  
Kanpur- 208014 (India)

<sup>2</sup>MGCGV, Chitrakoot  
Satna (M.P.) India

## Abstract

In Wireless Local Area Network data transfer from one node to another node via air in the form of radio waves. There is no physical medium for transferring the data like traditional LAN. Because of its susceptible nature WLAN can open the door for the intruders and attackers that can come from any direction. Security is the most important element in WLAN. MAC address filtering is one of the security methods for securing the WLAN. But it is also vulnerable. In this paper we will demonstrate how hackers exploit the WLAN vulnerability (Identity theft of legitimate user) to access the Wireless Local Area Network.

**Keywords:** - *WLAN, MAC address, Access Point, WNIC, Wi-Fi*

## Introduction

Wi-fi technology has played a very significant role in IT revolution and continues to do so. After 2 decades it is very popular among the It fraternity. Many companies, Educational Institutions, Airports as well as domestic users make use of the WLAN facility. Security is an important factor of Wireless Local

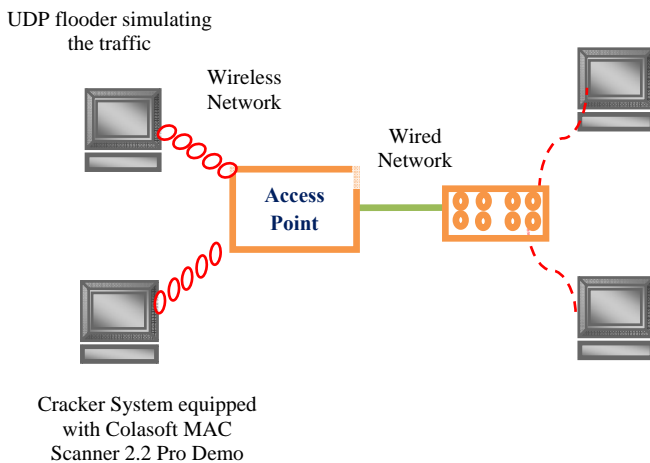
Area Network because of its nature. D-Link, Linksys are providing the WLAN security with the help of MAC address.[1] and WEP key. It is noted that the MAC address filtering is the gateway for hackers to enter and access the facility of Wireless Local Area Network.

## Material and methods

The research was carried out to reveal WLAN Security: Active Attack on WLAN Secure Network

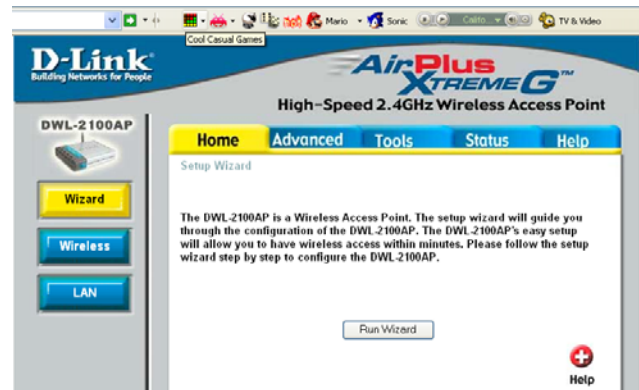
(Identity theft). The work was conducted at Department of Information Technology, Jagran Institute of Management. Materials used and the procedures employed are as follows: We can design a scenario after understanding the theory of WLAN security with the help of MAC address filtering. We have taken the Colasoft MAC Scanner 2.2 Pro Demo. There are hardware such as: HCL Desktop, Toshiba Laptops, AP (D-Link 2100 Series Access Point) and Wireless card (D-Link DWA 510).

Softwares such as: Operating System (Windows XP) and other application softwares. One client is used to communicate with Access Point. Another client is used to keep track of the network traffic as a hacker and listens to the WLAN. AP is linked to LAN with wires. Figure 1 is the illustration of Identity theft job.

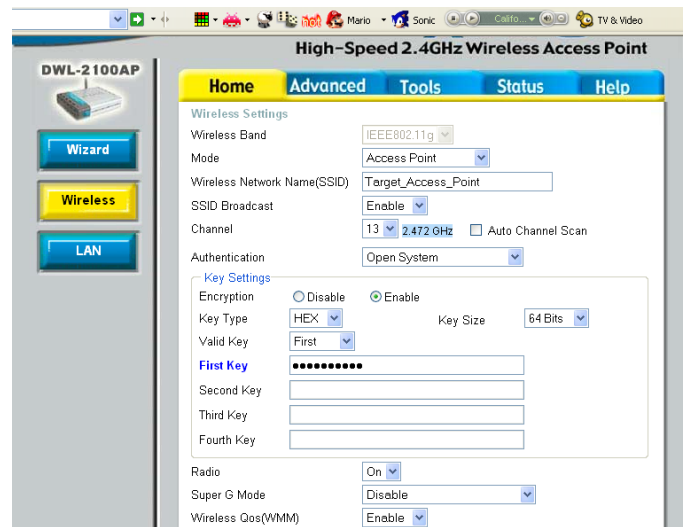


**Figure - 1 Identity Theft gears**

Open the internet explorer and type the IP of the access point 80.0.99.6 in address bar and press enter, Access Point will display the following window



Click on wireless tab



Click on advance tab, Click on filter, write the MAC address of legitimate user. Target searching the MAC address click on start, click on run, type cmd and

```

C:\WINDOWS\system32\cmd.exe

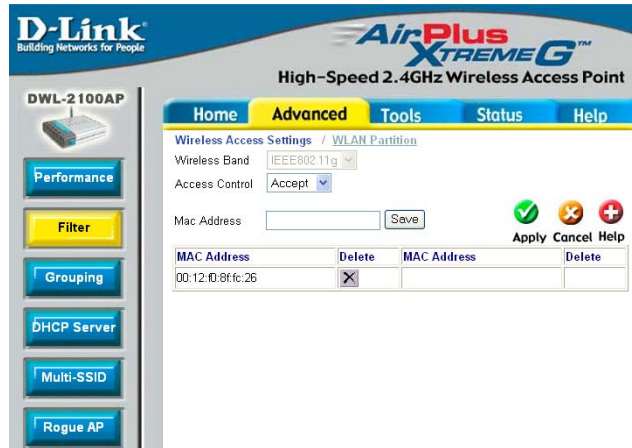
C:\Documents and Settings>getmac

Physical Address      Transport Name
-----
Disabled             Disconnected
00-12-F0-8F-FC-26    \Device\Ncpip_{DF77EF7A-761D-427B-A31A-ABC

C:\Documents and Settings>_
    
```

again type **getmac**, this command will display the MAC address of the WNIC

Click on access control displays three options namely disable, accept and reject, click on accept it means only authorized MAC address can access the WLAN, write the MAC address and save.

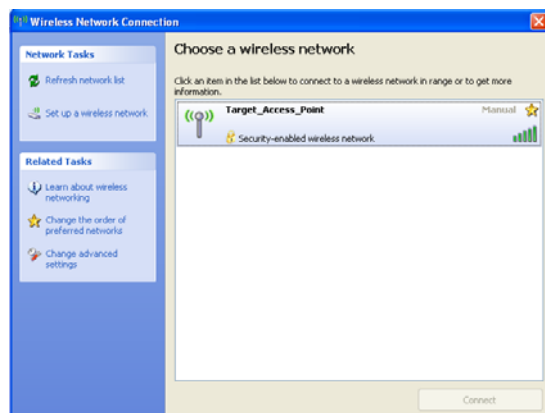


Now we are going to another computer to access the wireless local area network which MAC address is displaying below:

C:\>getmac

```
1C-AF-F7-0C-CC-8C \Device\Tcpip_{12361AAF-5538-4489-87B4-C9BB984E1299}
```

Now we are trying to connect the Target\_Access\_Point wireless network

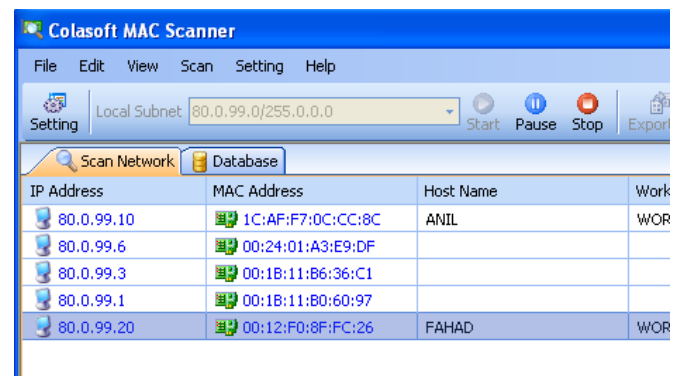


After that the system is not connected the wireless LAN.

Then we hack the MAC address of the legitimate user with the help of Cola soft MAC Scanner 2.2 Pro Demo [2].

After that the system will not be connected to the wireless LAN, and then we hack the MAC address of legitimate user with the help of Colasoft MAC Scanner 2.2 Pro Demo (it can be downloaded to [http://www.colasoft.com/mac\\_scanner/](http://www.colasoft.com/mac_scanner/)

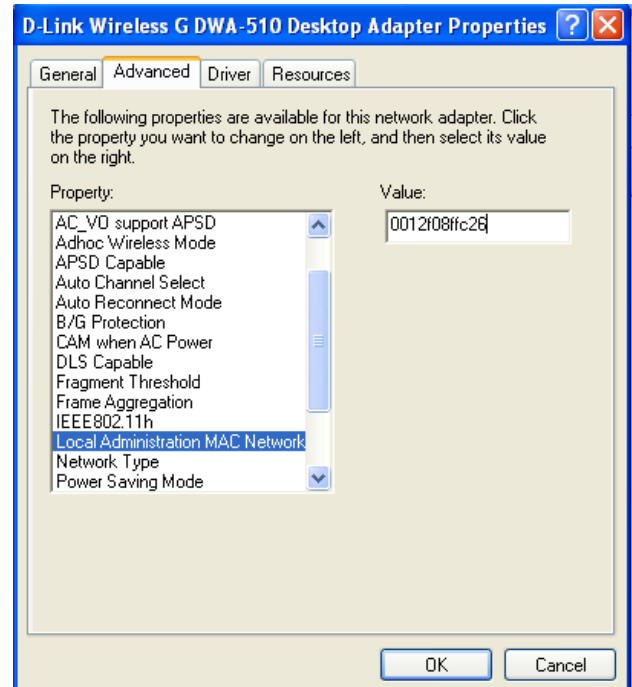
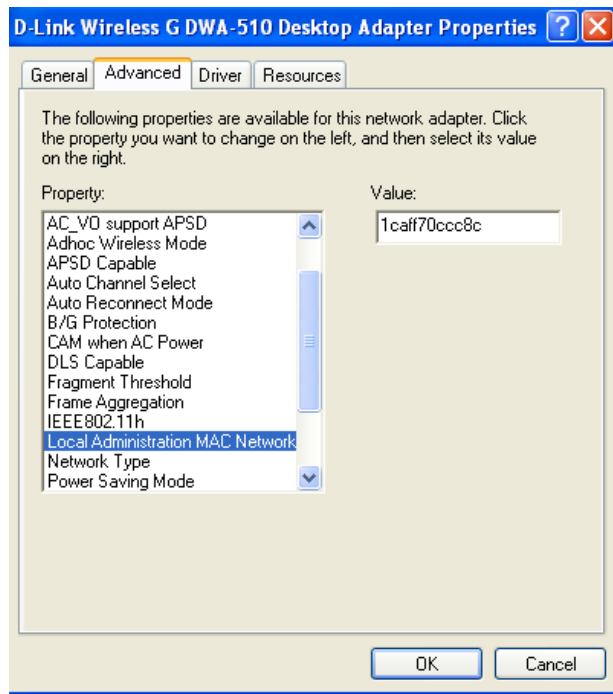
Double click on colasoft Scanner



Here you can see the highlighted MAC address. This is the identity of authorized user namely FAHAD.

Now we change the identity with the help of following process:

Click on start, Go to control panel, double click on Network connection, right click on Wireless Network connection, click on configure, click on advance



Select Local Administration MAC network, here you can see in value column it displays the MAC address. Now you can replace the identity of the existing user.

Type the MAC address in value column  
0012f08ffc26

After changing the MAC address the hacker can easily access the WLAN without any barrier.

## Conclusion

Due to the broadcast nature of the wireless communication, it becomes an easy prey for an attacker to capture wireless communication or to disturb the normal operation of the network by injecting additional traffic.

WLAN is also prone to unauthorized intervention by hackers where they create conditions for the theft of

the identity (MAC address) of an authorized user. The access point cannot filter the MAC address. Because it checks their database and matches the MAC address, if found it allows accessing the WLAN.

To avoid this type of vulnerability we will strongly recommend that the administrator should use the combination of enabling WEP key and MAC address filter security mechanism. [3]

## References

1. **Bradley Mitchell**, Enable MAC Address Filtering on Wireless Access Points and Routers Improve home network security
2. **Downloaded Cola soft MAC Scanner 2.2 Pro Demo by :**  
[http://www.colasoft.com/mac\\_scanner/](http://www.colasoft.com/mac_scanner/)
3. **Anil Kumar Singh, (2011)**, Wireless Local Area Network: Security from unauthorized access, proceedings of NCICT, Ewing Christian College Allahabad, Excel India Publishers New Delhi

**Anil Kumar Singh, MCA** – Asst. Professor, Jagran Institute of Management, Kanpur. Currently pursuing the Doctoral programme in WLAN Security Vulnerability Threats and Alternative Solution at MGCV Satna (M.P.)

**Dr. Bharat Mishra, Ph.D.**, Dept. of Physical Science. MGCGV Satna (M.P.)



# Mining databases on World Wide Web

MANALI GUPTA  
STUDENT M.TECH CS (1<sup>ST</sup> YEAR)  
AMITY UNIVERSITY,NOIDA

VIVEK TOMAR  
STUDENT M.TECH CS (1<sup>ST</sup> YEAR)  
AMITY UNIVERSITY,NOIDA

JAYA VERMA  
STUDENT M.TECH IT (2<sup>ND</sup> YEAR)  
G.G.S.I.P.U., Delhi.

SUDEEPA ROY  
STUDENT M.TECH CS (1<sup>ST</sup> YEAR)  
AMITY UNIVERSITY,NOIDA

**Abstract**— The power of the WWW comes not simply from static HTML pages - which can be very attractive, but the important first step into the WWW is especially the ability to support those pages with powerful software, especially when interfacing to databases. The combination of attractive screen displays, exceptionally easy to use controls and navigational aids, and powerful underlying software, has opened up the potential for people everywhere to tap into the vast global information resources of the Internet [1]. There is a lot of data on the Web, some in databases, and some in files or other data sources. The databases may be semi structured or they may be relational, object, or multimedia databases. These databases have to be mined so that useful information is extracted.

While we could use many of the data mining techniques to mine the Web databases, the challenge is to locate the databases on the Web. Furthermore, the databases may not be in the format that we need for mining the data. We may need mediators to mediate between the data miners and the databases on the Web. This paper presents the important concepts of the databases on the Web and how these databases have to be mined to extract patterns and trends.

**Keywords** - Data Mining, Web Usage Mining, Document Object Model, KDD dataset

## I. INTRODUCTION

Data mining slowly evolves from simple discovery of frequent patterns and regularities in large data sets toward interactive, user-oriented, on-demand decision supporting. Since data to be mined is usually located in a database, there is a promising idea of integrating data mining methods into Database Management Systems (DBMS) [6]. Data mining is the process of posing queries and extracting patterns, often previously unknown from large quantities of data using pattern matching or other reasoning techniques.

## II. CHALLENGES FOR KNOWLEDGE DISCOVERY

Data mining, also referred to as database mining or knowledge discovery in databases (KDD) is a research area that aims at the discovery of useful information from large datasets. Data

Mining [9] uses statistical analysis and inference to extract interesting trends and events, create useful reports, support decision making etc. It exploits the massive amounts of data to achieve business, operational or scientific goals. However based on the following observations the web also poses great challenges for effective resource and knowledge discovery.

- *The web seems to be too huge for effective data warehousing and data mining.* The size of the web is in the order of hundreds of terabytes and is still growing rapidly. many organizations and societies place most of their public-accessible information on the web. It is barely possible to set up data warehouse to replicate, store, or integrate of the data on the web.
- *The complexity of web pages is greater than that of any traditional text document collection.* Web pages lack a unifying structure. they contain far more authoring style and content variations than any set of books or other traditional text based documents. The web is considered a huge digital library; however the tremendous number of documents in this library is not arranged according to any particular sorted order. There is no index by category, nor by title, author, cover page, table of contents and so on.
- *The web is a highly dynamic information source.* Not only does the web grow rapidly, but its information is also constantly updated. News, stock markets, weather, airports, shopping, company advertisements and numerous other web pages are updated regularly on the web.
- *The web serves a broad diversity of user communities.* The internet currently connects more than 100 million workstations, and its user community is still rapidly expanding. Most users may not have good knowledge of the structure of the information network and may not be aware of the heavy cost of a particular search.

- *Only a small portion of the information on the web is truly relevant or useful.* It is said that 99 % of the web information is useless to 99 % of web users. Although this may not seem obvious, it is true that a particular person is generally interested in only a tiny portion of the web, while rest of the web contains information that is uninteresting to the user and may swamp desired such results.

These challenges have promoted search into efficient and effective discovery and use of resources on the internet. There are many index based **Web search engines**. These search the web, index web pages, and build and store huge keyword-based indices that help locate sets of web pages containing certain keywords [7]. However a simple keyword based search engine suffers from several deficiencies. First, a topic of any breadth can easily contain hundreds of thousands of documents. This can lead to a huge number of document entries returned by a search engine, many of which are only marginally relevant to the topic or may contain materials of poor quality. Second, many documents that are highly relevant to a topic may not contain keywords defining them. This is referred to as the **polysemy problem**. For example, the keyword Oracle may refer to the oracle programming language, or an island in Mauritius or brewed coffee. So a search based on the keyword, search engine may not find even the most popular web search engines like Google, Yahoo!, AltaVista if these services do not claim to be search engines on their web pages.

So a keyword-based web search engine is not sufficient for the web discovery, then Web mining should have to be implemented in it. Compared with keyword-based Web search, Web mining is more challenging task that searches for web structures, ranks the importance of web contents, discovers the regularity and dynamics of web contents, and mines Web access patterns. However, Web mining can be used substantially enhances the power of documents, and resolve many ambiguities and subtleties raised in keyword-based web search. Web mining tasks can be classified into three categories:

- Web content mining
- Web structure mining
- Web usage mining.

### III. WEB CONTENT MINING

The concept of web content mining is far wider than searching for any specific term or only keyword extraction or some simple statics of words and phrases in documents. For example a tool that performs web content mining can summarize a web page so that to avoid the complete reading of a document and save time and energy. Basically there are two models to implement web content mining. The first model is known as local knowledgebase model. According to this model, the abstract characterizations of several web pages are stored locally. Details of these characterizations vary on different systems [8]. For example, there are three categories of

web sites: games, educational and others. References to several web sites relating to these categories are stored in a database.

When extracting information, first the category is selected and then a search is performed within the web sites referred in this category. A query language enables you to query the database consisting of information about various categories at several levels of abstraction. As a result of the query, the system using this model for web content mining may have to request web pages from the web that matches the query. The concept of artificial intelligence is highly used to build and manage the knowledgebase consisting of information on various classes of web sites. The second approach is known as agent based model. This approach also applies the artificial intelligence systems, known as web agents that can perform a search on behalf of a particular user for discovering and organizing documents in the web.

### IV. WEB USAGE MINING

The concept of web mining that helps automatically discovering user access patterns. For example, there are four products of a company sold through the web site of a company. Web usage mining analyses the behavior of the customers [8]. This means by using a web usage mining tool the nature of the customers that is which product is most popular, which is less, which city has the maximum number of customers and so on.

### V. WEB STRUCTURE MINING

Denotes analysis of the link structure of the web. web structure mining is used for identifying more preferable documents. For example, the document A in web site X has a link to the document B in the web site Y [11]. According to Web structure mining concept, document B is important to the web site A, and contains valuable information. The hyperlink induced Topic search (HITS) is a common algorithm for knowledge discovery in the web.

### VI. MINING THE WEB PAGE LAYOUT STRUCTURE.

Compared with traditional plain text, a web page has more structure. Web pages are also regarded as semi-structured data. The basic structure of a web page is its DOM[3](Document object model) structure. The DOM structure of a web page is a tree structure where every HTML tag in the page corresponds to a node in the DOM tree. The web page can be segmented by some predefined structural tags. Useful tags include <P>(paragraph), <TABLE>(table), <UL>(list), <H1>~<H6>(heading) etc. Thus the DOM structure can be used to facilitate information extraction. Figure 1 illustrates HTML DOM Tree Example [2]:

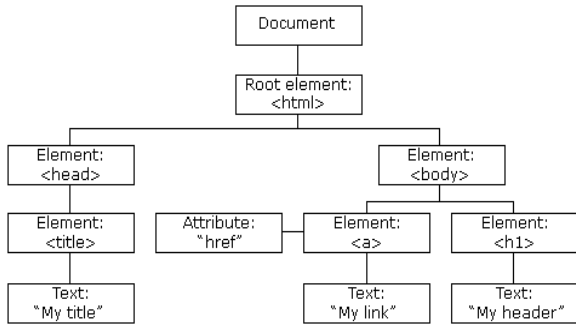


Figure 1. HTML DOM Tree Example

Here's the DOM object tree generated by the code for the TABLE element and its child elements [4]:

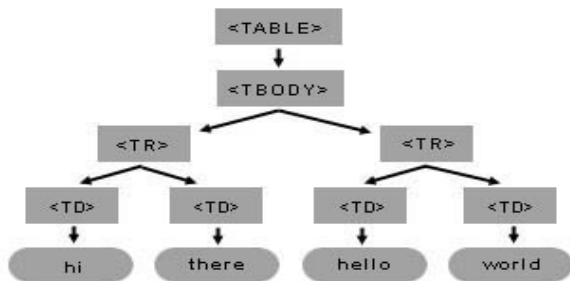


Figure 2. DOM Object Tree Example

Moreover, the DOM tree was initially introduced for presentation in the browser rather than the description of the semantic structure of the web page. For example, even though the two nodes in the Dom tree have the same parent, the two nodes might not be more semantically related to each other than to other nodes.

In the sense of human perception, people always view a web page as different semantic objects rather than as a single object. Some research efforts show that users always expect that certain functional parts of a web page appear at certain positions on the page. Actually, when a web page is presented to the user, the spatial and visual cues can help the user unconsciously divide the web page into several semantic parts. therefore it is possible to automatically segment the web pages by using the spatial and visual cues. Based on this observation there is an algorithm called Vision-based page segmentation (VIPS).VIPS aims to extract the semantic structure of a web page based on its visual presentation. Such semantic structure is a tree structure: each node in the tree corresponds how coherent is the content in the block based on visual perception. The VIPS algorithm makes full use of the page layout feature. It first extracts all of the suitable blocks from the HTML DOM tree, and then it finds the separators between these blocks. here

separators denote the vertical or horizontal lines in a web page that visually cross with no blocks. Based on the separators, the semantic tree of the web page is constructed. A web page can be represented as a set of blocks(leaf nodes of the semantic tree)Compared with the DOM- based methods, the segments obtained by VIPS are more semantically aggregated.

## VII. MINING THE WEB'S LINK STRUCTURES TO IDENTIFY AUTHORITATIVE WEB PAGES

Suppose to search for Web pages relating to a given topic, such as financial investing. In addition to retrieving pages that are relevant, the pages retrieved should be of high quality, or authoritative on the topic. the secrecy of authority is hiding in Web page linkages. The Web consists not only of pages, but also of hyperlinks pointing from one page to another. These hyperlinks contain an enormous amount of latent human annotation that can help automatically infer the notion of authority. When an author of a Web page creates a hyperlink pointing to another Web page, this can be considered as the author's endorsement of the other page. The collective endorsement of a given page by different authors on the Web may indicate the importance of the page and may naturally lead to the discovery of authoritative Web pages. Therefore, the tremendous amount of Web linkage information provides rich information about the relevance, the quality, and the structure of the Web's contents, and thus is a rich source for Web mining.

However, the Web linkage structure has some unique features. First, not every hyperlink represents the endorsement we seek. Some links are created for other purposes, such as for navigation or for paid advertisements. Yet overall, if the majority of hyperlinks are for endorsement, then the collective opinion will still dominate. Second, for commercial or competitive interests, one authority will seldom have its Web page point to its rival authorities in the same field. For example, Coca-Cola may prefer not to endorse its competitor Pepsi by not linking to Pepsi's Web pages. Third, authoritative pages are seldom particularly descriptive. For example, the main Web page of Yahoo! may not contain the explicit self-description "Web search engine."

These properties of Web link structures have led researchers to consider another important category of Web pages called a hub. A hub is one or a set of Web pages that provides collections of links to authorities. Hub pages may not be prominent, or there may exist few links pointing to them; however, they provide links to a collection of prominent sites on a common topic. In general, a good hub is a page that points to many good authorities; a good authority is a page pointed to by many good hubs. Such a mutual reinforcement relationship between hubs and authorities helps the mining of authoritative Web pages and automated discovery of high-quality Web structures and resources.

An algorithm using hubs, called HITS (Hyperlink-Induced Topic Search), is a common algorithm for knowledge discovery in the web. HITS is a web searching method where the searching logic partially depends on hyperlinks to identify

and locate the documents relating to a topic in the web. The HITS algorithm discovers the hubs and authorities of a community on a specific topic or query. In HITS algorithm the number of links between web sites is measured as weights. For a web site  $w$ , the weight of authority denotes the number of web sites containing a hyperlink to the web site  $w$ . Similarly the weight of the hub denotes the number of hyperlinks in the web site  $x$  pointing to other web sites.

The steps in the HITS algorithm are:-

- Accept the seed set,  $S$ , returned by a search engine. The set,  $S$  contains  $n$  number of web pages, where usually value of  $n$  lies between 0 to 200, means,  $n > 0$  and  $n \leq 200$ .
- Initialize the weight of the hub to 1 for each web page,  $p$  in the set,  $S$ . this means, assign  $hub\_weight(p)=1$ , for each  $p$ , where  $p \in S$ .
- Initialize the weight of the authority to 1 for each web page,  $p$  in the set,  $S$ . this means, assign  $authority\_weight(p)=1$  for each  $p$ , where  $p \in S$ .
- Let the expression  $p \rightarrow q$  denote that the web page  $p$  has a hyperlink to the web page  $q$ .
- Iteratively update weight of the authority, and weight of the hub for each page,  $p$  in the set,  $S$ . Repeat this step for a predetermined fixed number of times by calculating:
 
$$authority\_weight(p) = \sum_{q \rightarrow p} hub\_weight(q) \quad (1.1)$$

$$hub\_weight(p) = \sum_{p \rightarrow q} authority\_weight(q) \quad (1.2)$$
- Stop.

Equation (1.1) implies that if a page is pointed to by many good hubs, its authority weight should increase (i.e., it is the sum of the current hub weights of all of the pages pointing to it). Equation (1.2) implies that if a page is pointing to many good authorities, its hub weight should increase (i.e., it is the sum of the current authority weights of all of the pages it points to).

These equations can be written in matrix form as follows. Let us number the pages  $\{1, 2, \dots, n\}$  and define their adjacency matrix  $A$  to be an  $n \times n$  matrix where  $A(i, j)$  is 1 if page  $i$  links to page  $j$ , or 0 otherwise. Similarly, we define the authority weight vector  $a = (a_1, a_2, \dots, a_n)$ , and the hub weight vector  $h = (h_1, h_2, \dots, h_n)$ . Thus, we have

$$h = A \cdot a \quad (1.3)$$

$$a = A^T \cdot h, \quad (1.4)$$

where  $A^T$  is the transposition of matrix  $A$ . Unfolding these two equations  $k$  times, we have [3]

$$h = A \cdot a = AA^T h = (AA^T)h = (AA^T)^2 h = \dots = (AA^T)^k h \quad (1.4)$$

$$a = A^T \cdot h = A^T A a = (A^T A)a = (A^T A)^2 a = \dots = (A^T A)^k a. \quad (1.5)$$

According to linear algebra, these two sequences of iterations, when normalized, converge to the principal eigenvectors of  $AAT$  and  $ATA$ , respectively. This also proves that the authority and hub weights are intrinsic features of the linked pages collected and are not influenced by the initial weight settings.

Finally, the HITS algorithm outputs a short list of the pages with large hub weights, and the pages with large authority weights for the given search topic. Many experiments have shown that HITS provides surprisingly good search results for a wide range of queries. Although relying extensively on links can lead to encouraging results, the method may encounter some difficulties by ignoring textual contexts. The problems faced in the HITS are:-

- This algorithm does not have an effect of automatically generated hyperlinks.
- The hyperlinks pointing to the irrelevant or less relevant documents are not excluded and cause complications for updating hub and authority weights.
- A hub may contain various documents covering multiple topics. The HITS algorithm faces problem to concentrate on the specific topic mentioned by the query. This problem is called drifting.
- Many web pages across various web sites sometimes points to the same document. This problem is referred to as topic hijacking.

Such problems can be overcome by replacing the sums of Equations (1.1) and (1.2) with weighted sums, scaling down the weights of multiple links from within the same site, using anchor text (the text surrounding hyperlink definitions in Web pages) to adjust the weight of the links along which authority is propagated, and breaking large hub pages into smaller units.

By using the VIPS algorithm, we can extract page-to block and block-to-page relationships and then construct a page graph and a block graph. Based on this graph model, the new link analysis algorithms are capable of discovering the intrinsic semantic structure of the Web. Thus, the new algorithms can improve the performance of search in Web context. The graph model in block-level link analysis is induced from two kinds of relationships, that is, block-to-page (link structure) and page-to-block (page layout).

The block-to-page relationship is obtained from link analysis. Because a Web page generally contains several semantic blocks, different blocks are related to different topics.

Therefore, it might be more reasonable to consider the hyperlinks from block to page, rather than from page to page. Let  $Z$  denote the block-to-page matrix with dimension  $n \times k$ .  $Z$  can be formally defined as follows:

$$Z_{ij} = \begin{cases} 1/s_i & \text{if there is a link from block } i \text{ to page } j \\ 0, & \text{otherwise,} \end{cases} \quad (1.6)$$

where  $s_i$  is the number of pages to which block  $i$  links.  $Z_{ij}$  can also be viewed as a probability of jumping from block  $i$  to page  $j$ .

The block-to-page relationship gives a more accurate and robust representation of the link structures of the Web.

The page-to-block relationships are obtained from page layout analysis. Let  $X$  denote the page-to-block matrix with dimension  $k \times n$  [7]. As we have described, each Web page can be segmented into blocks. Thus,  $X$  can be naturally defined as follows:

$$X_{ij} = \begin{cases} f_{pi}(b_j), & \text{if } b_j \in p_i \\ 0, & \text{otherwise,} \end{cases} \quad (1.7)$$

where  $f$  is a function that assigns to every block  $b$  in page  $p$  an importance value. Specifically, the bigger  $f_p(b)$  is, the more important the block  $b$  is. Function  $f$  is empirically defined below,

$$f_p(b) = \alpha \times \frac{\text{the size of block } b}{\text{the distance between the center of } b \text{ and the center of the screen}} \quad (1.8)$$

where  $\alpha$  is a normalization factor to make the sum of  $f_p(b)$  to be 1, that is,

$$\sum_{b \in p} f_p(b) = 1$$

Note that  $f_p(b)$  can also be viewed as a probability that the user is focused on the block  $b$  when viewing the page  $p$ . Some more sophisticated definitions of  $f$  can be formulated by considering the background color, fonts, and so on. Also,  $f$  can be learned from some relabeled data (the importance value of the blocks can be defined by people) as a regression problem by using learning algorithms, such as support vector machines and neural networks. Based on the block-to-page and page-to-block relations, a new Web page graph that incorporates the block importance information can be defined as

$$W_p = XZ, \quad (1.9)$$

where  $X$  is a  $k \times n$  page-to-block matrix, and  $Z$  is a  $n \times k$  block-to-page matrix. Thus  $W_p$  is a  $k \times k$  page-to-page matrix.

## VIII. CONCLUSION

This paper has presented the details of tasks that are necessary for performing Web Usage Mining, the application of data mining and knowledge discovery techniques to WWW server access logs [5]. The World Wide Web serves as a huge, widely distributed, global information service center for news, advertisements, consumer information, financial management, education, government, e-commerce, and many other services. It also contains a rich and dynamic collection of hyperlink

information, and access and usage information, providing rich sources for data mining. Web mining includes mining *Web linkage structures*, *Web contents*, and *Web access patterns*. This involves mining the *Web page layout structure*, mining the *Web's link structures* to identify *authoritative Web pages*, mining *multimedia data* on the Web, *automatic classification of Web documents*, and *Web usage mining*. Data mining is an evolving technology going through continuous modifications and enhancements. Mining tasks and techniques use algorithms that are many a times refined versions of tested older algorithms. Though mining technologies are still in their infancies, yet they are increasingly being used in different business organizations to increase business efficiency and efficacy.

## REFERENCES

- [1] Definition of Data Mining  
<http://www.wdvl.com/Authoring/DB/>
- [2] HTML DOM Tutorial  
<http://www.w3schools.com/html/dom/default.asp>
- [3] <http://www.cs.cornell.edu/home/kleinber/ieee99-web.pdf>
- [4] Traversing an HTML table with DOM interfaces  
[https://developer.mozilla.org/en/traversing\\_an\\_html\\_table\\_with\\_javascript\\_and\\_dom\\_interfaces](https://developer.mozilla.org/en/traversing_an_html_table_with_javascript_and_dom_interfaces)
- [5] Web Usage Mining  
<http://maya.cs.depaul.edu/~mobasher/papers/webminer-kais.pdf>
- [6] Data Mining Within DBMS Functionality by Maciej Zakrzewicz, Poznan University.
- [7] Data Mining Concepts and Techniques By Jiawei Han and Micheline Kamber.  
<http://www.cs.uiuc.edu/~hanj/bk2/>
- [8] Data Mining by Yashwant Kanetkar.
- [9] Databases on web  
[www.ism-ournal.com/ITToday/Mining\\_Databases.pdf](http://www.ism-ournal.com/ITToday/Mining_Databases.pdf)
- [10] "Seamless Integration of DM with DBMS and Applications" by Hongjun Lu
- [11] "Mining the World Wide Web - Methods, Applications, and Perspectives".
- [12] Wiki links  
[http://en.wikipedia.org/wiki/Web\\_mining](http://en.wikipedia.org/wiki/Web_mining)



# Performance Analysis of IEEE 802.11 Non-Saturated DCF

Bhanu Prakash Battula<sup>1</sup>, R. Satya Prasad<sup>2</sup> and Mohammed Moulana<sup>3</sup>

<sup>1</sup>Asst. Professor, Dept. of CSE, Vignan's Nirula Institute of Technology and Science for Women, A.P., India.

<sup>2</sup>Professor, Acharya Nagarjuna University, Dept. of Computer Science and Engineering, A.P., India.

<sup>3</sup>Asst. Professor, Dept. of CSE, Vignan's Nirula Institute of Technology and Science for Women, A.P., India.

## Abstract

In the IEEE 802.11 MAC layer protocol, the basic access method is the Distributed Coordination Function which is based on the CSMA/CA. In this paper, we investigate the performance of IEEE 802.11 DCF in the non-saturation condition. We assume that there is a fixed number  $n$  of competing stations and packet arrival process to a station is a poisson process. We model IEEE 802.11 DCF in non-saturation mode by 3-dimensional Markov chain and derive the stationary distribution of the Markov chain by applying matrix analytic method. We obtain the probability generating function of packet service time and access delay, and throughput.

**Keywords:** DCF, Access delay, throughput.

## 1. Introduction

Recent years Wireless Local Area Networks have brought much interest to the telecommunication systems. IEEE 802.11 standards define a medium access control protocols. IEEE 802.11 MAC includes the mandatory contention-based DCF (Distributed Coordination Function) and the optional polling-based PCF (Point Coordination Function)[1]. Most of today's WLANs devices employ only the DCF because of its simplicity and efficiency for the data transmission process. The DCF employs CSMA/CA (Carrier-Sense Multiple Access with Collision Avoidance) protocol with binary exponential backoff. The DCF is relatively simple while it enables quick and cheap implementation, which is important for the wide penetration of a new technology.

We may classify arrival pattern of packets to the station into two modes: saturation mode and non-saturation mode. Saturation mode means that stations always have

packets to transmit. Non-saturation mode means that stations have sometimes no packets to transmit. Most of analytical models proposed so far for the IEEE 802.11 DCF focus on saturation performance. Unfortunately, the saturation assumption is unlikely to be valid in most real IEEE 802.11 networks. We note that most works ignore the effect of the queue at the MAC layer. There have not been many analytic works in the non-saturation mode due to mainly analytic complexity of models. The necessities of

analytic performance of IEEE 802.11 in non-saturation mode.

## 2. Overview of Medium Access Layer

Nowadays, the IEEE 802.11 WLAN technology offers the largest deployed wireless access to the Internet. This technology specifies both the Medium Access Control (MAC) and the Physical Layers (PHY) [1]. The PHY layer selects the correct modulation scheme given the channel conditions and provides the necessary bandwidth, whereas the MAC layer decides in a distributed manner on how the offered bandwidth is shared among all stations (STAs). This standard allows the same MAC layer to operate on top of one of several PHY layers.

Different analytical models and simulation studies have been elaborated the last years to evaluate the 802.11 MAC layer performance. These studies mainly aim at computing the saturation throughput of the MAC layer and focus on its improvement. One of the most promising models has been the so-called Bianchi model [2]. It provides closed form expressions for the saturation throughput and for the probability that a packet transmission fails due to collision. The modeling of the 802.11 MAC layer is an important issue for the evolution of this technology. One of the major shortcomings in existing models is that the PHY layer conditions are not considered. The existing models for 802.11 assume that all STAs have the same physical conditions at the receiving STA (same power, same coding, : :), so when two or more STAs emit a packet in the same slot time, all their packets are lost, which may not be the case in reality when for instance one STA is close to the receiving STA and the other STAs far from it [3]. This behavior, called the *capture effect*, can be analyzed by considering the spatial positions of the STAs. In [4] the spatial positions of STAs are considered for the purpose of computing the capacity of wireless networks, but only an ideal model for the MAC layer issued from the information theory is used. The main contribution of this paper is considering both PHY and MAC layer protocols to analyze

the performance of exciting IEEE 802.11 standard. Our work reuses the model for 802.11 MAC layer from [6], and extends it to consider interference from other STAs. We compute, for a given topology, the throughput of any wireless STA using the 802.11 MAC protocol with a specific PHY layer protocol. Without losing the generality of the approach, we only consider in this paper traffic flows sent from the mobile STAs in direction to the AP. The case of bidirectional traffic is a straight forward extension; we omit it to ease the exposition of our contribution. Further, we assume that all STAs use the Distributed Coordination Function (DCF) of 802.11 and they always have packets to send (case of saturated sources). We present an evaluation of our approach for 802.11b with data rates equal to 1 and 2 Mbps and the results indicate that it leads to very accurate results.

### 3. Importance Of Distributed Coordination Function (DCF)

Two forms of MAC layer have been defined in IEEE 802.11 standard specification named, Distributed Coordination Function (DCF) and Point Coordination Function (PCF). The DCF protocol uses Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) mechanism and is mandatory, while PCF is defined as an option to support time-bounded delivery of data frames. The DCF protocol in IEEE 802.11 standard defines how the medium is shared among stations. DCF which is based on CSMA/CA, consists of a basic access method and an optional channel access method with request-to-send (RTS) and clear-to-send (CTS) exchanged as shown in Fig. 1.

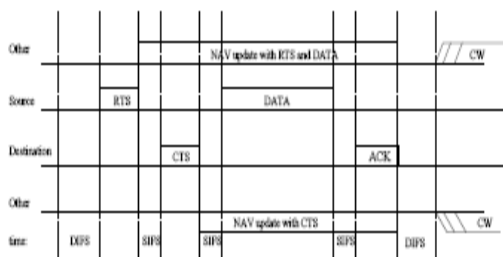


Figure 1. CSMA/CA with RTS/CTS exchange.

If the channel is busy for the source STA, a back off time (measured in slot times) is chosen randomly in the interval  $[0; CW)$ , where CW is called the contention window. This timer is decremented by one as long as the channel is sensed idle for a DIFS (Distributed Inter Frame Space) time. It stops when the channel is busy and resumes when the channel is idle again for at least DIFS time. CW is an integer with the range determined by PHY layer characteristics:  $CW_{min}$  and  $CW_{max}$ . CW will be doubled after each unsuccessful transmission, up to the maximum value which is determined by  $CW_{max} + 1$ . When the back

off timer reaches zero, the source transmits the data packet. The ACK is transmitted by the receiver immediately after a period of time called SIFS (Short Inter Frame Space) which is less than DIFS. When a data packet is transmitted, all other stations hearing this transmission adjust their Network Allocation Vector (NAV), which is used for virtual CS at the MAC layer. In optional RTS/CTS access method, an RTS frame should be transmitted by the source and the destination should accept the data transmission by sending a CTS frame prior to the transmission of actual data packet. Note that STAs in the sender's range that hear the RTS packet update their NAVs and defer their transmissions for the duration specified by the RTS. Nodes that overhear the CTS packet update their NAVs and refrain from transmitting. This way, the transmission of data packet and its corresponding ACK can proceed without interference from other nodes (hidden nodes problem).

Table 1 shows the main characteristics of the IEEE 802.11a/b/g physical layers. 802.11b radios transmit at 2.4GHz and send data up to 11 Mbps using Direct Sequence Spread Spectrum (DSSS) modulation; whereas 802.11a radios transmit at 5GHz and send data up to 54 Mbps using Orthogonal Frequency Division Multiplexing (OFDM) [1]. The IEEE 802.11g standard [1], extends the data rate of the IEEE 802.11b to 54 Mbps in an upgraded PHY layer named extended rate PHY layer (ERP).

PHY Layer Characteristic	Available in 802.11/a/b/g
Frequency	5, 2.4 GHz
Data Rates	1, 2, 5.5, 6, 9, 11, 12, 18, 22, 24, 33, 36, 48, 54 Mbps
Modulation	BPSK, DBPSK, QPSK, DQPSK, 16-QAM, 64-QAM, CCK
Error Correction Code	Convolutional codes 1/2, 2/3, 3/4

Table 1. PHY layer Characteristics in 802.11.

In each physical layer, there is a basic transmission mode (usually used to send ACK, RTS, CTS and PLCP header) which has the maximum coverage range among all transmission modes. This maximum range is obtained using BPSK or DBPSK modulations which have the minimum probability of bit error for a given SNR compared to other modulation schemes. It has the minimum data rate as well. As shown in Fig. 2, each packet may be sent using two different rates; the PLCP header is sent at the basic rate while the rest of the packet might be sent at a higher rate. The basic rate is 1 Mbps (with DBPSK modulation and CRC 16 bits) for 802.11b and 6 Mbps (with BPSK and FEC rate equal to 1/2) for 802.11a. The higher rate used to transmit the physical-layer payload (which includes the

MAC header) is indicated in the PLCP header. The PLCP Protocol Data Unit (PPDU) frame includes PLCP preamble, PLCP header, and MPDU. Fig. 3 shows the format for long preamble in 802.11b. The PLCP preamble contains the following fields: Synchronization (Sync)

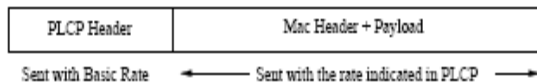


Figure 2. Packet format in IEEE 802.11.

and Start Frame Delimiter (SDF). The PLCP Header contains the following fields: Signal, Service, Length, and CRC. The short PLCP preamble and header may be used to minimize overhead and thus maximize the network data throughput. Note that the short PLCP header uses the 2 Mbps with DQPSK modulation and a transmitter using the short PLCP only can interoperate with the receivers which are capable of receiving this short PLCP format. In this paper we suppose that all stations use the long PPDU format in 802.11b. We evaluate our model in 802.11b where STAs use transmission rate equal to 1 and 2 Mbps. Our model can be employed for all other transmission modes for all standards if the packet error rate is calculated.

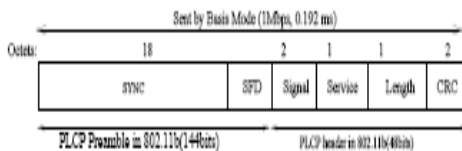


Figure 3. 802.11b long preamble frame format.

In this paper, we assume that the noise over the wireless channel is white Gaussian with spectral density equal to  $N_0=2$ . In our model we define  $N_0$  as the power of the thermal noise,

## References

[1] Wireless LAN MAC and PHY layer specifications, LAN MAN Standards Committee of the IEEE Computer Society Std., ANSI/IEEE 802.11, 1999.  
 [2] E. Winands, T.Denteneer, J.Resing, and R. Rietman, "A finite-source feedback queueing network as a model for the IEEE 802.11 DCF", in European Transactions On Telecommunications, Vol 16, 2005, pp. 77-89.  
 [3] LI BO and Roberto Battiti, "Performance analysis of an Enhanced IEEE 802.11 Distributed Coordination Function Supporting Service Differentiation", in International Workshop on Quality of Future Internet Service

$$N_o = N_f \cdot N_t = N_f \cdot kTW \quad (1)$$

where  $N_f$  denotes the circuit noise value,  $k$  the Boltzmann constant,  $T$  the temperature in Kelvin and  $W$  is the frequency bandwidth. For the BPSK modulation, the bit error probability is given:

$$P_b^{BPSK} = Q \left( \sqrt{2 \cdot \frac{E_b}{N_o}} \right) = Q \left( \sqrt{2 \cdot \frac{E_b}{N_o}} \right) \quad (2)$$

and for QPSK (4-QAM) is:

$$P_b^{QPSK} = Q \left( \sqrt{2 \cdot \frac{E_b}{N_o}} \right) - \frac{1}{2} Q^2 \left( \sqrt{2 \cdot \frac{E_b}{N_o}} \right) \quad (3)$$

## 4. Conclusion

There have been various attempts to model and analyze the saturation throughput and delay of the IEEE 802.11 DCF protocol since the standards have been proposed. As explained in the introduction there is different analytical models and simulation studies that analyze the performance of 802.11 MAC layer. As an example Foh and Zuckerman present the analysis of the mean packet delay at different throughputs for IEEE 802.11 MAC. Kim and Hou analyze the protocol capacity of IEEE 802.11MAC with the assumption that the number of active stations having packets ready for transmission is large. They have suggested some extensions to the model proposed to evaluate packet delay, the packet drop probability and the packet drop time. Since in our model we have used the Bianchi's model and its extension proposed.

QoFIS 2003, Sweden, Springer Lecture Notes on Computer Science LNCS volume 2811, 2004, pp. 152-161.  
 [4] LI BO and Roberto Battiti, "Achieving Maximum Throughput and Service Differentiation by Enhancing the IEEE 802.11 MAC Protocol", in WONS 2004, Springer Lecture Notes on Computer Science, Vol. 2928, pp. 285-301.  
 [5] IEEE 802.11 WG, part 11a/11b/11g, "Wireless LAN Medium Access Control (MAC) and Physical (PHY) specifications", Standard Specification, IEEE, 1999.  
 [6] Giuseppe Bianchi, "Performance Analysis of the IEEE 802.11 Distributed Coordination Function", IEEE Journal on Selected Areas in Communications, Vol. 18, Number 3, March 2000.

- [7] A. Kochut, A. Vasan, A. U. Shankar, A. Agrawala, "Sniffing out the correct Physical Layer Capture model in 802.11b", Proceeding of ICNP 2004, Berlin, Oct. 2004.
- [8] P. Gupta, P. R. Kumar "The Capacity of Wireless Networks", IEEE Transactions on Information Theory, Vol. 46, No. 2, March 2000.
- [9] C. H. Foh, M. Zukerman, "Performance Analysis of the IEEE 802.11 MAC Protocol", Proceedings of the EW 2002 Conference, Italy.
- [10] H. Wu, Y. Peng, K. Long, J. Ma, "Performance of Reliable Transport Protocol over IEEE 802.11 Wireless LAN: Analysis and Enhancement", Proc. of IEEE INFOCOM, vol.2, pp. 599-607, 2002.

## Authors profile

**Bhanu Prakash Battula** received Master's Engineering degree on Computer Science & Technology in 2008 from Acharya Nagarjuna University and also received another Master's degree on Computer Applications from Acharya Nagarjuna University. After Post graduation, He is working as a Asst.Professor in the Department of Computer Science and Engineering at Vignan's Nirula Institute of Technology and Science, Guntur, Andhra Pradesh. He published papers for International Journals. His research interests include Computer Security, Steganalysis and Image Processing.

**R.Satya Prasad** received PhD from Acharya Nagarjuna University in 2007. He is working as a Professor at Department of Science and Engineering, Acharya Nagarjuna University. He Published more than 15 National and International publications and his research interests include Computer Security, Software reliability and Image Processing

**Mohammed Moulana** received the Master's degree M.Sc Mathematics from Acharya Nagarjuna University, in 2004. He received M.Phil in Mathematics from Alagappa University in 2007. He received M.Tech Computer Science & Engineering in 2009 from JNTUK, Kakinada, A.P., India. He is currently an Asst. Professor, Department of Computer Science & Engineering, Vignan's Nirula Institute of Technology and Science for Women, Guntur Dist, A.P., India. His Area of Interest are Communication Networks, Wireless LANs & Ad-Hoc Networks.

# Enhancing the Capability of N-Dimension Self-Organizing Petrinet using Neuro-Genetic Approach

Manuj Darbari<sup>#</sup>, Rishi Asthana<sup>\*</sup>, Hasan Ahmed<sup>#</sup> Neelu Jyoti Ahuja<sup>#</sup>

<sup>#</sup>Department of Electrical and Information Technology,  
Babu Banarasi Das University,  
Lucknow, India.

<sup>\*</sup>Department of Computer Science,  
University of Petroleum and Energy Studies, Dehradun, India.

**Abstract**— The paper highlight intelligent Urban Traffic control using Neuro-Genetic Petrinet. The combination of genetic algorithm provides dynamic change of weight for faster learning and converging of Neuro-Petrinet.

**Keywords**— Neuro Petrinet, Urban Traffic Systems, Genetic Algorithm.

## I. INTRODUCTION

The previous models like developed for vehicular studies only considered a limited macro mobility, involving restricted vehicle movements, while little or no attention was paid to micro - mobility and its interaction. The research community could not provide the realistic environment[6] for modeling Urban Traffic which could simulate close to real time situations. Our papers extend the concept of Li, M and Change works of August oriented urban Traffic simulation using interaction agent in controlling and management of urban traffic systems. We use the concept of Neuro Genetic Networks on self organizing Petrinet to simulate the traffic condition.

## II. LITERATURE SURVEY

The dynamics of Urban traffic System[4] was observed by Tzes, Kim and Mc Shane[8] which explains about the timing plans of the traffic controlling junctions. While an example of coloured petrinet modeling of traffic light was proposed by Jenson [5]. Later on Darbari[2] and Medhavi also developed Traffic light control by Petrinet.

### A. Petrinet

Most recently List and Cetin [7] discussed the use of PNs in modeling traffic signal controls and perform a structural analysis of the control PN model by P-invariants, demonstrating how such a model enforces traffic operations safety.

List and Cetin [7] proposed different colour scheme to each vehicle entering the system, they modelled it by defining appropriate subnets modeling links at the intersections.

## III. BASICS OF PETRINET MODEL APPLICATION IN URBAN TRAFFIC MODELING

To start with , we described a simple pattern of PN using event relationship diagram. It shows that event e1 can cause event e2 within a time period [I1, I2] where T represents Transition.



Figure 1 : Simple Petrinet Representation

### A. Dynamics of Producer

Consumer Petrinet with the algorithm for Dynamic Producer-Consumer given as:

Step 1 : Initialise each of the Producers- Consumer situation (x). set the pattern rate as 'r'.

Step 2 : Set the control centre such that :

$$X_i = S_i$$

Step 3 : Let the Token release rate be given as 1/N, where N is defined as the number of producer - consumer initial states.

Step 4 : The release of Token are updated as :

$$x : (\text{producer} - \text{old}) = x; (\text{producer} - \text{new state}) + r$$

Step 5 : Stop when system has transferred all the tokens and traffic reaches a balancing state.

Assuming the initializing condition to be  $X_i$  and after successive training it reaches to 9. The stabilising condition is reached after 'n' iteration given as :

$$\{x; = t(x_1, \dots, x_n) \mid I \in \{1, \dots, n\}$$

if there are N - dimensional node the equation will become.

$$x_i = t(x_1, \dots, x_n)$$

.

..

$$x_n = t_n(x_1, \dots, x_n)$$

$x_i$  represents the recursion variables and  $t(x_1, \dots, x_n)$  shows the process terms with possible occurrence of the recursion variables.

#### IV. N- DIMENSIONAL SELF ORGANISED PETRINET MODEL OF URBAN TRAFFIC SYSTEM

Let '0' and '1' be defined as Low and High learning rate of the grid network of petrinet showing the simulation of Traffic in a mesh network.

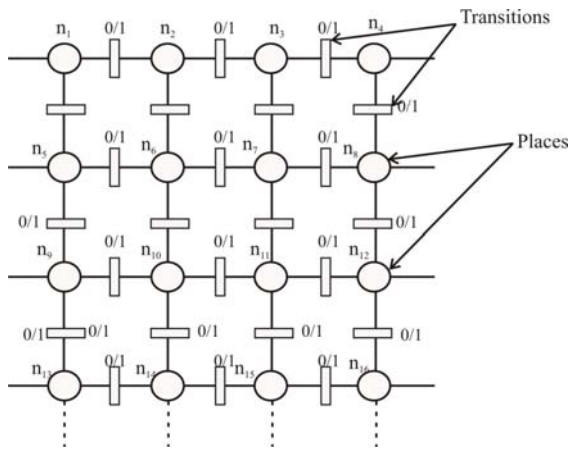


Figure 2 : Process Representation of Petrinet with 0/1 Learning rate in N-Dimensions.

We can define the movement of tokens and 0/1 learning rate by a single recursive equation as:

$$S_i = in(0) (X \parallel out(0)) + in(1) (X \parallel out(1)) \quad (1)$$

The process graph of Neural Petrinet Framework represents bisimilar relationship in Recursive mode. The Recursion can be achieved by using Genetic Algorithm[1] with learning rate of nodes  $[n_1, \dots, n_n]$  with a particular time frame.

Let the total function be defined as:

$$n_t = \int (n_{t-1}, n_{t-2}, n_{t-3}, \dots, n_{t-M}) \quad (2)$$

Which predicts the current value of node  $n^{\text{th}}$  from past input conditions.

The Nodes which will learn first will survive and based on them the traffic control network will converge. The Fitness[3] value (F.V.) is defined as :

$$\left[ F.V. = \frac{1}{N} \sum_{i=1}^N (f(n_i) - f(n_i))^2 \right] \quad (3)$$

Where:

$f(n_i)$  is the value of the function represented by GP individual (Geno-type). The algorithm for Procedure Trained Node selection by Genetic Algorithm[9,10] :

**BEGIN**

*Set the values to initial trained conditions.*

*Generate as many nodes as possible;*

*Evaluate each node in the set of Nodes selected;*

*WHILE termination NOT coverage DO*

**BEGIN**

*Select the Cube of Nodes with faster learning rates;*

*Generate offspring cube of Nodes by applying crossover and mutation on the selected cube or nearest neighbour cube;*

*Evaluate the equilibrium condition;*

*Generate new nodes to be trained further in combination with older node cube;*

**END**

*Return the best trained Cube Node from the Mesh;*

**END**

The first phase of the Algorithm deals with controlling parameters, such the population of Cube (P) and Offspring (O), the maximum number of crossover probability and mutations are set. The offspring Cube of Node is then evaluated and traffic is stabilized accordingly completing one cycle of operations. After several iterations the entire control Network converges to optimal solutions

#### V. CONCLUSIONS

The paper represents the dynamic control strategy of Urban Traffic System by combining Neuro-genetic approach on Petrinets. The use of genetic learning method performs rule discovery of larger system with rules fed into a conventional system. The main idea to use genetic algorithms with neural network is to use a genetic algorithm to search for the appropriate weight change in neural network which optimizes the learning rate of the entire network.

A good Genetic Algorithm can significantly reduce neuro-Petrinet in aligning with the traffic conditions, which other wise is a very complex issue.

#### REFERENCES

- [1] Baker, B.M. (2003). "A genetic algorithm for the Vehicle Routing Problem", Computers and operations Research, Vol. 30.
- [2] Darbari, M , Medhavi, S, "N-Dimensional Self Organizing Petrinet for Urban Traffic Modeling " IJCSI, Issue 4, No.2.
- [3] Deng, P.S., (2000), "Coupling Genetic Algorithm and Rule Based Systems for Complex Decisions", Expert Systems with Applications, Vol. 19, No. 3.
- [4] Grupe , F.H. (1998), "The Applications of Case - Based Reasoning to the Software Development Process," Information and Software Technology, Vol. 40, No. 9.



- [5] Jensen, K. (1992). "*Colored Petri nets: basic concepts, analysis methods and Practical use*", Vol. 1. New York: Springer.
- [6] Miller T.W. (2005), "Data and Text Mining: A Business Application Approach", Prentice Hall.
- [7] List G. F, Cetin M (2004), "Modeling Traffic Signal Control Using Petrinets", IEEE Transaction on Intelligent Transportation Systems, 5(3), 177-187.
- [8] Tzes, A., Kim, S., & McShane, W. R. (1996). Applications of Petri networks to transportation network modeling. *IEEE Transactions on Vehicular Technology*, 45(2), 391-400.
- [9] Wang Y. (2003), "Using Genetic Algorithm Models to Solve course scheduling Problems", Expert systems and applications, Vol. 25, No. 1.
- [10] Walbridge C.T., (1989), "Genetic Algorithms : What Computers can learn from Darwin", Technology Review.

## Vulnerabilities of Electronics Communication: solution mechanism through script

<sup>1</sup>Arun Kumar Singh

<sup>1,3,4</sup>Department of Computer Science and Engineering Motilal Nehru National Institute of Technology, Allahabad, Uttar Pradesh, 211004 India,

<sup>2</sup>Pooja Tewari

Computer Science and Engineering, I.M.S. Engineering College, Ghaziabad,

<sup>3</sup>Shefalika Ghosh Samaddar

<sup>4</sup>Arun K. Misra

### Abstract

*World trade and related business ventures are more or less dependent on communication. Information content of communication is to be protected as mis-communication or incorrect information may ruin any business prospect. Communication using Internet or any other electronic communication is having various kinds of threat and vulnerability. Information should be packaged for communication in such a way that these vulnerabilities are reduced to a minimum. With the increased use of networked computers for critical systems, network security is attracting increasing attention. This paper focuses on the most common attacks to paralyze computer and network resources, in order to stop essential communication services. The paper provides methods, ways and means for obtaining network traces of malicious traffic and strategies for providing countermeasures. Analysis of packet captured in a network traffic is a common method of deletion of countermeasure of communication based vulnerabilities. Analysis of http based network traffic allows to intercept sensitive information such as the user's name and password. The ideal approach for secured communication is to remove all security flaws from individual hosts. A tradeoff between overheads (computational and business) and efficiency of securing mechanism of communication may be achieved by using the script based solutions. This paper presents the communication based vulnerabilities and their script based solution.*

*Keywords: Computer Security, Network Security, Internet Security, Cryptography, Vulnerability, Firewalls, Attackers, Network Attacks*

## 1. Introduction

With the advent of more and more open systems, intranets, and the Internet, information systems and

need to assess and manage potential security risks on their network users are becoming increasingly aware of the networks and systems. Vulnerability assessment is the process of measuring and prioritizing these risks associated with network, host based systems and devices. A rational planning of technologies and activities will be able to manage business risk to a considerable extent. These tools allow customization of security measures, automated analysis of vulnerabilities, and creation of reports that effectively communicate security vulnerability. Detailed corrective actions to all levels of an organization may be automated.

The primary sources of information for vulnerable systems are network log data and system activity. Network-based systems look for specific patterns in a network traffic and host-based systems look for those patterns in log generated files. In general, network-based vulnerability can detect attacks that host-based systems can miss because they examine packet headers and the content of the payload, looking for commands or syntax used in specific attacks.

### 1.1 Vulnerability Assessment

Vulnerability assessment in a communication aims at identifying weaknesses and vulnerabilities in a system's design, implementation, or operation and management, which could be exploited to violate the system's security. The overall scope of vulnerability assessment is to improve information and system security by assessing the risks associated. Vulnerability assessment will set the guidelines to stop or mitigate any risk.

This paper focuses on a technical vulnerability assessment methodology, giving an exposure of the threats and vulnerabilities. Major Internet-based security issues and network threats are covered. Threats and their management requires performing assessment exercise.

#### 1.1.1 Host Based Vulnerability Assessment

---

<sup>1,3,4</sup>The first, second and third Authors are thankful to Information Security Education & Awareness Project (ISEA) of MCIT department of Information Technology, Govt. of India for the partial support to the research conducted.

Vulnerability Assessment is to identify what systems are “alive” within the network ranges for host based threats and what services they offer. Identifying the location of the establishment and cataloging its services are the two main elements of Vulnerability assessment. Assessment of vulnerability may lead to the deletion of a number of viruses, worms and Trojan horses.

A virus is a package of code that attaches itself to a host program and propagates when the infected program is executed in an indirect mode along with some other essential programs. Attracting a virus to system programs or commands is an easy way of propagating of the viruses. Thus, a virus is self-replicating and self-executing. Viruses are transmitted when included as part of files downloaded from the Internet or as e-mail attachments. Worms are independent programs that replicate by copying themselves from one system to another, usually over a network or through e-mail attachments. Many modern worms also contain virus code that can damage data or consume system resources that they render the operating system unusable.

A Trojan horse program (also known as a “back door” program) acts as a stealth server that allows intruders to take control of a remote computer without the owner’s knowledge. Greek mythical Trojan horses are analogous in attributes which these digital Trojan horses possess. These programs typically masquerade as benign programs and rely on gullible users to install them. Computers that have been taken over by a Trojan horse program are sometimes referred to as zombies. Armies of these zombies can be used to launch crippling attacks against Web sites.

Communication based vulnerability are a real time threats to computer’s security. Those may take the form of physical attacks, pilfered passwords, nosy network neighbors and viruses, worms, and other hostile programs. A number of manifestations of such vulnerability are seen these days e.g. Denial of service (DoS) attacks.

A denial-of-service (DoS) attack hogs or overwhelms a system’s resources so that it cannot respond to service requests. A DoS attack can be effected by flooding a server with so many simultaneous connection requests that it cannot respond. Another approach would be to transfer huge files to a system’s hard drive, exhausting all its storage space. A related attack is the distributed denial-of-service (DDoS) [ 1 ].

The Security Threat and the Response attack, is also an attack on a network’s resources. It is launched from a large number of other host machines. Attack software is installed on these host computers, unbeknownst to their

owners, and then activated simultaneously to launch communications to the target machine of a magnitude as to overwhelm the target machine.

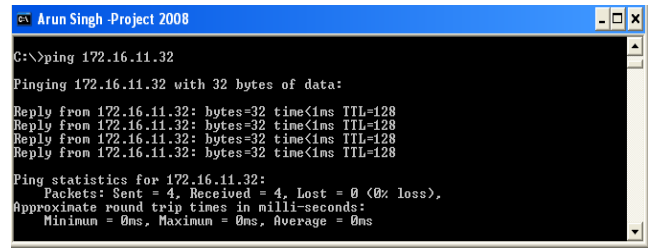


Figure –1 Ping command to check system is alive or not

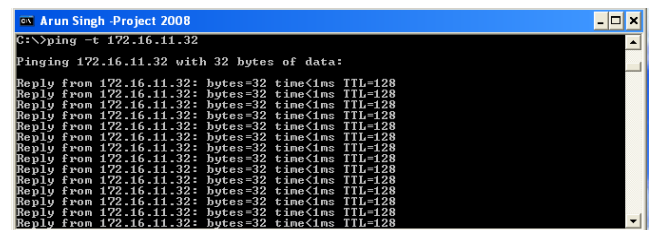


Figure –2 DoS Attack

Ping of Death is another flavour (Figure-1, Figure-2) of DDoS. Smurf Attack involves using IP spoofing and the ICMP to saturate a target network with traffic. It is then equivalent to launching a DoS attack. It consists of three elements: the source site, the bounce site, and the target site. The attacker (the source site) sends a spoofed ping packet to the broadcast address of a large network (the bounce site). This packet modified by the intruder contains the address of the target site. This causes the bounce site to broadcast the misinformation to all of the devices on its local network. All of these devices now respond with a reply to the target system, which is then saturated with those replies.

Spam is another malicious formulation in the arena of cyber crime. Responses to spam may lead to huge financial and material loss. Spam has the format of a e-mail message that are pushed to e-mail clients without their solicitation.

## 2.0 Related Work

Vulnerability assessment process is comprised of four phases, namely discovery, detection, exploitation, and analysis/recommendations [2]. Figure 3 identifies the relationships among the four phases, and the flow of information into the final report.

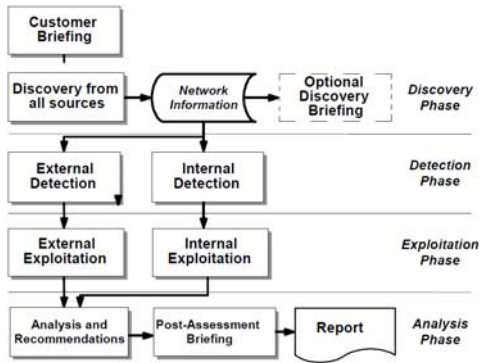


Figure-3 Vulnerability Assessment Process

[source: <http://www.oisssg.org/wiki/images/4/4b/Image001.png>]

Protocol based attack/Packet based attack has been studied from the very beginning of the study of security and related vulnerabilities. With rapid growth in both the number and sophistication of cyber attacks, it has become imperative that cyber defenders be equipped with highly effective tools that identify security Vulnerabilities before they are exploited [3]. Vulnerability can be defined as a set of conditions which if true, can leave a system open for intrusion, unauthorized access, denied availability of services running on the system or in any way violate the security policies of the system set earlier.

A breach of security occurs when a stated organizational policy or legal requirement regarding information security, has been contravened. However, every incident which suggests that the confidentiality, integrity and availability of the information has been inappropriately changed, can be considered a security vulnerability. Every security breach is always initiated via security vulnerability, only if confirmed does it become a security breach [4].

A denial of service (DoS) attack is a malicious attempt by one or many users to limit or completely disable the availability of a service. They cost businesses millions of pounds each year and are a serious threat to any system or network. These costs are related to system downtime, lost revenues, and the labour involved in identifying and reacting to such attacks [5]. DoS attacks were theorized years ago, before the mass adoption of current Internet protocols [6].

DoS is still a major problem today and the Internet remains a fragile place [6]. A large number of known vulnerabilities in network software and protocols exist; relating DoS. Sending enough data to consume all available network bandwidth (Bandwidth Consumption) is a DoS attack. Sending data in such a way as to consume a resource needed by the service (Resource Starvation) is another DoS attack. Exercising a software.bug. causing the software running the service to fail (Programming

Flaws) is the other type of the attack. Malicious use of the Domain Name Service (DNS) and Internet routing protocols leads to DoS. Many DoS attacks exploit inherent weaknesses in core Internet protocols. This makes them practically impossible to prevent, since the protocols are embedded in the underlying network technology and adopted as standards worldwide. Today, even the best countermeasure software can only provide a limiting effect on the severity of an attack [ 7]. An ideal solution to DoS will require changes in the security and authentication of these protocols [6].

In order to launch some DoS attacks, the programmer must be able to form raw packets. Using raw packets, the header information and data can be manipulated to form any kind of packet sequence. Hence techniques such as IP Spoofing and malformed ICMP Ping requests can be used [18]. This report will investigate the mechanism of DoS attacks and their countermeasures. Distributed denial of service attacks will also be investigated. A distributed DoS generally has the same effect as a single attack, with the disruption amplified by many systems acting together. These other systems are often compromised machines remotely controlled by the hacker [8].

With the rapid development of more complex systems, the chance of introduction of errors, faults and failures increases in many stages of software development life-cycle [9]. This class of system failures is commonly termed as software vulnerabilities. These security vulnerabilities violate security policies and can cause the system to be compromised leading to loss of information . Vulnerabilities can be introduced in a host system in different ways; via errors in the code of installed software, mis-configurations of the software settings that leave systems less secure than they should be (improperly secured accounts, running of necessary services, etc)

Network based vulnerability assessment gathers information of the system and services attached to the network and identifies weakness and vulnerabilities exploitable in the network. These vulnerabilities could be related to services, such as HTTP, FTP and SMTP protocol, running on the given network. A network-based scanning assessment may also detect extremely critical vulnerabilities such as mis-configured firewalls or vulnerable web servers in a De-Militarized Zone (DMZ), which could provide a security hole to an intruder, allowing them to compromise an organizations security [10]. Network assessment tools gather information and may also have network mapping and port scanning abilities [2]. The tools use for such purpose are Nmap etc. [2].

### 3.0 Design of the solution

Host-based vulnerability analysis has been taken up for design of solution along with a lot of potential for further research and development in many other fields including the field of vulnerability analysis. Plugging of the vulnerability is ensured by designing script based and command based codes sniffing a HTTP packet is shown in figure 4. Capturing a HTTP based e-mail password is shown in figure 5.

Sniffing HTTP packet and its result in figure 4 are roles worthy. Capturing a HTTP based mail Password in figure 5 is equally important from the point of view of vulnerabilities. The packet list pane shows that the HTTP protocol packets are being transmitted from source IP 172.31.132.59 to destination IP 172.31.100.29. The packets are being captured while transmitting from one mode to other. This particular packet gives the information that HTTP mail of this website *http://mail.mniti.ac.in* has been logged in by the source IP and its corresponding username and password are also captured under the heading of line-based text data in packet detail pane (figure-5).

```
Referer: http://mail.mniti.ac.in/webmail/src/login.php\r\n
Cookie: sqidentity=pooja; sqghash=aQfja2Vj; squirrelmail_language=en_US; SQMSESSID=5i5qki32
Proxy-Authorization: Basic YXJlbmtzZW5naDphcnVuc2luZ2gx\r\n
  Credentials: arunksingh:arunsingh1
Content-Type: application/x-www-form-urlencoded\r\n
Content-Length: 91
\r\n
Line-based text data: application/x-www-form-urlencoded
login_username=arun&secretkey=cracker&js_autodetect_results=1&just_logged_in=1&button=login
00 30 33 09 33 71 00 09 33 32 00 32 00 32 39 00 30
90 63 71 6a 6f 32 32 33 35 6f 67 36 0d 0a 50 72 6f
a0 78 79 2d 41 75 74 68 6f 72 69 7a 61 74 69 6f 6e
b0 3a 20 42 61 73 69 63 20 59 58 4a 31 62 6d 74 7a
c0 61 57 35 6e 61 44 70 68 63 6e 56 75 63 32 6c 75
d0 5a 32 67 78 0d 0a 43 6f 6e 74 65 6e 74 2d 54 79
e0 70 65 3a 20 61 70 70 6c 69 63 61 74 69 6f 6e 2f
f0 78 2d 77 77 72 66 6f 72 6d 2d 75 72 6c 65 6e
00 63 6f 64 65 64 0d 0a 43 6f 6e 74 65 6e 74 2d 4c
10 65 6e 67 74 68 3a 20 39 31 0d 0a 0d 0a 6c 6f 67
20 69 6e 5f 75 73 65 72 6e 61 6d 65 3d 61 72 75 6e
30 26 73 65 63 72 65 74 6b 65 79 3d 63 72 61 63 6b
40 65 72 26 6a 73 5f 61 75 74 6f 64 65 74 65 63 74
50 5f 72 65 73 75 6c 74 73 3d 31 26 6a 75 73 74 5f
60 6c 6f 67 65 64 5f 69 6e 3d 31 26 62 75 74 74
70 6f 6e 3d 6c 6f 67 69 6e
```

Figure-4 Capturing a HTTP based mail Password

Figure 6 shows that the username is *arun* and password is *cracker* which is given next to secret key. This is also shown in packet bytes pane in the right hand side of HEX numbers (Figure 4). Sometimes, when the password of a

user contains some special characters, they are written using special character that appears in the pane.

```
1 0.00000 172.31.100.29 172.31.132.59 HTTP Continuation or non-HTTP traffic
2 0.00010 172.31.132.59 172.31.100.29 TCP 33043 > ndf.aaa [ACK] Seq=1 Ack=54 Win=156 Len=0 TSV=3767564 TSN=937707469
3 0.675544 IntelCor_42f66c7 Broadcast ARP Who has 172.31.133.76? Tell 172.31.133.147
4 0.621166 Class_03c7a15 Spanning-tree (for-bf STP) Conf. Root = 32768/00:01:fa:6b:02:c8 Cost = 27 Port = 0a015
5 2.859502 Class_03c7a15 Spanning-tree (for-bf STP) Conf. Root = 32768/00:01:fa:6b:02:c8 Cost = 27 Port = 0a015
6 3.524834 172.31.132.59 172.31.100.29 TCP 4007 > ndf.aaa [FIN] Seq=0 Win=584 Len=0 MSS=1460 TSV=3768445 TSN=0 MS7
7 3.524796 172.31.132.59 172.31.132.59 TCP ndf.aaa > 4007 [FIN, ACK] Seq=1 Ack=1 Win=576 Len=0 MSS=1460 TSV=3768099 TSN=0
8 3.524802 172.31.132.59 172.31.100.29 TCP 4007 > ndf.aaa [ACK] Seq=1 Ack=1 Win=588 Len=0 TSV=3768445 TSN=3768099
9 3.524800 172.31.132.59 172.31.100.29 HTTP POST http://mail.mniti.ac.in/webmail/src/receive.php HTTP/1.1 (application/
10 3.529553 172.31.100.29 172.31.132.59 TCP ndf.aaa > 4007 [ACK] Seq=1 Ack=823 Win=752 Len=0 TSV=303780095 TSN=37684
11 3.477289 IntelCor_42f66c7 Broadcast ARP Who has 172.31.133.76? Tell 172.31.133.147
12 4.281473 172.31.133.147 172.31.135.255 NMG Name query NB SQLSERVER.NET:00
13 4.281475 172.31.133.147 172.31.135.255 NMG Name query NB CIFS\CORP.MNITI.COM:00
14 4.281486 172.31.133.147 172.31.135.255 NMG Name query NB WGSYS.MNITI.COM:00
15 4.281488 172.31.133.147 172.31.135.255 NMG Name query NB WGSYS.MNITI.COM:00
Referer: http://mail.mniti.ac.in/webmail/src/login.php\r\n
Cookie: sqidentity=pooja; sqghash=aQfja2Vj; squirrelmail_language=en_US; SQMSESSID=5i5qki32&js_autodetect_results=1&just_logged_in=1&button=login
Proxy-Authorization: Basic YXJlbmtzZW5naDphcnVuc2luZ2gx\r\n
  Credentials: arunksingh:arunsingh1
Content-Type: application/x-www-form-urlencoded\r\n
Content-Length: 91
\r\n
Line-based text data: application/x-www-form-urlencoded
login_username=arun&secretkey=cracker&js_autodetect_results=1&just_logged_in=1&button=login
0 65 6e 67 74 68 3a 20 39 31 0d 0a 0d 0a 6c 6f 67
1 69 6e 5f 75 73 65 72 6e 61 6d 65 3d 61 72 75 6e
2 26 73 65 63 72 65 74 6b 65 79 3d 63 72 61 63 6b
3 65 72 26 6a 73 5f 61 75 74 6f 64 65 74 65 63 74
4 5f 72 65 73 75 6c 74 73 3d 31 26 6a 75 73 74 5f
5 6c 6f 67 65 64 5f 69 6e 3d 31 26 62 75 74 74
6 f 6e 3d 6c 6f 67 69 6e
```

Figure-5 Proxy Authorization

There are a number of tools available for such purpose. Wireshark is able to sniff the proxy password as illustrated in figure-5. This is done in the same way as capturing of username and password of a mail user as shown in figure 6. Proxy password is also obtained in packets detail pane under the Proxy-Authorisaton. In this figure, proxy username is 'arunksingh' and password is 'arunsingh1' which is shown next to Credentials. This is how sniffing is being done over HTTP connection in LAN.

```
Referer: http://mail.mniti.ac.in/webmail/src/login.php\r\n
Cookie: sqidentity=pooja; sqghash=aQfja2Vj; squirrelmail_language=en_US; SQMSESSID=5
Proxy-Authorization: Basic YXJlbmtzZW5naDphcnVuc2luZ2gx\r\n
  Credentials: arunksingh:arunsingh1
Content-Type: application/x-www-form-urlencoded\r\n
Content-Length: 91
\r\n
Line-based text data: application/x-www-form-urlencoded
login_username=arun&secretkey=cracker&js_autodetect_results=1&just_logged_in=1&button=login
0250 61 47 46 6a 61 32 56 79 38 2d 73 71 75 69 72 72
0260 65 6c 6d 61 69 6c 5f 6c 61 6e 67 75 61 67 65 3d
0270 65 6e 5f 55 53 3b 20 53 51 4d 53 45 53 53 49 44
0280 3d 35 69 39 71 68 69 39 32 6b 32 6b 32 39 6b 38
0290 63 71 6a 6f 32 32 39 35 6f 67 36 0d 0a 50 72 6f
02a0 78 79 2d 41 75 74 68 6f 72 69 7a 61 74 69 6f 6e
02b0 3a 20 42 61 73 69 63 20 59 58 4a 31 62 6d 74 7a
02c0 61 57 35 6e 61 44 70 68 63 6e 56 75 63 32 6c 75
02d0 5a 32 67 78 0d 0a 43 6f 6e 74 65 6e 74 2d 54 79
02e0 70 65 3a 20 61 70 70 6c 69 63 61 74 69 6f 6e 2f
02f0 78 2d 77 77 72 66 6f 72 6d 2d 75 72 6c 65 6e
0300 63 6f 64 65 64 0d 0a 43 6f 6e 74 65 6e 74 2d 4c
0310 65 6e 67 74 68 3a 20 39 31 0d 0a 0d 0a 6c 6f 67
0320 69 6e 5f 75 73 65 72 6e 61 6d 65 3d 61 72 75 6e
0330 26 73 65 63 72 65 74 6b 65 79 3d 63 72 61 63 6b
0340 65 72 26 6a 73 5f 61 75 74 6f 64 65 74 65 63 74
0350 5f 72 65 73 75 6c 74 73 3d 31 26 6a 75 73 74 5f
0360 6c 6f 67 65 64 5f 69 6e 3d 31 26 62 75 74 74
0370 6f 6e 3d 6c 6f 67 69 6e
```

Figure-6 Capturing the Content of Message sites

Wireshark is able to capture the username and password of mail user in the same way it does for message websites like [www.160by2.com](http://www.160by2.com) or [www.way2sms.com](http://www.way2sms.com). Figure-6 shows the capturing of a message packet being sent from the message website [www.160by2.com](http://www.160by2.com) as shown in figure-6. This figure shows that the user whose IP address



is 172.31.132.59 when logs the message website, the packet is sent to the destination IP address 172.31.100.29 which capture the HTTP packet and the corresponding information to this is given in info 'POST' as <http://www.160by2.com/logincheck>.

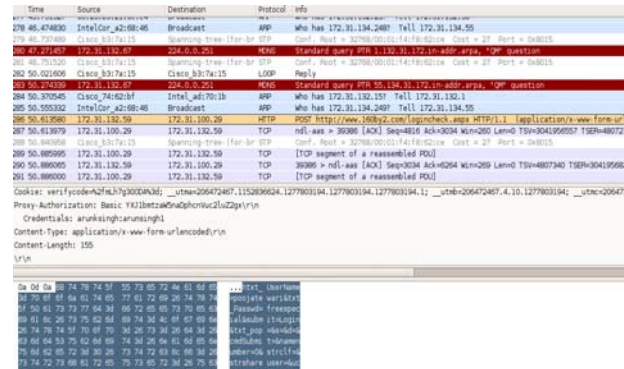


Figure-7 Message Captured by Wireshark

Sending a secret message to anyone by these types of message sites has its own liabilities because Wireshark can easily capture his message. Example of this sent message is as shown in figure-8. This captured packet is analyzed by TCP stream.



Figure-8 Content of the message

It shows the content of a message is seen clearly and also the contact number of the person to whom it has been sent. The message content written next to the text message heading is *hi+ dear+ hw+ r+ u*. This is the original message content as *hi dear hw r u* was being sent from this website.

TShark is a network protocol analyzer and a command-line version of Wireshark, which captures the live packet data from a live network, or read packets from a previously saved capture file. By default, tshark prints the summary line information to the screen. This is the same information contained in the top pane of the Wireshark

GUI. The default tshark output is shown below in figure-9.

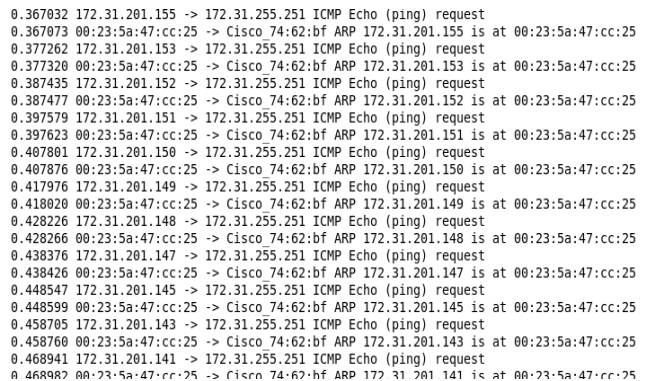


Figure-9 Capturing Password by Tshark

This paper is focused on the data communication over the HTTP connections in LAN, which are not secure and important information maybe sniffed in the form of packet when passing through multiple stations to a destined one. In figure-9, it is illustrated that when the user logs the message website, then his password can be sniffed as shown in the right hand side of the column in the last 9th line. The username is 'poojatewari' and password is 'passhacked' when the user logs the message website (figure 10).

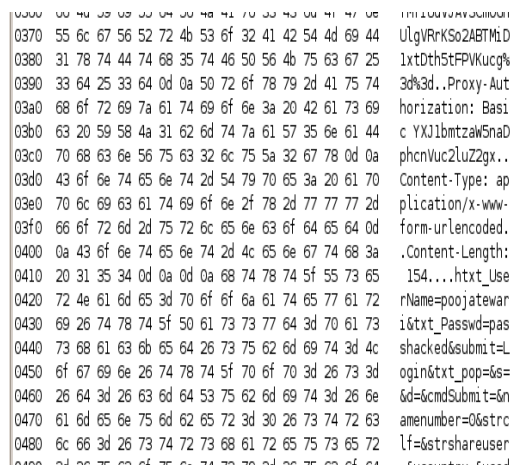


Figure-10 Capturing the data of a Message Website

Tshark can also capture the sent message from a message website like [www.160by2.com](http://www.160by2.com) or [www.way2sms.com](http://www.way2sms.com). Capturing of the sent message from [www.160by2.com](http://www.160by2.com) is illustrated in figure 11.

When the source IP address 172.31.132.59 sends a packet containing the data content to the destination IP address 172.31.100.14, it can be sniffed as shown in the figure. It



shows the contact number and the message sent to that contact. Here, the captured content is *hello++ hwz+ u* next to text message heading as in Wireshark. This original message content sent is “*hello hwz u*” sent from this website.

```

-----
) 30 6a 44 6c 4e 31 39 32 77 45 72 78 4d 65 4f 4f 0jDlN192wErXMe00
) 6f 68 31 47 51 74 35 61 56 42 47 5a 64 54 25 32 oh1GQt5aVBGDZt%2
) 62 77 25 33 64 25 33 64 0d 0a 50 72 6f 78 79 2d bw%3d%3d..Proxy-
) 41 75 74 68 6f 72 69 7a 61 74 69 6f 6e 3a 20 42 Authorization: B
) 61 73 69 63 20 59 58 4a 31 62 6d 74 7a 61 57 35 asic YXJlbmtzaW5
) 6e 61 44 70 68 63 6e 56 75 63 32 6c 75 5a 32 67 naDphcnVuc2luZ2g
) 78 0d 0a 43 6f 6e 74 65 6e 74 2d 54 79 70 65 3a x..Content-Type:
) 20 61 70 70 6c 69 63 61 74 69 6f 6e 2f 78 2d 77 application/x-w
) 77 77 2d 66 6f 72 6d 2d 75 72 6c 65 6e 63 6f 64 ww-form-urlencoded
) 65 64 0d 0a 43 6f 6e 74 65 6e 74 2d 4c 65 6e 67 ed..Content-Leng
) 74 68 3a 20 31 31 32 0d 0a 0d 0a 75 73 65 72 65 th: 112....usere
) 6d 61 69 6c 73 3d 76 69 73 68 75 2b 25 33 43 39 mails=vishu+%3C9
) 34 35 33 31 39 32 32 36 37 25 33 45 26 74 78 74 453192267%3E6txt
) 5f 6d 73 67 3d 68 65 6c 6f 2e 2e 2e 2e 2e 2e _msg=hello.....
) 2e 2b 2b 68 77 7a 2b 75 2b 25 33 46 25 33 46 25 .+hwz+u+%3F%3F%
) 33 46 26 68 66 5f 6d 73 67 3d 26 69 73 6c 61 6e 3F&hf_msg=6islan
) 67 3d 26 61 63 74 5f 6d 6e 6f 73 3d 39 31 39 34 g=6act_mnos=91194
) 35 33 31 39 32 32 36 37 25 32 43 53192267%2C
-----
869674 172.31.132.59 -> 172.31.100.14 HTTP GET http://www.160by2.com/css/innerpage
    
```

Figure-11 Captured content of the message

Before doing arpspoofing, IP forwarding is enabled so that all the traffic passes through the attacker’s system. The attacker determines whether the IP forwarding is enabled in the system or not by the command ‘cat /proc/sys/net/ipv4/ip forward’ If the IP forwarding is disabled in the system then the output is 0 else the output is 1. When the system has its IP forwarding disabled then it is enabled by the following command as given in figure-12.

```

echo 1 > /proc/sys/net/ipv4/ip forward

ip_dynaddr      ip_local_port_range ip_no_pmtu_disc
root@pooja-laptop:/home/pooja# cat /proc/sys/net/ipv4/ip_forward
1
root@pooja-laptop:/home/pooja#
    
```

Figure-12 IP forwarding

The communication between the host and a gateway is achieved in a defined manner Computer A whose IP address is 172.31.132.42 and MAC address is 00:24:be:b5:a6:73 wants to communicate with gateway whose IP address is 172.31.132.1 and MAC address is 00:1b:d4:74:62:bf to access Internet. Computer A sends out ARP request to gateway requesting MAC address. Switch receives request (which is broadcasted) and passes this request along to every connected computer. Switch also updates its internal MAC address to port table.

Gateway receives ARP request from Computer A, and replies with MAC address. Gateway updates internal ARP table with MAC address and IP address of Computer A. Switch receives ARP reply to Computer A, checks its table, and finds Computer A’s MAC address listed at port 1. It passes this information to port 1 and then updates MAC table with MAC address from gateway. Computer A receives ARP information from gateway, and it updates its ARP table with this information. Computer A sends information out to gateway using updated MAC address information, and communication channel is established. ARP spoofing is now done after the IP forwarding is enabled to sniff all the packets going between a host IP 172.31.132.49 and gateway IP 172.31.132.1, which is being sent to the internet as illustrated in figure 13.

```

arpspoof -t 172.31.132.42 172.31.132.1 & > /dev/null
    
```

Figure 13 illustrates that all the packets that were destined to 172.31.132.1 are rerouted to the system running this command. The system whose IP address is 172.31.132.42 and MAC address 0:24:be:b5:a6:73 is being spoofed by the attacker’s system whose IP address is 172.31.132.59 and MAC address is :23:5a:47:cc:21. The system running ARP spoof whose MAC address is 0:23:5a:47:cc:21 broadcasts the ARP reply that it has the IP address 172.31.132.42. The victim’s MAC address is spoofed by the attacker’s MAC address.

```

0:23:5a:47:cc:21 ff:ff:ff:ff:ff:ff 0806 42: arp reply 172.31.132.42 is-at 0:23:5a:47:cc:21
0:23:5a:47:cc:21 ff:ff:ff:ff:ff:ff 0806 42: arp reply 172.31.132.42 is-at 0:23:5a:47:cc:21
0:23:5a:47:cc:21 ff:ff:ff:ff:ff:ff 0806 42: arp reply 172.31.132.42 is-at 0:23:5a:47:cc:21
0:23:5a:47:cc:21 ff:ff:ff:ff:ff:ff 0806 42: arp reply 172.31.132.42 is-at 0:23:5a:47:cc:21
0:23:5a:47:cc:21 ff:ff:ff:ff:ff:ff 0806 42: arp reply 172.31.132.42 is-at 0:23:5a:47:cc:21
0:23:5a:47:cc:21 ff:ff:ff:ff:ff:ff 0806 42: arp reply 172.31.132.42 is-at 0:23:5a:47:cc:21
0:23:5a:47:cc:21 ff:ff:ff:ff:ff:ff 0806 42: arp reply 172.31.132.42 is-at 0:23:5a:47:cc:21
0:23:5a:47:cc:21 ff:ff:ff:ff:ff:ff 0806 42: arp reply 172.31.132.42 is-at 0:23:5a:47:cc:21
0:23:5a:47:cc:21 ff:ff:ff:ff:ff:ff 0806 42: arp reply 172.31.132.42 is-at 0:23:5a:47:cc:21
0:23:5a:47:cc:21 ff:ff:ff:ff:ff:ff 0806 42: arp reply 172.31.132.42 is-at 0:23:5a:47:cc:21
    
```

Figur-13- ARP request

### 3.1 Capturing of WebPages Visited

Dsniff is a tool that extracts information about the webpages visited by the victim. Let us consider the following case study as conducted in the Information Security Laboratory. The victim’s MAC address 00:24:be:b5:a6:73 has been spoofed by the attacker’s MAC address 0:23:5a:47:cc:21. In figure-14, victim’s IP 172.31.132.42 has been spoofed and IP forwarding has already been enabled to get the whole traffic between the victim and the gateway IP. It shows all the webpages

which has been visited by victim in the system who is running the dsniff tool. Here, the system whose IP address is 172.31.132.59 and MAC address is 0:23:5a:47:cc:21 dsniffs all the webpages visited by the victim's system. Hacker-Arun first connects to the web site *www.google.com* on date 07-07-10 at the time 15:13:05 and then to the mail.mnnit.ac.in after 2 minutes 15:15:53 on the same day.

```
-----
07/07/10 15:13:05 tcp Hacker-Arun.local.43924 -> 172.31.100.14.3128 (http)
CONNECT www.google.com:443 HTTP/1.1
Host: www.google.com
Proxy-Authorization: Basic YXJ1bmtzaW5naDphcnVuc2luZ2gx [arunksingh:arunsingh1]
-----
07/07/10 15:15:53 tcp Hacker-Arun.local.60382 -> 172.31.100.14.3128 (http)
GET http://mail.mnnit.ac.in/webmail/images/draft.png HTTP/1.1
Host: mail.mnnit.ac.in
Proxy-Authorization: Basic YXJ1bmtzaW5naDphcnVuc2luZ2gx [arunksingh:arunsingh1]
-----
GET http://mail.mnnit.ac.in/webmail/images/senti.png HTTP/1.1
Host: mail.mnnit.ac.in
Proxy-Authorization: Basic YXJ1bmtzaW5naDphcnVuc2luZ2gx [arunksingh:arunsingh1]
```

Figure-14 Capturing of Webpage Visited

### 3.2 Denial Of Services

In a denial-of-service (DoS) attack, an attacker attempts to pre-vent legitimate users from accessing information or services. It is an action or set of actions that prevent any part of a system from functioning as it should. This includes the actions that causes unauthorized destruction, modification, or delay of service. DoS results in the loss of a service in a particular network or temporary loss of services in all the network services. It does not usually used to sniff the data and information passing through the network traffic over the HTTP connection in LAN. By targeting victim's computer and its network connection, an attacker may be able to prevent him from accessing email, websites, online accounts (banking, etc.) or other services that rely on the affected computer. When a person connects to a website into the browser, he is sending a request to that site's computer server to view the page. There is a limit to the number of the requests which can be accessed at a given time. So, the attacker overloads the server with requests, which in turn can not process the victim's request.

DOS includes sending oversized ICMP echo packets which increases the payload and results in Denial of Services for the client.

### 4.0 Countermeasures for Network Attacks

Static ARP table is a one way to prevent the ARPspoofing. The ARP table is generated using the command `arp -s IPaddress MAC address` This will add static entries to the table i.e. unchanging entries which

prevents attacker from adding spoofed ARP entries as illustrated in figure-17. This detects if a new Ethernet device is added to an existing network, but it has no method of predefining an acceptable IP address. In this figure, a static entry to the ARP table is added by `arp -s 172.31.152.45 00:1B:D4:74:62:BF`.

### 4.1 Static ARP Table

The table will record this IP address and MAC address. As a result no ARP spoofing can be done. Whenever there is any data communication in between the hosts over the HTTP connection in LAN, it will check whether the table has the particular IP address or not before broadcasting the ARP request to each hosts on the network. So, no ARP broadcasts request is sent which prevents the ARP spoofing. ARP table shows IP address, MAC address, interface and flag

```
root@pooja-laptop:/home/pooja# arp -e
Address      HWtype  HWaddress      Flags Mask    Iface
172.31.100.14 ether    00:1B:D4:74:62:BF CM            eth0
root@pooja-laptop:/home/pooja# arp -s 172.31.152.45 00:1B:D4:74:62:BF
root@pooja-laptop:/home/pooja# arp -e
Address      HWtype  HWaddress      Flags Mask    Iface
172.31.152.45 ether    00:1B:D4:74:62:BF CM            eth0
172.31.100.14 ether    00:1B:D4:74:62:BF CM            eth0
```

Figure-15- Adding Static entry to the ARP table

Mask in figure 15. If any static entry is added to the ARP table, then the corresponding IP/MAC address is marked and remains unchanged until the system shuts down.

### 4.2 ARPwatch

ARPwatch is a program which works by monitoring an interface in promiscuous mode and recording MAC and IP address pairings over a period of time. When it sees anomalous behavior in case of change to one of the MAC and IP address pairs that it has received, it will send an alert in the form of a warning to the user. ARPwatch runs by selecting one of the inter- face from multiple interfaces on the command line. It runs and records the IP and MAC address by `arpwatch -d` and gives the information about hostname, host IP address, interface, Ethernet address and time when it is recorded as illustrated in figure-16. The system running the ARPwatch gets the details of MAC and the corresponding IP addresses. In the presented simulation, the system *pooja-laptop* is running the ARPwatch and gets the information about the unknown host name whose IP address is 172.31.134.126, interface is *eth0* and has an ethernet address 0:13:20:b1:3d:8. It again records that the host name '*Hacker-Arun*' whose IP address is 172.31.132.42 , interface is *eth0* and has its corresponding MAC address 0:24:be:b5:a6:73. A file

arp.dat is created so as to record the MAC/IP address of the system in that network.

```

rom: arpspoof (Arpspoof pooja-laptop)
o: root
ubject: new station eth0

    hostname: <unknown>
    ip address: 172.31.134.126
    interface: eth0
    ethernet address: 0:13:20:b1:3d:8
    ethernet vendor: <unknown>
    timestamp: Monday, July 5, 2010 12:39:42 +0530

rom: arpspoof (Arpspoof pooja-laptop)
o: root
ubject: new station (Hacker-Arun.local) eth0

    hostname: Hacker-Arun.local
    ip address: 172.31.132.42
    interface: eth0
    ethernet address: 0:24:be:b5:a6:73
    ethernet vendor: <unknown>
    timestamp: Monday, July 5, 2010 12:40:38 +0530
    
```

Figure 16- Record of MAC and IP addresses made by ARPwatch

This file is reloaded every time a new pair of MAC and IP address becomes known. Whenever there is any change found in MAC and IP address, then ARPwatch alerts the person that ARPspoofing of a particular MAC is done as shown in figure-17. The system executing this program as this simulated attack is that pooja-laptop gets to know that the host-name 'Hacker-Arun' whose IP address is 172.31.132.42, interface eth0 and has now changed its MAC address from 0:24:be:b5:a6:73 to 0:30:65:24:21:36. Detection of ARP spoofing ARPwatch by first finding all of the current ARP entries by the command arp -a sends an alert. Then, one among them is selected for ARPspoofing which spoofs the victim's MAC address by the attacker's MAC address. This is detected by ARPwatch and it shows the alert by showing the old ethernet address and current ethernet address as illustrated in figure-18. arp -a command finds the current ARP entry which has the IP address 172.31.132.49 and MAC address 00:16:35:ae:56:14 which is shown in the right hand side of the figure 18.

```

delta: 39 minutes

From: arpspoof (Arpspoof pooja-laptop)
To: root
Subject: changed ethernet address (Hacker-Arun.local) eth0

    hostname: Hacker-Arun.local
    ip address: 172.31.132.42
    interface: eth0
    ethernet address: 0:30:65:24:21:36
    ethernet vendor: Apple Computer, Inc.
old ethernet address: 0:24:be:b5:a6:73
old ethernet vendor: <unknown>
    timestamp: Monday, July 5, 2010 16:11:27 +0530
previous timestamp: Monday, July 5, 2010 12:40:38 +0530
    delta: 3 hours
    
```

Figure- 17 Alert when change in IP and MAC address Then, the ARP spoofing is done which is illustrated in the above side of the figure. The system whose MAC address 0:23:5a:47:cc:21, is running the ARPspoof broadcasts an ARP reply that the system having IP address 172.31.132.49 is at 0:23:5a:47:cc:21. This ARP spoofing is detected by this tool ARPwatch which is shown in the left hand side of the figure 20 i.e. hostname 'niraj-desktop' is having IP address 172.31.132.49, interface eth0, whose old ethernet address was 00:16:35:ae:56:14, is now changed to 0:23:5a:47:cc:21, the attacker's MAC address running the ARPspoof.

```

oot@pooja-laptop:/usr/sbin# ./arpspoof 172.31.132.49
:23:5a:47:cc:21 ff:ff:ff:ff:ff:ff 0806 42: arp reply 172.31.132.49 is-at 0:23:5a:47:cc:21
:23:5a:47:cc:21 ff:ff:ff:ff:ff:ff 0806 42: arp reply 172.31.132.49 is-at 0:23:5a:47:cc:21
:23:5a:47:cc:21 ff:ff:ff:ff:ff:ff 0806 42: arp reply 172.31.132.49 is-at 0:23:5a:47:cc:21
:23:5a:47:cc:21 ff:ff:ff:ff:ff:ff 0806 42: arp reply 172.31.132.49 is-at 0:23:5a:47:cc:21
:23:5a:47:cc:21 ff:ff:ff:ff:ff:ff 0806 42: arp reply 172.31.132.49 is-at 0:23:5a:47:cc:21
:23:5a:47:cc:21 ff:ff:ff:ff:ff:ff 0806 42: arp reply 172.31.132.49 is-at 0:23:5a:47:cc:21
:23:5a:47:cc:21 ff:ff:ff:ff:ff:ff 0806 42: arp reply 172.31.132.49 is-at 0:23:5a:47:cc:21
    
```

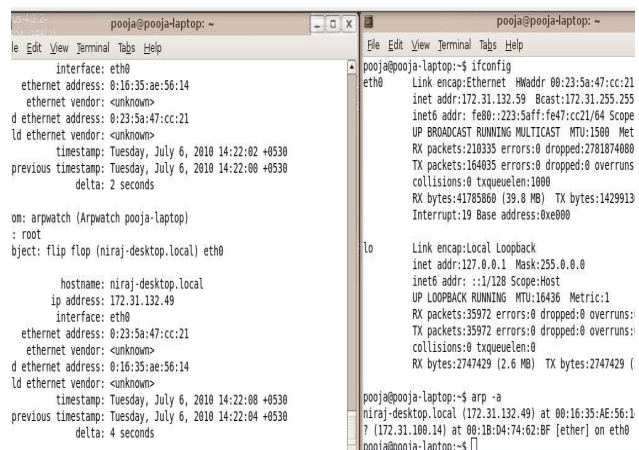


Figure-18 ARP spoofing Detected by ARPwatch

The Security Threat and the Response attack, is also an attack on a network's resources, but is launched from a large number of other host machines. This is a type of DOS attack. Attacking software is installed on these host computers, unbeknownst to their owners, and then activated simultaneously to launch communications to the target

machine of such magnitude as to overwhelm the target machine (figure 19, figure 20).

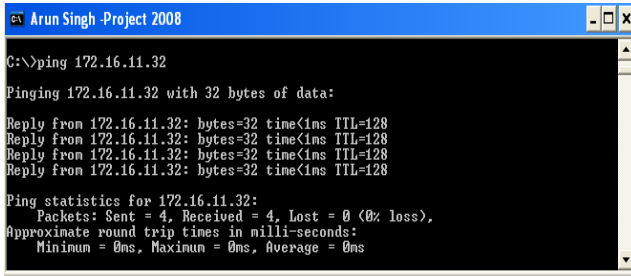


Figure- 19 Ping command to check system is alive or not

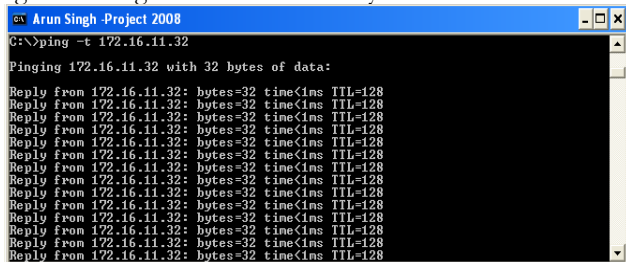


Figure -20 Dos Attack

The proposed solution which is given for ARP poisoning is to have control of the user over the ping reply i.e. if the user wants to reply the ping then only he or she can reply else not. Control of the user can be of two types either he or she ignores all the ICMP echo packets or accepts all. In the first one, the user will ignore all the ICMP echo packets i.e. the other system user will not be able to detect whether the system is host or not even if the system is actually hosting up. In the second one, the user will accept all the ICMP echo packets i.e. if any other system pings the user's system, it will reply the number of times it is asked to do so. This will increase the payload on the user's system. which may lead to crash. The proposed solution gives a way to have control on this payload which in turns, benefits the user to reply to the system once when it is pinged by another system and then stops for some time and then continue again. This pattern may be repetitive. Such repetitive pattern may be indicative of a network attack or vulnerabilities. This will reduce the payload to a very great extent which was the disadvantage of accepting all ICMP echo packets and will also inform the other users that the host is up. By this way, the proposed solution will overcome both the problems arising earlier. The solution is designed using shell script. If the user is busy and does not want to reply then it will ignore all the ICMP echo packets and continue doing his or her work even if the other system pings the user's system. But, if the user is not busy and wants to reply the trusted system so that no

ARP poisoning could take place, then he/she may choose to reply to the system requesting.

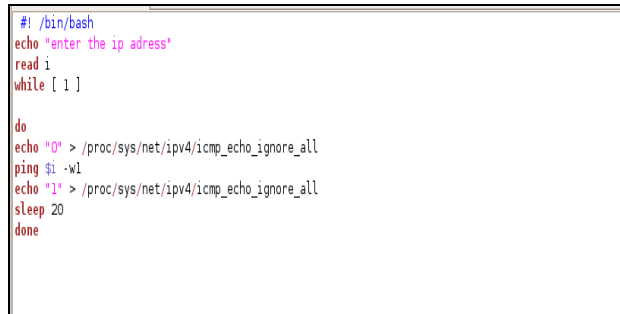


Figure-21 Shell script preventing DoS



Figure- 22 Shell script to prevent payload

## 5.0 Conclusion and Future Direction of Work

Security threats and breaches in an organization's network infrastructure can cause critical disruption of business processes and lead to information and capital losses. A potent security system is imperative for an enterprise networks and vulnerability assessment is an important element for the same.

A host-based vulnerability scanning system informs about the vulnerabilities that the respective host carries. This paper provides a review of the current research related to host-based vulnerability assessment followed by avenues for further research. It is important to make a distinction between penetration testing and network security assessments. Some of the simulators have strong resemblance with the penetration testing but these differ for their purpose. The purpose has carefully been taken care to simulate the attacked and its successful solution by writing scripts for these attacks. A network security or vulnerability assessment may be useful to a degree, but do not always reflect the extent to which hackers will go to exploit a vulnerability. Penetration tests attempt to emulate a 'real world' attack to a certain degree. The penetration testers will generally compromise a system with vulnerabilities that they successfully exploited.

If the penetration tester finds several holes in a system to get in this does not mean that hackers or external



intruder will not be able to find more than the holes deleted earlier. Hackers and intruders need to find only one hole to exploit whereas penetration testers need to possibly find all if not as many as possible holes that exist. This is a daunting task as penetration tests are normally done within a certain time frame.

A penetration test alone provides no improvement in the security of a computer or network. Action taken to address these vulnerabilities that is found as a result of conducting the penetration test are not the part of penetration listing. Security is an ever-changing arena. Hackers are constantly adapting and exploring new avenues of attack. The technology is constantly changing with new versions of operating systems and applications. The result of all this change is an increased risk to the typical workstation based on popular operating system. Increased upgrades and patches are a result of the need to propagate fixes to security vulnerabilities. The quick fixes of vulnerabilities presented in this paper provide a readymade solution.

This paper also provides an overview of Network Security Monitoring (NSM) which involves network analysis through NMAP, sniffing of the packets across the traffic over the HTTP connection in LAN by Wireshark, Tshark, Dsniff. Analysis and detection of ARP poisoning are discussed briefly to highlight the vulnerabilities in the data communication over the HTTP connection in LAN. The proposed solution given in this paper to stop the MITM attack in LAN is simple to understand and provides the user to have a control over the ping reply given by its system. It allows the user to defend from the attack of ARP poisoning. The future work can extend this bash shell script to block the particular IP address if it pings the system many times and does not allow any system to send the packet with a greater size than that has been sent the first time. Analysis of data communication over HTTPS connections in LAN and secure routing of the network data communication over HTTP and HTTPS in LAN or Wi-Fi are the other area having applicability of the present research.

## 6.0 Acknowledgement

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions. The research reported here is fully supported by the ISEA Project, DIT, MCIT, and Government of India.

## References

- [1]. Cryptography and Network Security Principles and Practices, Fourth Edition, By William Stallings, Prentice Hall publication, 2006
- [2]. Arpspoof a arp poisoning tool available at, <http://monkey.org/dugsong/dsniff/> [Accessed on May 20, 2010].

- [3]. Ettercap a arp poisoning tool, <http://ettercap.sourceforge.net/> [Accessed on May 20, 2010].
- [4]. HTTPS sniffing through ssniff, <http://thoughtcrime.org> [Accessed on May 20, 2010].
- [5]. <http://www.blackhat.com/presentations/bh-dc-09/Marlinspike/BlackHat-DC-09-Marlinspike-Defeating-SSL.pdf> [Accessed on May 20, 2010].
- [6]. tshark command manual at <http://www.wireshark.org/docs/manpages/tshark.html> [Accessed on May 20, 2010].
- [7]. B. Ross, C. Jackson, N. Miyake, D. Boneh, and J. C. Mitchell, Stronger Password Authentication Using Browser Extensions, Proceedings of the 14th Usenix Security Symposium, 2005
- [8]. Nmap Security Scanner For Network Exploration & Security <http://nmap.org/>
- [9]. Wireshark. [Online document] Available: <http://www.wireshark.org/>
- [10]. US-CERT Technical Cyber Security Alert <http://www.us-cert.gov/cas/tips/ST04-015.html> [Last Accessed on 5th July, 2010]
- [11]. Wireshark, <http://www.wireshark.org/docs/manpages/tshark.html>, [Last Accessed on 8th July, 2010]
- [12]. Douglas E. Comer, Internetworking with TCP/IP Principles, Protocols and Architecture, Fifth Edition,
- [13]. Pearson Prentice Hall Publications, 2006 Angela Orebaugh, Wireshark & Ethereal Network
- [14]. Protocol Analyzer Toolkit, Syngress Publication, 2007
- [15]. Chris Sanders, Practical Packet Analysis using Wireshark to solve real world Network Problems, William Pollock Publications, 2007
- [16]. David Slee, Common Denial of Service Attacks, July 10, 2007.
- [17]. Renaud Bidou, Denial of Service Attacks Joe Habraken, Absolute Beginner's Guide to Networking, Fourth Edition, Que Publication, 2003.
- [18]. Arun Kumar Singh, Lokendra Kumar Tiwari, Shefalika Ghosh Samaddar and C.K Dwivedi, Security Policy & Its Scope in Research Area, accepted in International Conference on Strategy and Organization, ICSO 2010 on 14 & 15 May-2010, Institute of Management Technology, Ghaziabad, Uttar Pradesh, India.
- [19]. Lokendra Kumar Tiwari, Arun Kumar Singh, Shefalika Ghosh Samaddar and C.K Dwivedi, Recovery Evidentiary files using Encase Ver 6.0, accepted and presented in National conference & Workshop on High Performance & Applications, 08-10 February, Banaras Hindu University, Varanasi, Uttar Pradesh, India, pp-8.
- [20]. Lokendra Kumar Tiwari, Arun Kumar Singh, Shefalika Ghosh Samaddar and C.K Dwivedi, Evidentiary Usage of E-mail Forensics: Real Life Design of a Case, First International Conference on Intelligent Interactive Technologies and Multimedia (IITM-2010) page 219-223, on Dec 28-30, 2010, Indian Institute of Information Technology Allahabad, Uttar Pradesh, India..

- [21]. Arun Kumar Singh, Pooja Tewari, Shefalika Ghosh Samaddar and A.K. Misra, Communication Based Vulnerabilities and Script based Solvabilities, International Conference on Communication, Computing & Security (Proceedings by ACM with ISBN-978-1-4503-0464-1) on 12-14 Feb-2011, National Institute of Technology Rourkela Orissa, India.
- [22]. Arun Kumar Singh, Pooja Tewari and Shefalika Ghosh Samaddar, A. K. Misra, Vulnerabilities of Electronics Communication: solution mechanism through script, International Journal of Computer Science Issues (IJCSI), Volume 8, Issue 3, 2011 (IN Press).
- [23]. Arun Kumar Singh, Lokendra Tiwari, Vulnerability Assessment and penetration Testing, National Conference on Information & Communication Technology (NCICT-2011), ISBN: 978-93-80697-77-2, 5th-6th March, 2011, Centre for Computer Sciences Ewing Christian College Allahabad-211003 Uttar Pradesh, India.
- [24]. Lokendra Kumar Tiwari, Arun Kumar Singh, Shefalika Ghosh Samaddar and C.K Dwivedi, An Examination into computer forensic tools, accepted and to be presented in 1st International Conference on Management of Technologies and Information Security (ICIMS 2010) page 175-183, on 21-24 of January 2010, Indian Institute of Information Technology Allahabad, Uttar Pradesh, India. ([http://icmis.iiita.ac.in/TOOL\\_FORENSIC.ppt](http://icmis.iiita.ac.in/TOOL_FORENSIC.ppt)).



**Corresponding Author:** Arun Kumar Singh received his B.Tech in Electronics and Communication from SRMCEM College, Lucknow, Uttar Pradesh, India in 2005. He received his MS degree in Information Security from Indian Institutes of Information Technology, Allahabad, Uttar Pradesh, India in 2008. Currently, he is pursuing the Ph.D. degree in Computer Sciences and Engineering at the Motilal Nehru National Institute of Technology (MNNIT), Uttar Pradesh, India. He also is working as a Research Associate at the MNNIT. His research interests include network security, network protocol design and verification, in network security, Cryptography and Computer Forensic fields.



# Image Registration in Digital Images for Variability in VEP

<sup>1</sup> N.Sivanandan , Department of Electronics, Karpagam University , Coimbatore, India.

<sup>2</sup> Dr.N.J.R.Muniraj, Department of ECE, Anna University,KCE, Coimbatore, India.

## Abstract

The visually evoked potential (VEP) is the measure of cortically evoked electrical activity that provides information about the integrity of the optic nerve and the primary visual cortex. The analysis of P-100 latency and amplitude measurement variability based on visual pathway conduction in VEP has been shown to have clinical utility. The reliable measurement of VEP techniques to do are less well developed. This work presents a technique for a reliable extraction P-100 latency and amplitude using a wavelet based technique. The challenge of image registration (the process of correctly aligning two or more images accounting for all possible source of distortion) is of general interest in image processing. Several types of VEPs are routinely used in a clinical setting. These primarily differ in a mode of stimulus presentation.. This registration can be carried out for VEP waveforms of the same subject taken at different times, waves taken under different modalities, and wave pattern which have only a partial overlap area. This research focused on investigating potential registration algorithms for transforming partially overlapping VEP waves which have only a partially overlapping waveform of the retina into a single overlapping composite waveform to aid physicians in assessment of retinal health, and on registering vectors from known common points in the images to be registered. All potential transforms between waveforms are generated, with the correct registration producing a tight cluster of data points in the space of transform coefficients. The technique has been applied to different types of retinal waveforms – B/W checker board (pattern reversal),B/W checker board (flash),LED Goggles (pattern reversal) and LED Goggles(flash) stimulations and the technique can be readily used to provide cross – modal.

**Keywords:**

*Electro diagnostic instrument, VEP stimulator, Image Registration, Discrete wavelet transform and VEP signals.*

## 1.Introduction

The VEP is the measure of cortically evoked electrical activity that provides information about the integrity of the optic nerve and the primary visual cortex. The optic nerve joins the retina with the brain. On giving pattern or flash stimulation, not only there is increased metabolism in primary visual area but also in the visual association areas. The VEP studies in patients with well defined cortical lesions provide additional about its generator sources. It is important that the infant or child does fix on the stimulus. In the children below five years, the pattern reversal is first carried out; if the potentials are un recordable then a flash VEP should be undertaken. The pattern reversal useful as these assess the visual acuity whereas the flash VEP determines the presence or absence of light perception.

It is important to check the variability of a number of above mentioned parameters for reliable interpretation of VEP. The P-100 latency increases with the decrease of luminance. The reduction of contrast between black and white squares results in increased latency and decreased amplitude of P-100. Usually black and white checks or gratings are employed in clinical practice. Use of colors such as green-black or red-black increase the frequency of VEP abnormalities. The pattern reversal frequency if

increased from 1Hz to 4Hz , the P-100 latency increases by 4.8 m sec. At a faster rate, the waveforms become less distinct and stimulation above 8-10 Hz results in a steady state VEP. The

VEP is not influenced by the direction of pattern shift.

## 2.Methods of Visual Evoked Potential

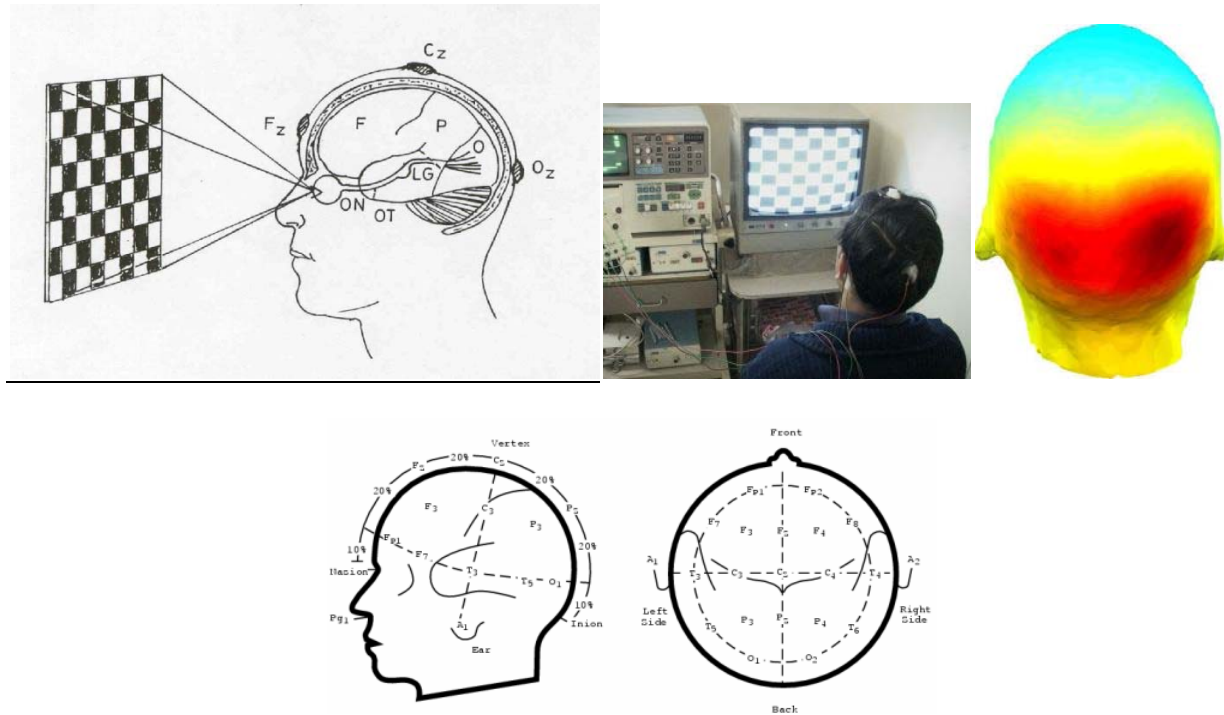


Fig 1. Basic VEP different parts

For the best results of VEP testing, patient should be explained about the test to ensure full cooperation and should avoid hair spray or oil after the last hair wash. The usual glasses if any should be put on during the test. The results of ophthalmological examination such as visual acuity, papillary diameter and field changes should be review before starting the test. The VEP recording were performed in a dark and sound attenuated room in a laboratory. Subject was asked to sit comfortably in front of the checker board pattern at an eye screen distance of 100 cm. The preferred stimulus for clinical investigation of the visual pathways is a reversal of a black and white checker board pattern, as it tends to evoke larger and clear responses than other patterns. For VEP, standard disc EEG electrodes are used. The recording electrode is places at Oz using conducting jelly or electrode paste as per 10-20 international system of EEG electrode placement. The reference is placed at FPz or 12 cm above the nasion. The ground electrode is placed at the vertex at Cz. The electrode impedance should be kept below 5 kilo ohms.

The square checks alternate from black/white to white/black at a specified rate without change in the overall luminance of the screen. This is accomplished by displaying 8\*8 checker board pattern on the computer screen using visual basic software. These stimuli elicit VEP responses in the visual cortex. From the VEP recordings, measured values of P-100 component because of waveform consistency and reliability among the normal subjects. Normal amplitude component are much less useful interpretive tools than latencies because of variation in results obtained from normal subject.

## 3.Introduction to Image Registration

Image registration is the process of transforming the different sets of data into one coordinate system. Registration is necessary in order to be able to compare or

integrate the data obtained from different measurements. Medical imaging registration for data of the same patient taken at different points in time often additionally involves elastic registration to cope with elastic deformations of the body parts imaged. The original image is often referred to as the reference image and the image to be mapped onto reference image is referred to as the target image. Image similarity based are broadly used in medical imaging. A basic image similarity based method consist of a transformation modal, which is applied to reference image coordinates to locate their corresponding coordinates in the target image space, an image similarity metric, which quantifies the degree of correspondence between features in both image spaces achieved by a given transformation and an optimization algorithm which tries to maximize image similarity by changing the transformation parameters.

The choice of an image similarity measure depends on the nature of the images to be registered. Common examples of image similarity measures include cross-correlation, Mutual information, Mean square difference and ratio image uniformity. Mutual information and its variant ,normalized registration of multimodality images. Cross-correlation ,mean square difference and ratio image uniformity are commonly used for registration of images of same modality.

#### 4. Discrete wavelet transform (DWT)

If the function being expanded is discrete (ie, a sequence of coefficients.If the function being expanded is discrete ie,a sequence of numbers) the resulting coefficients are called the discrete wavelet transform (DWT).

If  $f(n) = f(X_0 + \Delta N_x)$  for some  $X_0, \Delta X$  and  $n=0, 1, 2, 3, \dots, M-1$ , the wavelet series expansion coefficients for  $f(x)$  defined by

$$C_{jo}(k) = \langle f(x), \phi_{jo,k}(x) \rangle = \int f(x) \phi_{jo,k}(x) dx \text{ and} \text{-----}(1)$$

$$D_j(k) = \langle f(x), \psi_{j,k}(x) \rangle = \int f(x) \psi_{j,k}(x) dx \text{-----}(2)$$

Become the forward DWT coefficients for sequence  $f(n)$ :

$$w_{\phi}(j_0, k) = 1/\sqrt{m} \sum f(n) \phi_{j_0, k^{(n)}} \text{-----}(3)$$

$$w_{\psi}(j, k) = 1/\sqrt{m} \sum f(n) \psi_{j, k^{(n)}} \text{ for } j \geq j_0 \text{-----}(4)$$

The  $\psi(0), k^{(n)}$  and  $\psi_j, k^{(n)}$  in the equations are sampled versions of basic functions  $\phi_{j_0, k^{(n)}}$  and  $\psi_j, k^{(n)}$ .

If  $\psi_{j_0, k^{(n)}} = \phi_{j_0, k}(x_s + \Delta x_s)$  for some  $x_s$ , equally spaced samples over the support of the basic functions. In accordance with equations

$$f(x) = \sum_{j_0} \psi_{j_0, k^{(n)}}(x) \sum_{j_0} \sum_{k^{(n)}} d_{j_0, k^{(n)}} \text{-----}(5)$$

Normally, we let  $j_0=0$  and select  $M$  to be a power of 2. So that summations in equations (3) through (5) are performed over  $n=0, 1, 2, \dots, M-1, j=0, 1, 2, \dots, j-1$  and  $k=0, 1, 2, \dots, 2^j - 1$ . The  $W_{\phi}(j_0, k)$  and  $W_{\psi}(j, k)$  in equations (3) to (5) correspond to the  $C_{j_0, k^{(n)}}$  and  $d_{j_0, k^{(n)}}$  of the wavelet series expansion. Note that the integration in the series expansion have been replaced by summations and a  $1/\sqrt{m}$  normalizing factor, reminiscent of the DFT.

Using the equations (3) through (5), consider the discrete functions of four points in VEP study, ie,  $f(0), f(1), f(2)$  and  $f(3)$ . Where  $f(0)$  is the checker board pattern reversal,  $f(1)$  is the checker board flash,  $f(2)$  is the LED Goggles pattern reversal and  $f(3)$  is LED Goggles flash stimulation. These four points to be considered as  $f(0)=1, f(1)=4, f(2)=-3$  and  $f(3)=0$ . Because  $m=4, j=2$  and with  $j_0=0$  and summations are performed over  $x=0, 1, 2, 3, j=0, 1$  and  $k=0$  for  $j=0$  or  $k=0, 1$  for  $j=1$ .

We will use the Hear scaling and wave let functions and assume that the four samples of  $f(x)$  are distributed over the support of the basic function, which is in width. Substituting the four samples into equations (3), we find that

$$\begin{aligned} W_{\psi}(0, 0) &= 1/2 \sum f(n) \psi_{0, 0^{(n)}} \\ &= 1/2 [107.5 \text{ m sec} + 113.1 \text{ m sec} - 113.1 \text{ m sec} + 116.9 \text{ m sec}] \\ &= 1/2 [224.8 \text{ m sec}] \\ &= 112.4 \text{ m sec} \end{aligned}$$

Because  $\psi_{0, 0^{(n)}}=1$  for  $n=0, 1, 2, 3$  note that we have employed uniformly spaced samples of the Hear transmission matrix. Therefore the P-100 latency of four point stimulations of DWT are uniformly spaced samples of the scaling and wave let functions are used in the computation of the inverse.

The four point DWT in the VEP P-100 latency measurement of a two scale decomposition of  $f(x)$ , ie,  $j=\{0, 1\}$ . To underlying assumption was that starting scale  $J_0$  was zero but other starting scales are possible.

#### 5. Experimental procedure

The VEP recording were performed in a dark and sound attenuated room in a laboratory. Subject was asked to sit comfortably in front of the checker board pattern at an eye screen distance of 100 cm. The preferred stimulus for clinical investigation of the visual pathways is a reversal of a black and white checker board pattern, as it tends to evoke larger and clear responses than other patterns. The

stimulus pattern was a black and white checkerboard displayed on a computer screen. The checks alternate from black/white and white/black at a rate approximately of twice per second. The subject was instructed to gaze at a colored dot on the center of the checkerboard pattern. Every time the pattern alternates, the patient visual system generates an electrical response and was recorded using electrodes. Signal acquisition and stimulus presentation

was synchronized using software program. The starting point of VEP waveform is stimulus onset. The VEP waveform recording is done over a period of 250 m sec. More than 100 epochs were averaged to ensure a clear VEP waveform. For judging the reproducibility, the waveform is recorded twice and superimposed. A typical averaged various types of stimulations like

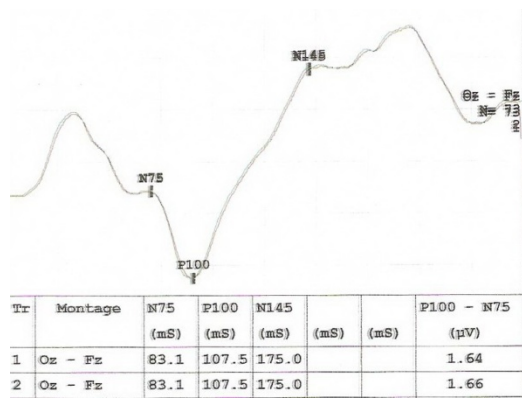


Fig 2. B/W Checker board (Pattern Reversal)

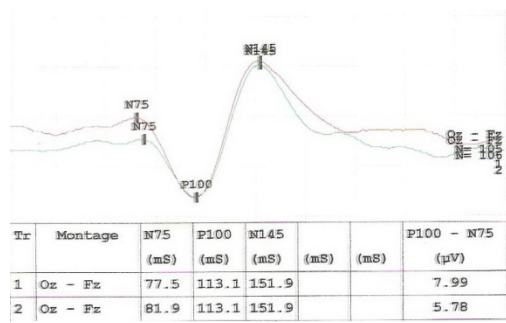


Fig 3. B/W Checker board-(flash)

different types of retinal images – B/W checker board (pattern reversal),B/W checker board (flash),LED Goggles (pattern reversal) and LED Goggles(flash) stimulations are recorded with same subject with variability P-100 latencies are amplitudes are noted in the following superimposed waveforms and results. The VEP signal has been labeled to indicate the N75,P100 and N145 marks, the corresponding latencies for the subject being

83.1 ms,107.5 ms and 175 ms for B/W pattern reversal checker board,77.5 ms,113.1 ms,151.9ms for B/W flash

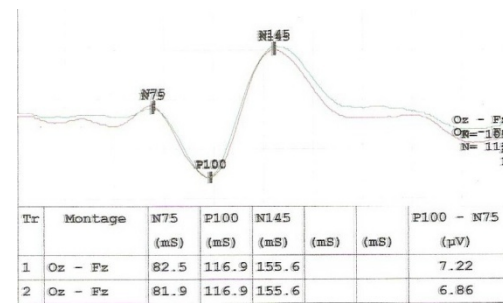


Fig 4. LED Goggles(Pattern Reversal)

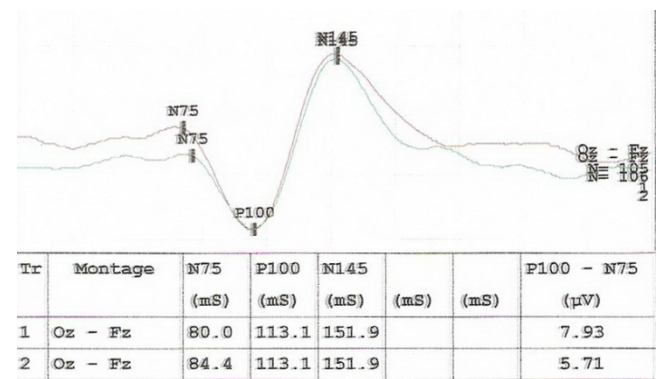


Fig 5. LED Goggles(Flash)

checkerboard stimulation,82.5 ms,116.9 ms and 155.6 ms for LED Goggles pattern reversal and 80.0 ms,113.1 ms and 151.9 ms noted for LED Goggles flash stimulation.

Table :1 Developed accuracy P-100 latency measurement

Montage	Tr	N75 ms)	P100(ms)	N145(ms)
Oz-Fz	1		112.4	
Oz-Fz	2		112.4	

Finally, all the potential transforms between images are generated, with the correct registration producing accurate P-100 latency measurement of 112.4 m sec with different types of retinal waves as shown in the above table (V).

## 6. Results and conclusion

The developed accurate P-100 latency measurement evaluated using different types of stimulations under the process of image registration correctly aligning of the same subject taken at different stimulations and different modalities. The tables I,II,III and IV were variability of P-100 latency and table V showed accurate P-100 latency using discrete wavelet transform. However, there is a small difference in the P-100 latency measurement because of the subjective behavioral factors, like the quality of the cooperation in fixation and accommodation. Diagnosis of optic nerve diseases for the recorded VEP signals is performed (P-100) on the basis of established for that neuro diagnostic laboratory.

## 7. References

- 1.Cinical Neurophysiology-Misra and Kalitha
- 2.Biomedical Signal Analysis- Rangaraj M and Rangayan,John Wiley and sons, singapore.
- 3.Chiappa K.H,Evoked potential in clinical testing'churchill Livingstone (1997)

- 4.FitzGerald,M.J.T and Folan-Curran,J.Clinical Neuroanatomy and Related Neuro science,W.I.FitzGerald
- 5.Biomedical Instrumentation and Measurements-Leslie Cromwell,Fred J.Weibell,Erich A.Pfeiffer-2<sup>nd</sup> edition.
6. Digital Image Processing and Analysis-  
B.Chanda,D.Dutta Majumder
- 7.Digital Image processing,3<sup>rd</sup> edition-Rafael C. Gonzalez, Richard E. Woods
- 8.Sun India's Digital Image Processing- M.C.Trivedi
- 9.Clinical Neuro anatomy and Related Neuro science, W.I. FitzGerald,M.J.T. and FolanCurran,J,Edinburgh,2002.

**N.Sivanandan** has received his M.Sc Applied Electronics from PSG college of Arts & Science, Coimbatore, India .He has 8 years experience in medical technical department in KMCH hospital, Coimbatore, India and 5 years experience in teaching and his area of interest are Biomedical and Digital medical Image processing. Currently, he is doing his Ph.D program at Karpagam University, Coimbatore, India.

**Dr.N.J.R.Muniraj**, Professor and Head, Department of ECE, Karpagam college of Engineering, Anna University, Coimbatore, India.



# WiMAX (Worldwide Interoperability for Microwave Access): a Broadband Wireless Product in Emerging Markets

Komal Chandra Joshi, M.P.Thapliyal

Lecturer, Computer Science Department, Kumaon University, Shriram Institute of Management & Technology  
Kashipur (Udham Singh Nagar), Uttarakha, 244713, India

Associate professor, Computer Science & Engg. Department, HNB Garhwal University,  
Srinagar (Garhwal), Uttarakhand-246174, India

## Abstract

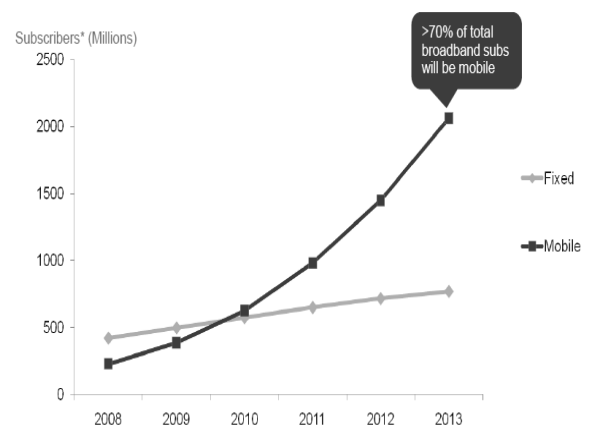
Wireless access networks like WiMAX provide an excellent opportunity for operators to participate in the rapid growth opportunities that exist in emerging markets. Emerging markets are hungry for fixed broadband services; however characteristics of ADSL (Asymmetric Digital Subscriber Line) limit the even distribution of fixed broadband services to encompass urban and rural areas. WiMAX-based access networks will enable local operators to cost-effectively reach millions of new potential customers and provide them with traditional voice and broadband data services. Which are still not possible? Although these markets have all the attributes required for a winning business world, they are not without challenges. But after launching the WiMAX over ADSL in the market, the result are encouraging because with respect to other services the WiMax getting better responses in term of WiMAX's users and operators.

**Keywords:** WiMax, ADSL, Wireless Network, Fixed Broadband, Emerging Market

## 1. Fixed- and Mobile-WiMAX?

WiMAX is a standards-based wireless broadband technology [3], also known by the IEEE standard, 802.16, offering high-speed wireless access over long distances. A WiMAX system have a radio tower, similar to a cellular base station, and a WiMAX antenna and receiver at the customer-end, which can be a modem, PC (personal computer) data card or even a mobile handset. Fixed-WiMAX, often referred to as 802.16d standard, was released in 2004. Pre-standard CPEs (Customer Premises Equipment) are available including outdoor, directional antennas as well as indoor modems. The standard is specified to allow nomad city, where users can access the service from various locations covered by the network. However, in the absence of portable devices currently, present deployments offer only fixed wireless access, meaning users can access the service only from their home location, where the CPE is installed. Mobile-WiMAX, referred to by the standard 802.16e, adds mobility to the WiMAX specifications, such that seamless handover and roaming are possible when users move from one cell site area to another. The specifications for 802.16e standard were released in December 2005. PC data cards, mobile handsets and laptops with embedded WiMAX chips are being planned by vendors on this standard. In countries where regulation prohibits full mobility for alternative wireless technologies such as WiMAX, operators can also deploy 802.16e networks for fixed and nomadic access.

We analyzed what can be realistically expected from WiMAX deployments as well as compared its proposition against DSL (digital subscriber line) and 3G (third generation). Wireless Broadband Races to Substitute ADSL[1] For many years, users around the world have relied on fixed Internet connections, from the humble beginnings of dial-up to more generous portions of bandwidth though broadband connectivity. Recently, the emergence of wireless broadband has begun to challenge the landscape of fixed broadband. Though during its infancy stage, wireless broadband was regarded as a complimentary technology to empower mobile broadband, a sector outside the service perimeters of fixed line broadband operators, the scene has now changed to a competing one. Especially in emerging markets, wireless broadband technologies such as WiMAX are now in direct competition with fixed (ADSL) operators, where wireless broadband is positioned for fixed, indoor use that caters for home and small office users. WiMAX holds this advantage as it began as a fixed wireless broadband connectivity (IEEE's 802.16d standard). According to Qualcomm, the year 2010 will see the number of wireless broadband subscribers overtake fixed broadband subscribers as shown in Figure 1 below.



Source: Qualcomm  
Fig 1: Rapid growth of global wireless broadband subscribers

## 2. Characteristics of Emerging Markets



The definition of an emerging market [4] is a nation having an economy with a very low current gross domestic product per capita (GDP) with an above-average economic growth potential. The annual GDP per capita for China and India for example is under \$1,000, whereas the United States, Japan, and countries in Western Europe have GDPs per capita ranging from \$24,000 to \$36,000 per year. The above-average growth potential in emerging markets makes these countries attractive for investment but the low current GDP creates one of the major initial challenges. In terms of broadband services the low discretionary income per household has the following impact:

- Lower revenues (ARPU) (average revenue per user) for broadband services.
- Fewer customers can afford to purchase their own customer premise equipment.
- Higher churn and higher percentage of bad debts can result in higher operating expense
- Lower percentage of households own personal computers thus reducing the size of the addressable market for broadband services.

On a more positive note there are a number of favorable attributes in addition to the above-average economic growth that make these markets particularly attractive for communications network investment. These attributes are summarized in the following table

Table 1: Characteristics & Impact

Characteristics of Emerging Markets	Impact on WiMAX Operator
Support of government telecom regulators	<ul style="list-style-type: none"> <li>• Spectrum available at low or no cost</li> <li>• Facilitated licensing process</li> </ul>
Very high household (HH) density in metro areas	<ul style="list-style-type: none"> <li>• Lower infrastructure CAPEX (capital expenditure) per HH passed</li> </ul>
Limited wire-line competition	<ul style="list-style-type: none"> <li>• Gain higher penetration of addressable market</li> </ul>
High pent-up demand	<ul style="list-style-type: none"> <li>• Rapid market adoption rate (1 to 2 years instead of 3 to 5 years)</li> </ul>

### 3. WiMAX Technology Forecast

Wireline technologies are slow and costly to roll out - even in some parts of developed nations. Cellular technology is often too costly to use, does not deliver true broadband speed and does not scale to the capacity of an all-IP media-centric network. Therefore it is assumed that, throughout the forecast period, particularly aggressive WiMAX growth [2] will take place in countries such as Brazil, China, India and Russia; and in regions such as the Americas, Middle East/Africa, Eastern Europe and Developing Asia Pacific.

Initial forecasting assumptions are based on current penetration levels and potential total penetration levels, which take into account current and future economic development potential in each world region.

The WiMAX penetration rates in these forecasts vary significantly by region and are based on the following assumptions:

- After launching of WiMAX services the market potential depends on the availability of suitable spectrum in each region.
- WiMAX penetration will increase as equipment costs—and particularly device costs—decrease, with the rate of penetration in each region depending on the wider broadband market (with compared to other broadband devices) as well as macroeconomic factors such as consumer purchasing power
- WiMAX will have higher growth and penetration rates where penetration of alternative fixed and mobile broadband systems is low.
- WiMAX will have higher growth rates in regions where major operators are already committed to deploying the technology. We are talking to those operators where large number of users migrates to WiMAX.
- WiMAX penetration will increase as service costs decrease, with the exact rate depending on the wider broadband and economic landscape of each region.
- WiMAX penetration rates in each region have been benchmarked against comparable historical penetration rates in the fixed broadband, mobile, and mobile broadband markets. More detail on these penetration rates will be available in future reports. In future forecast revisions our intention is to introduce a dual methodology that includes both a tops-down and a bottom-up approach based on actual deployment data. This will allow for growth assumptions to be tied more closely to the number and growth of national and major regional operators. User Growth Forecasts [8]

The WiMAX subscription model is same as of fixed broadband in that there are multiple business and consumer users connecting per each CPE subscription. The forecasts in Table 1 below take this into account and accordingly show a higher number of users than subscribers. Table 1 set out the user numbers by major world region.

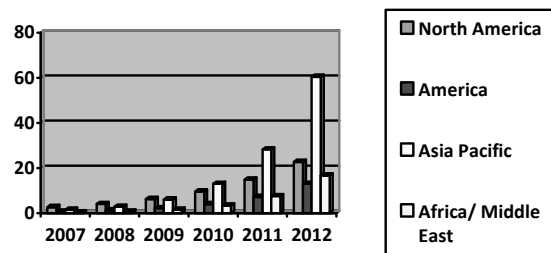


Fig 2: WiMAX Users by Region 2007-2012

Table 2: WiMAX Users by Region (millions) 2007-2012

Users = subscribers adjusted to reflect multiple users per subscription

Region	2007	2008	2009	2010	2011	2012
North America	2.61	4.03	6.25	9.59	14.79	22.62
Americas	0.66	1.18	2.14	3.92	7.17	12.97
Asia Pacific	1.39	2.84	5.99	12.96	28.17	60.45
Europe	1.35	2.34	4.07	7.08	12.23	21.01
Africa/Middle East	0.30	0.65	1.46	3.32	7.50	16.60
TOTAL	6.32	11.04	19.91	36.88	69.87	133.66

Fixed WiMAX device subscriptions—for example by outdoor or indoor CPE—will on average service more than one user. By 2012 the Asia Pacific region will lead the market in total actual users, with North America in second place followed by Europe, Africa/Middle East and the Americas. User numbers in India will overtake those in the USA in 2012, and it is estimated that by then China will have almost as many users as the whole of the Americas region (Latin America & the Caribbean).

#### 4. WiMAX: Country Growth & Operator

The numbers of WiMAX operators and countries shown in Figure 3 are those in which WiMAX service has commenced. Those currently in deployment but not yet implemented in their account for the forecasts, with other operators and status which will adopt WiMAX technology in future.

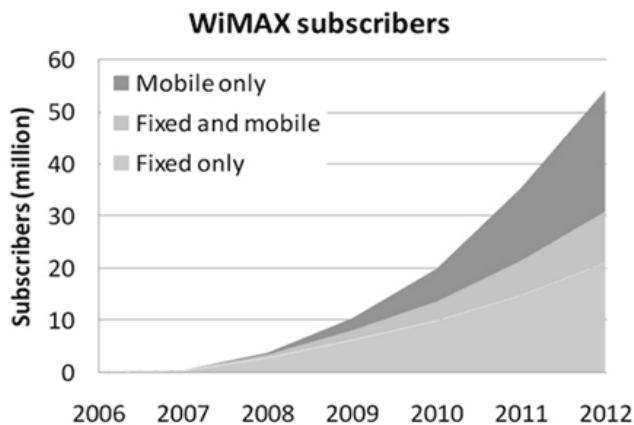


Fig 3: WiMAX Subscribers 2006-2012

The end of 2007 showed a total of 181 WiMAX operators globally. This number is expected to rise to 538 operators by 2012. The number of countries with WiMAX is anticipated to rise from 94 (out of a total 234 countries) at the end of 2007[5] to 201 in 2012.

Europe is anticipated to have the largest number of operators, followed by Asia Pacific, Africa/Middle East, Americas and North America. However, Africa/Middle East is expected to have the highest number of WiMAX operator countries, followed by Europe, Americas, Asia Pacific and North America.

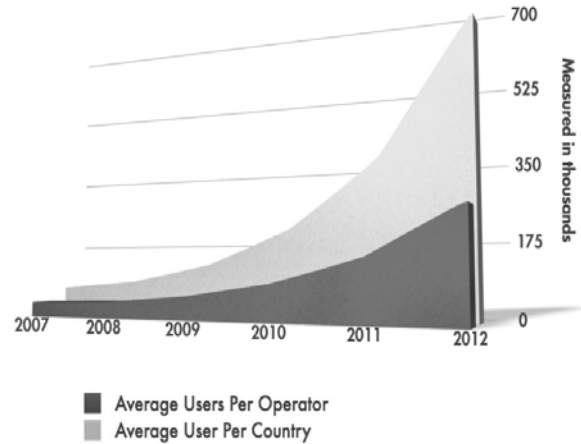


Fig 4: Average WiMAX Users by Operator & Country 2007-2012

#### Competing technologies

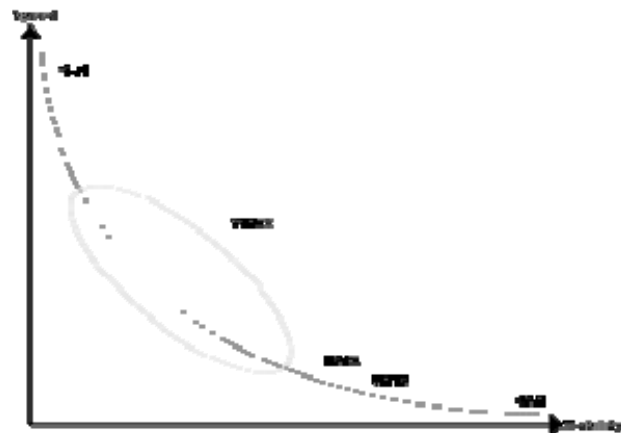


Fig 5: Speed vs. Mobility of wireless systems

Wi-Fi, HSPA (high speed downlink packet access), UMTS (universal mobile telecommunication system), and GSM (global system for mobile communication) within the marketplace, WiMAX's main competition comes from existing, widely deployed wireless systems such as UMTS, CDMA2000 (code division multiple access), existing Wi-Fi and mesh networking.

In the future, competition [7] will be from the evolution of the major cellular standards to so-called 4G, high-

bandwidth, low-latency, all-IP networks with voice services built on top. The worldwide move to 4G for GSM/UMTS is the 3GPP (third generation partnership project) Long Term Evolution effort. However, it has been noted that the likely performance difference between WiMAX as it stands today and LTE (long term evolution) [6] when it is eventually commercially available in 2–3 years time, will be negligible.

LTE is expected to be ratified at the end of 2010, with commercial implementations becoming viable within the next two years. End of 2009 TeliaSonera started commercial deployment in Oslo and Stockholm, In Denmark the 3 big telecoms are upgrading their network, and will make LTE available during 2010.

In some areas of the world, the wide availability of UMTS and a general desire for standardization has meant spectrum has not been allocated for WiMAX: in July 2005, the EU-wide frequency allocation for WiMAX was blocked.

## 5. Conclusion

The purpose of this forecast is to provide the WiMAX Forum prediction of the ecosystem's worldwide growth over the next five years. The forecast covers WiMAX deployments globally and is broken down by major regions – North America, Asia-Pacific, Europe, and Middle East/Africa. This also includes major country or sub-regional breakouts for the USA (united state of America), Canada, Japan, China, Korea, India, the Rest of Asia-Pacific Developed, the Rest of Asia-Pacific Developing, Western Europe, Eastern Europe, Africa and the Middle East.

### Assumptions

Worldwide access to Broadband Internet is vital for economic growth and development. All governments must work to ensure that their nations are able to realize the benefits associated with a strong communications infrastructure. Therefore this report assumes that many countries will adopt WiMAX as a wireless Broadband Internet technology to facilitate rapid economic development. It is also assumed that the move to WiMAX, a technology that is ready for deployment now, will be preferable to waiting for alternative technologies that may not be available for three or more years.

We can assume the growth of WiMAX technology, because we have seen the results the other related technologies by rapid growth, WiMAX user growth, the worldwide WiMAX operator growth, average WiMAX user by operator and country 2007 to 2012 and finally other competing technology of Fig 1, Fig 2, Fig 3, Fig 4 and Fig 5 respectively .

So, we can conclude that WiMAX's operator and product has a vital role for country and their operators' growth to cost-effective reach million of traditional voice and broadband data services.

## Reference

- [1] Wimax in emerging markets, monica paolini trendsmediatelebriefing, april 19, 2006

- [2] <http://www.slideshare.net/greenpacket/wi-max-a-wireless-solution-for-Fixed-Wireless-access-in-emerging-markets>
- [3] WiMAX: The Last Mile Winner? Telecom & Media Insights Issue 12, January 2006
- [4] WiMAX: The Business Case for Fixed Wireless Access in Emerging Markets June 2005
- [5] WiMAX Forum WiMAX Technology Forecast (2007-2012)
- [6] WiMAX An Efficient Tool To Bridge The Digital Divide, Guy Cayla, Stephane Cohen and Didier Guigon (on behalf of WiMAX Forum)
- [7] [http://en.wikipedia.org/wiki/Wimax#WiMAX\\_Forum](http://en.wikipedia.org/wiki/Wimax#WiMAX_Forum)
- [8] WiMAX Forum WiMAX Technology Forecast (2007-2012)

# Simulation and Optimization of MQW based optical modulator for on chip optical interconnect

Sumita Mishra<sup>1</sup>, Naresh k. Chaudhary<sup>2</sup> and Kalyan Singh<sup>3</sup>

<sup>1</sup> E & C Engineering Department, Amity School of Engineering and Technology, Amity University  
Lucknow, Uttar pradesh, India

<sup>2</sup> Electronics Department, Dr. RML Awadh University  
Faizabad, Uttar pradesh, India

<sup>3</sup> Electronics Department, Dr. RML Awadh University  
Faizabad, Uttar pradesh, India

## Abstract

Optical interconnects are foreseen as a potential solution to improve the performance of data transmission in high speed integrated circuits since electrical interconnects operating at high bit rates have several limitations which creates a bottleneck at the interconnect level. The objective of the work is to model and then simulate the MQWM based optical interconnect transmitter. The power output of the simulated modulator is then optimized with respect to various parameters namely contrast ratio, insertion loss and bias current. The methodology presented here is suitable for investigation of both analog and digital modulation performance but it primarily deals with digital modulation. We have not included the effect of carrier charge density in multiple quantum well simulation.

**Keywords:** *Optical interconnect, MQW, CR (contrast -ratio), IR (insertion loss).*

## 1. Introduction

The microprocessor industry has developed at an incredible pace, particularly in the past decades. Transistor scaling has been the crux of the rapid growth in processing power over the past forty years [1]. Scaling process has a large impact on electrical parameters of metallic interconnections which are responsible for transporting data within the microprocessor and between the microprocessor and memory and consequently, interconnect has become the dominant factor determining speed. Two fundamental interconnection limits encountered as the density of transistors increase one related to speed and the other to the number of input/output channels. Consequently as integrated circuit technology continues to scale if the interconnect problem

is not addressed it will not be possible to achieve the exponential speed increases we have come to expect from the microprocessor industry. Optical interconnects have the potential to address this problem by providing both greater bandwidth and lower latency than electrical interconnects. Advantages offered by optical interconnects provide strong motivation to further develop methodologies for analysing optical interconnect links.

There have been several attempts at optimizing optical interconnect links using software tools such as Microsim P-Spice. [3-6]. P-Spice is designed for EDA and is not optimized for optical networks hence at very high frequencies precision of simulated circuit reduces. Thus for analysing the behaviour of high speed optical interconnects MATLAB and Simulink may be a more powerful tool since it offers multi-domain simulation environment and model-based design which can accurately model the behaviour of optical sub systems making it a good platform for optimization of optical interconnect link.

## 2. Background

For optical transmitters, VCSELs and MQWMs are the two primary optical sources for high density optical interconnects. However VCSEL's use is limited due to self-heating and device lifetime concerns [5].

Quantum-well modulators have so far been the devices most extensively used in demonstrating actual dense interconnects to and from silicon CMOS chips. [6,9]

These devices have successfully been made in large arrays and solder bonded to the circuits. Also, Multiple quantum well (MQW) modulators offer an advantage over other light emitters in terms of signal and clock distribution. Furthermore, the electrical signals can be sampled with short optical pulses to improve the performance of receivers. MQWM based link requires that an external beam be brought onto the modulator. This facilitates to generate and control one master laser beam which allows centralized clocking of the entire system, and the use of modulators, as described above, allows the retiming of signals, especially if the master laser operates with relatively short optical pulses. Thus QWM based approach, besides yielding lower transmitter on-chip power dissipation can be more conducive to monolithic integration. This was the motivation for simulating a MQWM based optical interconnect link.

### 3. Modeling and Simulation methodology

In this section we describe the methodology used for modelling and simulation of optical interconnect transmitter.

The simulated laser diode is an InGaAs–Al–GaAs–GaAs quantum-well separate confinement heterostructure. We considered only the internal parasitics assuming a low-parasitics assembly scheme. The simulated modulator structure is reflective mode (RMQWM).

For simulation of the dynamic response of MQW laser a rate equation model has been used [7]. In this model we have not included the effect of carrier dynamics in the quantum wells yielding the following set of equations

$$\frac{dN(t)}{dt} = \frac{I(t)}{qV_a} - \beta_0 \frac{N(t) - N_0}{1 + \alpha S(t)} S(t) - \frac{N(t)}{\tau_n} \quad (1)$$

$$\frac{dS(t)}{dt} = \beta_0 \Gamma \frac{N(t) - N_0}{1 + \alpha S(t)} S(t) - \frac{S(t)}{\tau_p} + \Gamma \beta \frac{N(t)}{\tau_n} \quad (2)$$

$$\frac{d\phi}{dt} = \frac{\alpha}{2} \left[ \Gamma \beta_0 [N(t) - N_0] - \frac{1}{\tau_p} \right] \quad (3)$$

With

$$\beta(t) = \frac{S(t)\eta/h\nu V_a}{2\Gamma\tau_p} \quad (4)$$

Where N(t) is a the carrier density in the in the quantum wells, S is the photon density in the laser cavity,  $\phi$  is the

phase of the optical field, I is the injection current, q is the electronic charge,  $N_0$  is the carrier density in the quantum wells for the reference bias level, p is the power output .physical meaning and values of various other coefficients can be found in ref [7] . Simulated Laser power output was then fed to the modelled integrated surface-normal reflective electroabsorption mqw modulators. Quantum well absorption data for three quantum wells is taken from the literature for well width of 95 Å, and the Al0.3Ga0.7As barrier thickness of 30 Å. An electroabsorption modulator using the quantum-confined Stark effect is formed by placing an absorbing quantum well region in the intrinsic layer of a pin diode. Doing so creates the typical p-i-n photodiode structure and enables large fields to be placed across the quantum wells without inducing large currents. By applying a static reverse bias across the diode, photogenerated carriers are efficiently swept out of the intrinsic region and the device acts as a photodetector. Varying this bias causes a modulation in the optical absorption, resulting in an optical modulator. The modulator is characterized by its capacitance, Insertion Loss and Contrast Ratio. An ideal modulator has minimum optical power loss during the "on" state (IL), and largest possible optical power ratio between the "on" and the "off" states (CR). Typically, there is a trade-off between these parameters for a given value of the ratio between maximum ( $\alpha_{max}$ ) and minimum ( $\alpha_{min}$ ) absorption . The IL/CR relation for a simple RMQW structure in a reverse biased PIN configuration is given below

$$CR = \frac{R_{on}}{R_{off}} = \frac{e^{-\alpha_{min}l}}{e^{-\alpha_{max}l}} = (1 - IL)^{1-\eta} \quad (5)$$

Here  $R_{on}$  and  $R_{off}$  are the modulator reflectivities in the less absorbing and more absorbing states, respectively. CR decreases significantly at low operating voltages.

The modulator power output consists of the dynamic component including the capacitance of the driver chain and the modulator and the static component due to the absorbed optical power in the “on” and the “off” state. At low voltages, the dynamic component is small. The static power is calculated in terms of the CR and IL by multiplying the current in each binary state by its respective voltages and taking an average [9]

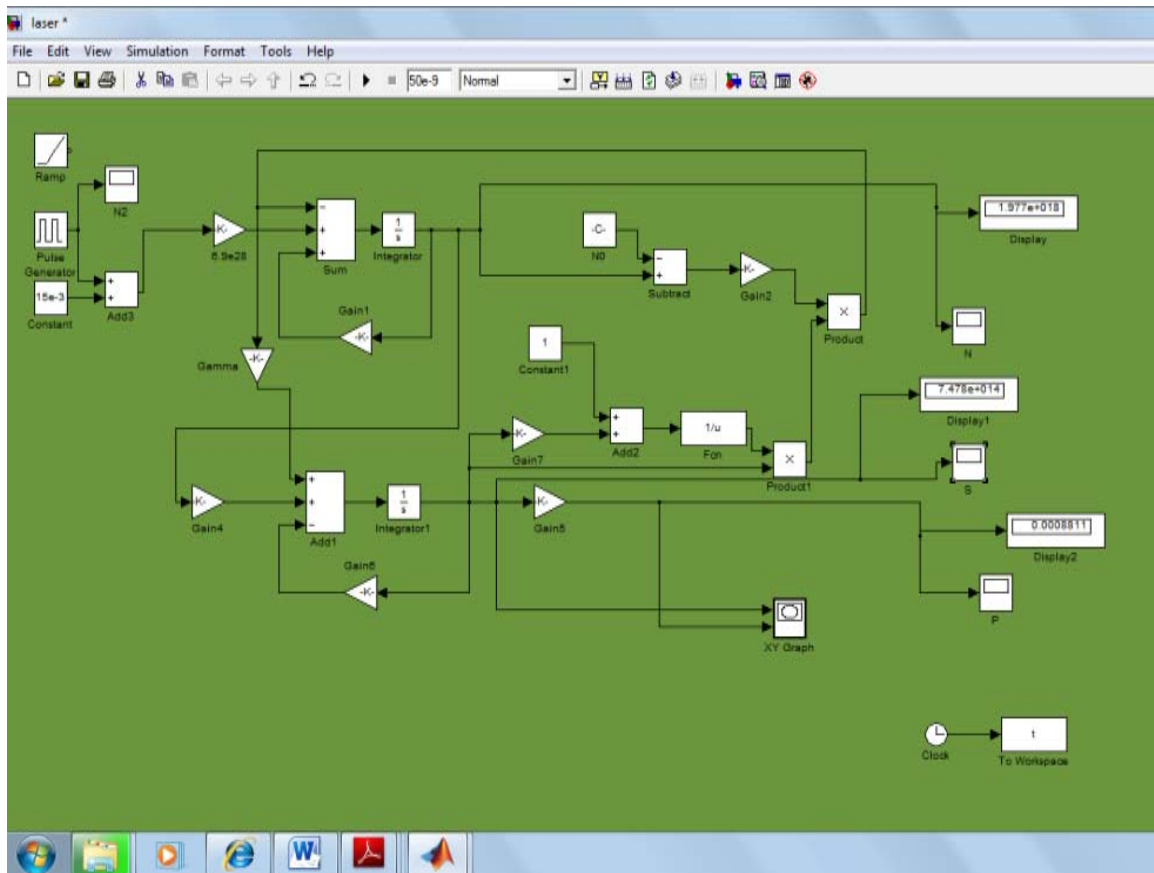
$$P = .5R_s \text{ mod} P_i [IL(V_{bias} - V_{dd}) + (1 - \frac{1-IL}{CR})V_{bias}] = \eta_{mod} P_i \quad (6)$$

Here  $\eta_{mod}$  is a dimensionless efficiency factor  $R_s$  is the modulator responsivity  $P_i$  is the input laser power to the modulator,  $V_{bias}$  the pre-bias voltage and  $V_{dd}$  is the supply voltage small compared to the static power of the modulator.

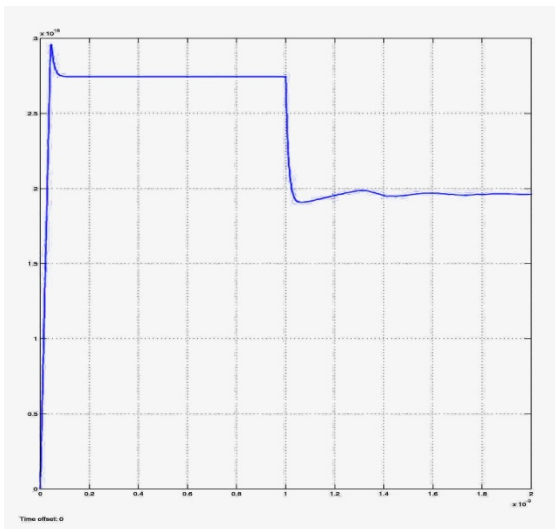
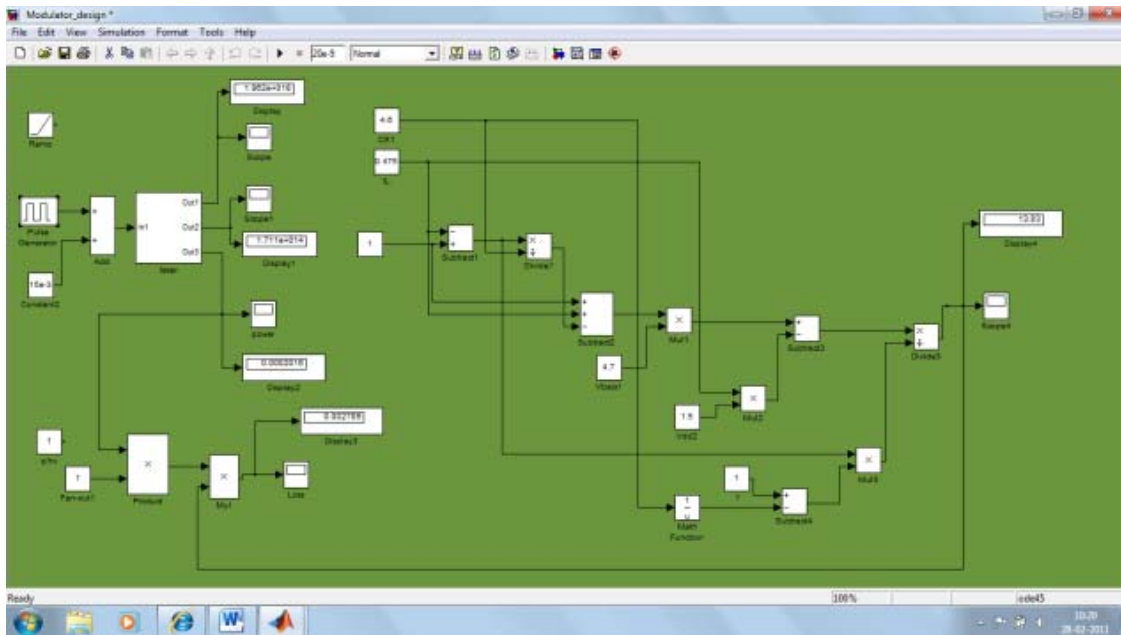
#### 4. Model description and Results

Simulation was carried out in two stages. In the first stage the rate equations were implemented in simulink as shown in fig -1. Laser power output was then coupled to external modulator. Simulink model of MQWM modulator is shown in fig 2. Simulated Laser diode photon density for 1ns pulse is shown in Figure 3. The simulated power

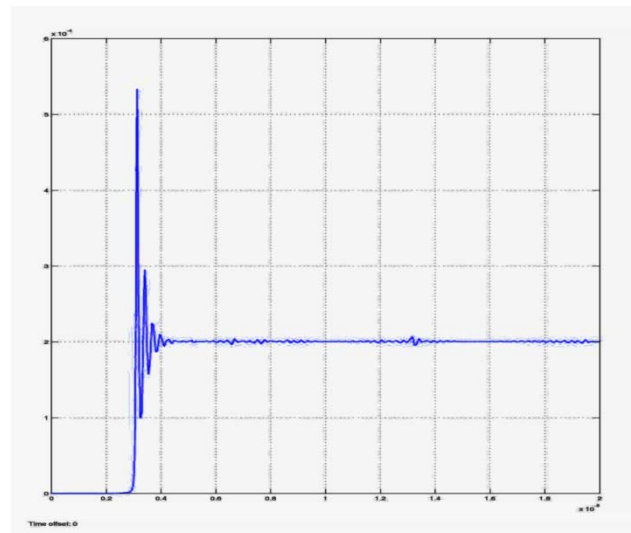
output response of MQWM modulator is shown in fig 4. Simulated optical photon density output of MQWM Modulator with ramp input and bias current=2mA is shown in fig 5. Minimum interconnect power is observed as a function of bit rate. We further study the change in the minimum interconnect power as a function of parameter X, which is dictated by bias current. It was observed that response of model worsens with increase in bias current. We have not included the effect of pattern jitters and crosstalk. All the simulations were run over a time period that was several orders of magnitude longer than the fixed step size chosen so that turn-on transient effects that happen near threshold can be avoided. All simulations were carried out using standard 4th-order Runge-Kutta algorithm with a fixed step size.







**Fig.3 Simulated photon density of Laser with bias current=1mA**



**Fig.4 Power Dissipation of MQWM**



**Fig. 5 Simulated photon density of MQWM Modulator with ramp input and bias current=2mA**

## 5. Conclusions

The work describes a methodology to model, simulate and then optimize the MQWM based optical interconnect transmitter power output with respect to various parameters namely contrast ratio, insertion loss and bias current. The methodology presented here is suitable for investigation of both analog and digital modulation performance but it primarily deals with digital modulation. The modulator was simulated on MATLAB Simulink tool and model response was obtained for 1- 20Gbps bit rate. The simulated model can achieve error-free operation under 16 Gbps data rate. It was observed that Modulator output worsens with increase in bias current. These results are based on simplified cases excluding pattern jitters, crosstalk and the effect of carrier charge density in multiple quantum well. However, the effect of pattern jitters and bandwidth limits of each device will become increasingly important as the density of an interconnect array becomes higher. These are subjects for further study. The model can be further improved by addressing these issues.

## Acknowledgments

The first author Sumita Mishra is grateful to Maj. Gen. K.K. Ohri , Prof. S.T.H. Abidi and Brig. U. K. Chopra of Amity University, Lucknow for their support during the research work.

## References

- [1] David A. B. Miller "Rationale and Challenges for Optical Interconnects to Electronic Chips" PROCEEDINGS OF THE IEEE, VOL. 88, NO. 6, JUNE 2000.
- [2] Daniel S. Chemla, David A B Milar, Peter W Smith "Room Temperature Excitonic Non linear absorption and Refraction in GaAs/AlGaAs Multiple Quantum Well Structure" IEEE JOURNAL OF QUANTUM ELECTRONIC, VOL -QE-20, NO.3 MARCH 1984.
- [3] Samuel Palermo, Azita Emami-Neyestanak,, and Mark Horowitz, "A 90 nm CMOS 16 Gb/s Transceiver for Optical Interconnects" IEEE JOURNAL OF SOLID-STATE CIRCUITS, VOL. 43, NO. 5, MAY 2008.
- [4] Azad Naeemi, E, Reza Sarvari, and James D. Meindl "Performance Comparison Between Carbon Nanotube and Copper Interconnects for Gigascale Integration" IEEE ELECTRON DEVICE LETTERS, VOL. 26, NO. 2, FEBRUARY 2005.
- [5] Y. Liu et al, "Numerical investigation of self-heating effects of oxide confined vertical-cavity surface-emitting lasers," IEEE J. of Quantum Electron., Vol. 41, No. 1, pp. 15-25, Jan. 2005.
- [6] O. Kibar, D. A. A. Blerkon, C. Fan, and S. C. Esener, "Power minimization and technology comparison for digital free-space optoelectronic interconnects," J. Lightw. Tech., vol. 17, no. 4, pp. 546-555, Apr. 1999.
- [7] A. Javro & S.M. Kang, transforming Tucker's Linearized Laser Rate Equations to a Form that has a Single Solution regime," Journal of Lightwave Technology, vol.13, No.9, pp.1899-1904, September 1995.
- [8] A. V. Krishnamoorthy and D. A. B. Miller, "Scaling optoelectronic- VLSI circuits into the 21st century: A technology roadmap," IEEE J. Select. Topics Quantum Electron., vol. 2, pp. 55-76, Apr. 1996.
- [9] Hoyeol Cho, Pawan Kapur, and Krishna C. Saraswat "Power Comparison Between High-Speed Electrical and Optical Interconnects for Interchip Communication" JOURNAL OF LIGHTWAVE TECHNOLOGY, VOL. 22, NO. 9, SEPTEMBER 2004.
- [10] J. J. Morikuni, A. Dharchoudhury, Y. Leblebici, and S. M. Kang "Improvements to the Standard Theory for Photoreceiver Noise" JOURNAL OF LIGHTWAVE TECHNOLOGY, VOL. 12, NO. 4, JULY 1994.
- [11] Kyung-Hoae Koo, Hoyeol Cho, Pawan Kapur, and Krishna C. Saraswat, "Performance Comparisons Between Carbon Nanotubes, Optical, and Cu for Future High-Performance On-Chip Interconnect Applications" IEEE TRANSACTIONS ON ELECTRON DEVICES, VOL. 54, NO. 12, DECEMBER 2007
- [12] C. L. Schow, J. D. Schaub, R. Li, J. Qi, and J. C. Campbell, "A 1-Gb/s Monolithically Integrated Silicon NMOS Optical Receiver" IEEE JOURNAL OF SELECTED TOPICS IN QUANTUM ELECTRONICS, VOL. 4, NO. 6, NOVEMBER/DECEMBER 1998.

- [13] H. Zimmermann, and T. Heide “A Monolithically Integrated 1-Gb/s Optical Receiver in 1-micron CMOS Technology” IEEE PHOTONICS TECHNOLOGY LETTERS, VOL. 13, NO. 7, JULY 2001.
- [14] Hoyeol Cho, Kyung-Hoae Koo, Pawan Kapur, and Krishna C. Saraswat, “Performance Comparisons Between Cu/Low- $\kappa$ , Carbon-Nanotube, and Optics for Future On-Chip Interconnects” IEEE ELECTRON
- [15] Osman Kibar, Daniel A. Van Blerkom, Chi Fan, and Sadik C. Esener, “Power Minimization and Technology Comparisons for Digital Free-Space Optoelectronic Interconnections” JOURNAL OF LIGHTWAVE TECHNOLOGY, VOL. 17, NO. 4, APRIL 1999.
- [16] Emel Yuceturk, Sadik C. Esener “Comparative study of very short distance electrical and optical interconnects based on channel characteristics” IEEE MICRO TRENDS IN OPTICS AND PHOTONICS SERIES, VOL.90, 17, pp. 48-56, 1997.



**Sumita Mishra** has done her post-graduation in Electronics Science from Lucknow University, Lucknow. She has also done M. Tech (Optical Communication) from SGSITS, Indore in 2004. Thereafter, she was appointed as a lecturer in Electronics and Communication Engineering Department at Amity School of Engineering and Technology (New Delhi). Currently, She is working as a lecturer in ECE department at Amity University (Lucknow campus), she is also pursuing doctoral degree in Electronics at DRML Awadh University, India. She is a member of IEC, Oxford Journals, ABI Research, Transmission & Distribution World IEEE, ACM, IEEE (institutional membership), PCPro, IACSIT and VLSI Jagrati. Her current research interests include Fibre Optic CDMA, Optical interconnects and Machine vision. Her research papers (6) have been presented in various IEEE international and National conferences. She has two publications in international journals.

**Naresh K Chaudhary** has done his Ph.D in Physics from Lucknow University, Lucknow in 2005 , Thereafter, he joined as a lecturer in Institute of Engineering and Technology Resura Sitapur UP INDIA. Where he worked till November 2006 .Then he was appointed as Assistant professor in Department of Electronics and Physics at Dr RML Avadh University Faizabad in Nov 2006. He has seven publications in various national and international journals.

**Kalyan singh** Dr Kalyan Singh is Professor and head Department of Physics and Electronics Dr RML Avadh University Faizabad. He has 37 year of post Ph.D experience .He has guided 11 Research Scholars and published 26 papers at national and international level.

# Determination of the Complex Dielectric Permittivity Industrial Materials of the Adhesive Products for the Modeling of an Electromagnetic Field at the Level of a Glue Joint

Mahmoud Abbas<sup>1</sup>, Mohammad Ayache<sup>2</sup>

<sup>1</sup>Department of Electronics, Lebanese University  
Saida, Lebanon

<sup>2</sup>Department of Biomedical, Islamic University of Lebanon  
Khalde Highway, Lebanon

## Abstract

To achieve out this work we were interested in the study of microwaves techniques and also of the measurement of complex relative dielectric permittivity. It is important to measure this dielectric permittivity for the used adhesive before subjecting it to electromagnetic energy.

This prediction enables us to avoid an exothermic phenomenon due to the brutal rise in the temperature in the joint of adhesive. However, these results are used calculations program, to trace cartography of the electric field and of a temperature gradient in standardized test-tubes. At the end of this step, we have physical and experimental tools that can be used in the study of an optimized process using the microwaves. We check also the strong absorption of energy on the level of joint of adhesive (attenuation electric field), that the microwaves make it possible to well polymerize the adhesives with less times and low energy consumption without rise in prejudicial temperature of the parts to be stucked.

**Keywords:** Microwave, Coaxial line, dielectric, permittivity, glue, adhesive.

## 1. Introduction

The dielectric heating concerns dielectric body, the body that is bad electrically conductive is generally bad driver of heat. In general, such a body contains molecules or groups polar. These charges tend to align with the electric field within the material. In the case where an electric field at low frequency is imposed, alignment can occur with a lag which is a loss of electromagnetic energy and thus heating of the material. The choice of the working frequency is regulated to avoid interfaces with telecommunications; some bands are released for industrial, scientific and medical use (ISM). The interaction of electromagnetic waves and materials transforms electromagnetic energy into thermal energy which is reflected in both the ionic conductivity and

dielectric relaxation. Therefore, the dissipative properties and the complex properties of the materials are determined by the conductivity dielectric permittivity. Depending on their values, these factors characterize the absorbency of the product subjected to radiation.

When an electromagnetic wave comes into contact with a dielectric, a part of the wave is reflected and a part enters the material. The energy of transmission, in the sample to be treated, decreases exponentially transforming itself into heat. The attenuation factor depends on the physical characteristics of the local environment and the frequency and it is represented in the following equation [1].

$$E = E_0 e^{-\alpha x} \quad (1)$$

$E_0$  represents the amplitude of the field at the internal surface of the dielectric environment, the attenuation coefficient is given by [2, 3]:

$$\alpha = \frac{c}{2\pi \sqrt{\frac{\epsilon_r}{2}} (-1 + \sqrt{1 + tg^2 \Delta})} \quad (2)$$

$\alpha$  depends on the physical characteristics of the local environment and the frequency.  $\Delta$  is the angle losses and it is given by:

$$tg \Delta = \frac{\epsilon_r''}{\epsilon_r'}, \quad (3)$$

$C$  is the speed of the electromagnetic wave in vacuum; ( $C = 3 \times 10^8 \text{ms}^{-1}$ ),  $\epsilon_r'$  and  $\epsilon_r''$  are the real and imaginary parts of the complex relative permittivity.

## 2. Technical Results of Measurement

The automotive industry introduces more and more plastic in these fabrications. This material replaces the metal, due to its low density, its chemical qualities of neutrality with respect to a wide range of corrosive agents and ways to implement it in simple process for some manufacturing. The addition of the reinforcing fibers can obtain parts whose mechanical properties are comparably seen superior to those of metals. We are also interested in the timber industry where we have made standard specimens exhibiting remarkable mechanical properties, after the treatment by microwave energy. The operation of traditional collage, produced with the help of a conformer metal movement of warm air when heated by resistance, lasts at least 3 minutes, the time required for the rise in temperature of the glue, because this should be restricted to avoid the distortion of parts to be assembled. Today, the demands of the production rates of the automobile industry need to realize this bonding time to less than one minute. In recent years, the use of microwaves in the collage of wood and in the automotive industry was growing significantly, but many principles were still poorly understood, which justifies our work. This work is to study the behavior of industrial adhesives during the cross linking reaction, polymerization is activated by an electromagnetic wave. In order to achieve this work, it is necessary to know in advance the dielectric behavior as a function of frequency and temperature of the various used glues. Unexpected reactions may occur such as shifting towards maximum absorption frequencies or thermal caused by a sudden increase in permittivity as soon as the temperature rises.

The measurement technique is to use a coaxial standard radius cell from a section of coaxial guide  $50 \Omega$  and exterior radius  $b$  (Fig. 1). This cell is terminated by driver infinity and must be immersed in the product. The network analyzer is used to measure the coefficient of reflection module and phase which is linked complex dielectric permittivity ( $\epsilon_r'$  and  $\epsilon_r''$ ) [5].

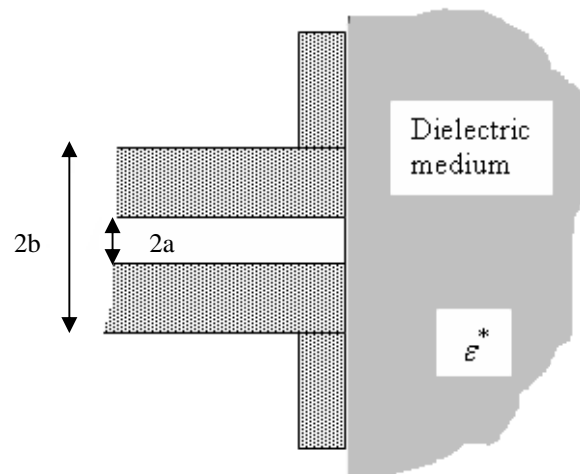


Fig 1: Coaxial Cell  $50 \Omega$

The main considered adhesives are: ALFO vinyl, epoxy glue FIX TOP 62 NA polyurethane varnish and glue XPU 4727 AC / BC, which have been developed and marketed by the company ELF ATOCHEM-France. For example, we present the results of measuring the dielectric permittivity of a polyurethane adhesive (XPU 4727 AC / BC). This is adhesive glue that can be used cold or hot. It is ideally suited for the assembly of structural composite materials made from polyester resins reinforced fiberglass, with thermoplastics. It is glue for the automobile industry, for a collage of body parts in SMC, BMC or pack. Figure 2 and Figure 3 show the curve of the glue XPU 4727 AC / BC according to the frequency and the temperature. We notice that the effect of temperature is dominant in high frequencies.

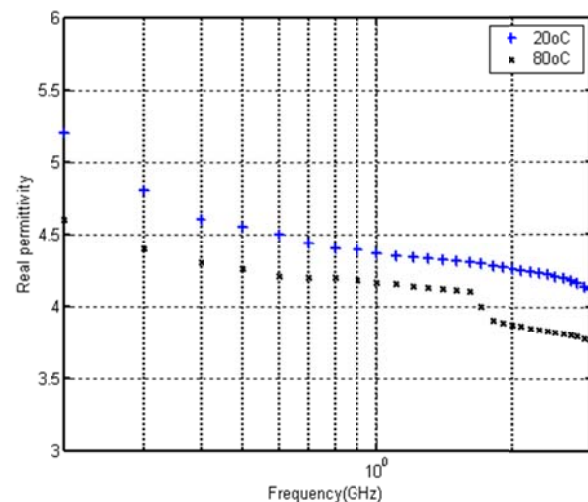


Fig 2: Real permittivity versus the frequency and the temperature



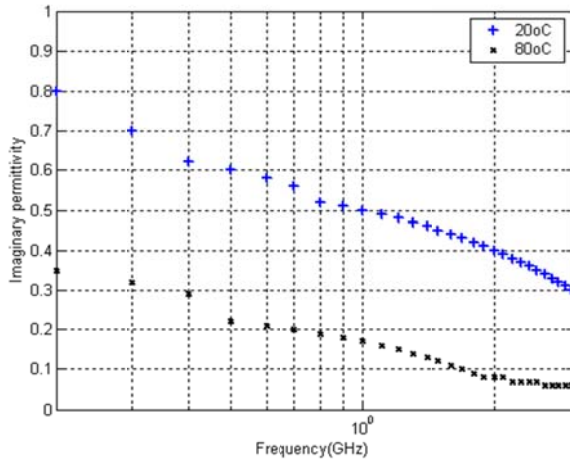


Fig 3: Imaginary permittivity versus the frequency and the temperature

The adhesive is fabricated using a mixture of equal volume of two components A and B, which results in a permittivity of 4.08- j0.30 at 2.45 GHz frequency. The permittivity of BMC, for example, at the same frequency is 4.3 - j 0.01. Figure 4 shows the temperature rise through the joint in terms of time. It has reached 150 ° C in less than 40 seconds. The measured temperatures in the socket BMC and glue highlights is one of the advantages of microwave, the maximum difference in temperature between the adhesive and support is 110 ° C at time t = 35 seconds [9]. These results are obtained from a test microwave containing:

- Power generator that can be adjusted from 0 to 1.2 kW, f = 2450 MHz.
- A pump for measuring the reflected power and protecting the generator reflected wave.
- An applicator consisting of a portion of a rectangular waveguide.

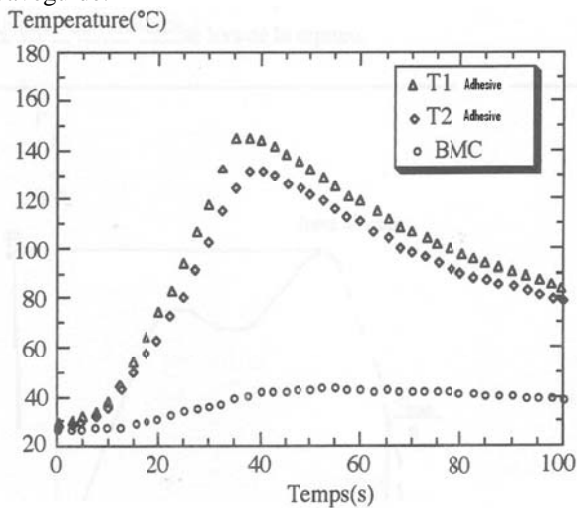


Fig 4: Evolution of the temperature through the joint glue and the BMC

Figure 4 shows, according to the curve of temperature mounted for different glued pieces that the warming of the joint adhesive is faster than the substrate. We can say that the parallel bonding field allows the reduction of processing time (electric field parallel to seal glue). This is an important point because production in a series, at a speed industry can be useful.

Tests of pure mechanical shearing, with a speed of 5 mm / min, depending on the extension, has been made on this standard test, 48 hours after bonding. The results are shown in the table below. By applying these tests, we noticed a breach of the bracket (BMC) with an average equal to 3.2 MPascal.

Table 1: Constraint observed on a series of tests of the adhesive polyurethane XPU 4727 AC/BC.

	Essay N°	Traction τ(MPascal)
Colle polyurethane ACBC 4727	1	3,0
	2	3,1
	3	3,0
	4	2,6
	5	3,0

### 3. Electromagnetic Field and Heat Modeling at Adhesive Joint [6, 7, 8]

It is often important to model the electric field and map temperature in the test tube before realizing an experimental microwave applicator. This allows us to optimize the energy absorbed by the sample. The development of a numerical model requires a physical analysis of the diffusion equation of heat and Maxwell's equations.

$$\vec{\nabla} \wedge \vec{E} = -\mu \frac{\partial \vec{H}}{\partial t} \quad (4)$$

$$\vec{\nabla} \wedge \vec{H} = \varepsilon \frac{\partial \vec{E}}{\partial t} + \sigma \vec{E} + \vec{J}_{Source} \quad (5)$$

$$\rho C_p \frac{\partial T}{\partial t} - \text{div}(\lambda \nabla T) = P_d \quad (6)$$



$\sigma$ : Electrical conductivity

$\rho$  : bulk density of the material

$C_p$ : Specific heat ( $J.g^{-1}.K^{-1}$ )

$\lambda$ : thermal conductivity ( $W.cm^{-1}.K^{-1}$ )

$P_d$  : Power absorbed by the material ( $W.cm^{-3}$ )

The energy is radiated by the antenna. We have a complex permittivity and a complex permeability. The source term is  $\cos(\omega t)$ ,  $\omega = 2 \pi f$  ( $f = 2.45$  GHz). This term source is placed at the  $\omega \pi t$ , with two locations of the antenna magnetron. Therefore, thermal and electromagnetic phenomena are coupled, firstly by the density of power which is a function of H and E, and secondly, by electrical, which depend on the temperature. The modeling of these equations, using the finite volume, determines the distribution of the electric field and the temperature during the process of heating of materials. We give as an example, the distribution of electric field into a vacuum microwave cavity, Figure 5, shows the electric field on the propagation mode TE013. In figures 6 and 7, we present the results obtained by modeling glue XPU 4727 AC / BC. The simulation of this joint glue was conducted using samples from BMC used in the automotive industry. We note, however, a reduction in the field slightly stronger in the joint adhesive. This is normal because the angle of loss is slightly lower for the BMC than for the glue XPU 4727.

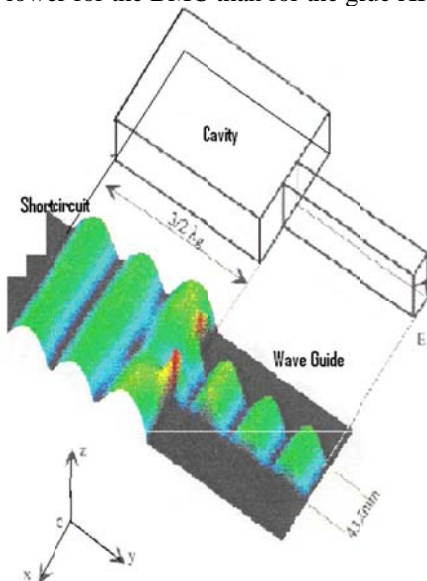


Fig 5: Cavity with vacuum, mode TE013, slice of the electric field in the yoz plan in 3D.

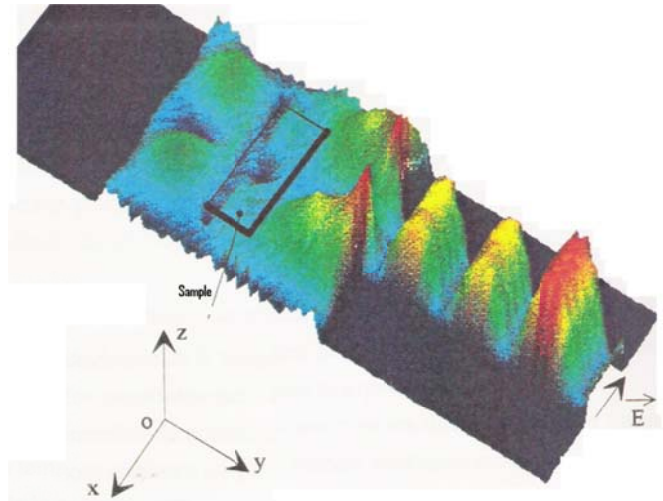


Fig 6: Cavity with sample, adhesive polyurethane XPU 4727 AC/BC, = 4, 08-j 0, 38 slice of the electric field in the xoy plan (in the joint of adhesive).

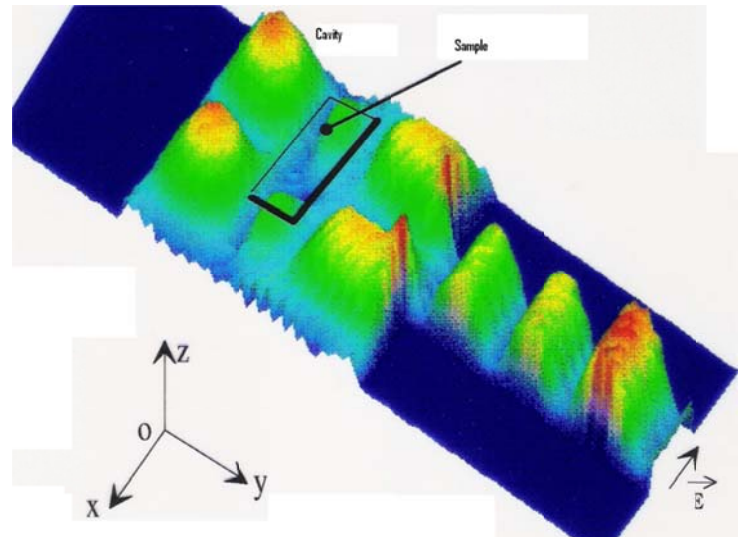


Fig 7: Cavity with sample, adhesive polyurethane XPU 4727AC/BC, slices electric field in the xoy plan (in the BMC).

## 4. Conclusion

By studying of the results produced through mechanical dimensioning of the applicator, we conclude that the couple modeling experimentation constitutes a solid basis and effective to apprehend the problems of bonding under microwave. We also check the strong absorption of energy at the level of glue seals (electric field attenuated) and the microwave can polymerize well the adhesives with reduced time and low energy consumption without fever parts pasting.

These results of measurement on the dielectric parameters will give us numerical data and give the thermal parameters using the Maxwell's equations and the equation

of the conduction of heat. This is connected to model the distribution of electromagnetic field and the topography of the temperature in a joint of adhesive. Finally, these thermal and dielectric parameters of the adhesive products give us a quite precise idea on the choice of an applicator microwaves.

## References

- [1] G.B. GAJDA, S.S. STUCHLY, "Numerical Analysis of Open-Ended Coaxial Lines", IEEE Trans. On Microwave Theory and Techniques, vol. MTT-31, N°5, May 1983.
- [2] MARIA A. STUCHLY, MICHAEL M. BRADY, G.GAJDA, "Equivalent circuit of an open-ended coaxial line in a lossy dielectric", IEEE Trans. Instrum. Meas., vol. IM-31 N°2, June 1982.
- [3] M.ABBAS, P.A. BERNARD, Cl. MARZAT, Microwave bonding in continuous devices for reprocessing little of wood and converting them into solid pieces. EUROCOAT 1994, Sitges (Barcelone), vol.1, pp. 99, 27-30 Septembre 1994.
- [4] M.ABBAS, P.A. BERNARD, Cl. MARZAT, Collage industriel par micro-ondes- diélectrique des colles en fonction de la fréquence et de la température. "Matériaux & Techniques", N° 10-11, pp.9, 1994.
- [5] PATANKAR S.V., Numerical heat transfer and fluid flow. Hemispher. 1980 New York.
- [6] M. ABBAS, P.A. BERNARD, Cl. MARZAT, Microwave bonding in continuous devices for processing little of wood and converting them into solid pieces. Double liaison, physique et chimie des peintures et adhesives- N°466-1994.
- [7] M. ABBAS, P.A. BERNARD, Cl. MARZAT, B.HAMDOUN, Modélisation électromagnétique et thermique d'un micro-ondes à 2054 GHz afin d'optimiser la répartition au niveau du joint de colle. Matériaux et techniques, N° 10-11-12, PP.27,2003.
- [8] M. ABBAS, B.HAMDOUN, Measurement of complex permittivity of adhesive materials using a short open-ended coaxial line probe, journal of microwave and optoelectronics, volume3, number6, October 2004.
- [9] M. ABBAS, J. CHARARA, Thermal characterization of industrial adhesives used to glue composite materials and wood by microwaves at 2.45 GHz, Matériaux et techniques 94, PP.165-169,2006.

**Mahmoud Abass** received the Ph.D degree in Electronics from the University of Bordeaux, France, in 1995. His research interests include modeling and optimization of microwave devices and electronic circuits.

**Mohammad Ayache** received the Ph.D degree in medical Image Processing from the University of Tours, France, in 2007. He is the coordinator of the department of biomedical at the faculty of engineering at the Islamic University of Lebanon. His research interests include advanced neural networks software development and advanced signal and image processing techniques.

# Power Aware Routing in Wireless Sensor Network

Rajesh Sahoo<sup>1</sup>, Satyabrata Das<sup>2</sup>, D.P.Mohapatra<sup>3</sup> & M.R.Patra<sup>4</sup>

<sup>1</sup>Department of Computer Science & Engineering  
Ajay Binay Institute of Technology, Cuttack, Odisha, India

<sup>2</sup>Department of Computer Science & Engineering  
College of Engineering, Bhubaneswar, Odisha, India

<sup>3</sup>Department of Computer Science & Engineering  
National Institute of Technology, Rourkela, Odisha, India

<sup>4</sup>Department of Computer Science & Engineering  
Berhampur University, Berhampur, Odisha, India

## Abstract

*The efficient node-energy utilization in wireless sensor networks has been studied because sensor nodes operate with limited battery power. To extend the lifetime of the wireless sensor networks, we reduced the node energy consumption of the overall network while maintaining all sensors balanced node power use. Since a large number of sensor nodes are densely deployed and interoperated in wireless sensor network, the lifetime extension of a sensor network is maintained by keeping many sensor nodes alive. In this paper, we submit power aware routing protocol for wireless sensor networks to increase its lifetime without degrading network performance. The proposed protocol is designed to avoid traffic congestion on specific nodes at data transfer and to make the node power consumption widely distributed to increase the lifetime of the network. The performance of the proposed protocol has been examined and evaluated with the NS-2 simulator in terms of network lifetime and end-to-end delay.*

*Keywords: wireless sensor networks, power aware routing protocol, NS-2*

## 1. Introduction

A wireless sensor network is one of the ad hoc wireless telecommunication networks, which are deployed in a wide area with tiny low-powered smart sensor nodes. An essential element in this environment, this wireless sensor network can be utilized in a various information and telecommunication applications. The sensor nodes are small smart devices with wireless communication capability, which collects information from light, sound, temperature, motion, etc., processes the sensed information and transfers it to other nodes.

A wireless sensor network is typically made of many sensor nodes for sensing accuracy and scalability of sensing areas. In such a large scale of networking environment, one of the most important networking factors are self-organizing capability for well adaptation of dynamic situation changes and interoperating capability between sensor

nodes [1]. Many studies have shown that there are a variety of sensors used for gathering sensing information and efficiently transferring the information to the sink nodes.

The major issues of such studies are protocol design in regards to battery energy efficiency, localization scheme, synchronization, and data aggregation and security technologies for wireless sensor networks. In particular, many researchers have great interest in the routing protocols in the network layer, which considers self-organization capabilities, limited battery power, and data aggregation schemes [2, 3].

A wireless sensor network is densely deployed with a large number of sensor nodes, each of which operates with limited battery power, while working with the self-organizing capability in the multi-hop environment. Since each node in the network plays both terminal node and routing node roles, a node cannot participate in the network if its battery power runs out. The increase of such dead nodes generates many network partitions and consequently, normal communication will be impossible as a sensor network. Thus, an important research issue is the development of an efficient battery-power management to increase the life cycle of the wireless sensor network [4].

In this paper, we proposed an efficient energy aware routing protocol, which is based upon the on-demand ad hoc routing protocol AODV [5, 6], which determines a proper path with consideration of node residual battery powers. The proposed protocol aims to extend the lifetime of the overall sensor network by avoiding the unbalanced

exhaustion of node battery powers as traffic congestion occurs on specific nodes participating in data transfer.

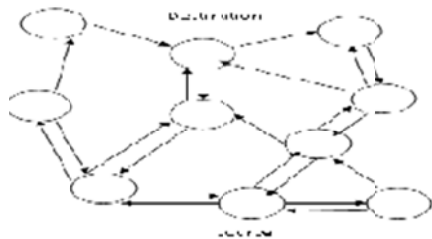
In section 2 of this paper, we describe the well-known AODV routing protocol and show some difficulties in adapting the protocol for wireless sensor network. In section 3, we propose an efficient routing protocol, which considers the node residual battery power while extending the lifetime of the network. Section 4 discusses the NS-2 simulation performance analysis of the routing protocols along with final conclusions and future studies.

## **2. Related Study and Problems defined**

The AODV (Ad hoc On-demand Distance Vector) protocol is an on-demand routing protocol, which accomplishes the route discovery whenever a data transfer is requested between nodes. The AODV routing protocol searches a new route only by request of source nodes. When a node requests a route to a destination node, it initiates a route discovery process among network nodes. The protocol can greatly reduce the number of broadcasts requested for routing search processes, when compared to the DSDV (Destination Sequenced Distance Vectors) routing protocol, which is known to discover the optimum route between source and destination with path information of all nodes. Additionally, since each node in the DSDV routing protocol maintains a routing table - data, which includes complete route information - the AODV protocol greatly improves some drawbacks of DSR (Dynamic Source

Routing) protocol such as the overhead incurred at data transfer.

Once a route is discovered in the AODV routing protocol, the route will be maintained in a table until the route is no longer used. Each node in the AODV protocol contains a sequence number, which increases by one when the location of a neighbor node changes. The number can be used to determine the recent route at the routing discovery. Figure-1



The AODV protocol utilizes a similar routing discovery process as the DSV protocol but uses a different process to maintain and manage a routing table. The nodes of the DSV protocol

Figure 1. Flooding of RREQ messages maintains all routing information between source and destination but the nodes of the AODV protocol have path information in a brief routing table, which stores the destination address, destination sequence number, and next hop address. Each entry of a routing table has a lifetime field, which is set when its routing information is updated and changed. An entry will be removed from the routing table when its lifetime is expired. Moreover, to maintain a routing table, the AODV protocol periodically exchanges routing messages between neighbor nodes. Such processes typically raise significant overhead and wastes available bandwidth. However, the AODV protocol reduces the latency time

of the routing discovery and determines efficient routes between nodes. Figure 2. A routing establishing flow between source and destination.



(fig-2)

The route discovery process of the AODV protocol is similar to that of DSR. A source node broadcasts a RREQ (Route REQuest) packet to find a route to a destination node. When a neighbor node receives the RREQ packet, it rebroadcasts the packet to intermediate nodes until the packet arrives at a destination node. At the same time, the intermediate node or the destination node, which receives a RREQ packet, replies a RREP (Route reply) packet back to the source node. The destination node collects all RREQ messages during a time interval, determines a least hop-count route, and then sends a RREP message to the source node.

The sequence number of a RREQ packet can eliminate a loop generation and make an intermediate nodes reply only on recent route information. When an intermediate node forwards a RREQ packet to neighbor nodes, the receiver node records the intermediate node into the routing information in order to determine the forwarding path. Such processes repeat until arriving at the destination. Then the destination node sends a RREP message, which includes the routing, to the source via the reverse path. In the case that a node receives duplicated RREQ messages, it uses only the first message and ignores the rest. If

errors occur on a specific link of the routing path, either a local route recovery process is initiated on a related node or a RERR (Route Error) message will be issued to the source for a source route recovery process. In such cases, the intermediate nodes receiving the RERR message eliminate all routing information related to the error link.

The AODV routing protocol determines a least hop-count path between a source and a destination, thus minimizing the end-to-end delay of data transfer. Since the protocol uses the shortest route for end-to-end data delivery, it minimizes the total energy consumption. However, if two nodes perform data transfer for long time on the specific path, nodes belonging in this path use more battery power than other nodes, resulting in earlier powering out of nodes. The increase of power-exhausted nodes creates partitions in the wireless sensor network. The nodes belonging to these partitions cannot transfer any further data, thus killing the lifetime of the network. In order to extend the lifetime of the network, one possible solution is to make equally balanced power consumption of sensor nodes. Since AODV routing mechanism does not consider the residual energy of nodes at the routing setup, and since it considers only routing hop count as a distance metric, such unbalanced node energy consumptions occurs. An efficient routing algorithm is proposed, which considers both node hop-count and node energy consumption in section 3

### 3. Problem Formulation

#### 3.1. Proposed Routing Protocol

In this paper, we describe a routing protocol, which considers a residual

energy of sensor nodes to avoid unbalanced energy consumption of sensor nodes. The proposed protocol is based upon a reactive ad hoc AODV routing algorithm. The protocol can make the node energy consumption balanced and extend overall network lifetime without performance degradation such as delay time, compared to the AODV routing algorithm.

#### 3.2. Operations of the proposed routing protocol

The proposed protocol performs a route discovery process similar to the AODV protocol. The difference is to determine an optimum route by considering the network lifetime and performance; that is, considering residual energy of nodes on the path and hop count. In order to implement such functions, a new field, called Min-RE (Minimum Residual Energy) field, is added to the RREQ message as shown in Figure 3. The Min-RE field is set as a default value of -1 when a source node broadcasts a new RREQ message for a route discovery process.

Type	J	R	G	D	U	Reserved	Hop Count
RREQ ID.							
Destination IP Address							
Destination Sequence Number							
Originator IP Address							
Originator Sequence Number							
Min-RE(Added)							

Figure 3.1. A RREQ message format for our proposed protocol

To find a route to a destination node, a source node floods a RREQ packet to the network. When neighbor nodes



receive the RREQ packet, they update the Min-RE value and rebroadcast the packet to the next nodes until the packet arrives at a destination node. If the intermediate node receives a RREQ message, it increases the hop count by one and replaces the value of the Min-RE field with the minimum energy value of the route. In other words, Min-RE is the energy value of the node if Min-RE is greater than its own energy value; otherwise Min-RE is unchanged.

Although intermediate nodes have route information to the destination node, they keep forwarding the RREQ message to the destination because it has no information about residual energy of the other nodes on the route. If the destination node finally receives the first RREQ message, it triggers the data collection timer and receives all RREQ messages forwarded through other routes until time expires. After the destination node completes route information collection, it determines an optimum route with use of a formula shown in 3.2 and then sends a RREP message to the source node by unicasting. If the source node receives the RREP message, a route is established and data transfer gets started. Such route processes are performed periodically, though node topology does not change to maintain node energy consumption balanced. That is, the periodic route discovery will exclude the nodes having low residual energy from the routing path and greatly reduce network partition

### 3.3. Determination of routing

The optimum route is determined by using the value of  $\alpha$  described in formula (1). The destination node calculates the values of  $\alpha$  for received all route information and choose a route that has

the largest value of  $\alpha$ . That is, the proposed protocol collects routes that have the minimum residual energy of nodes relatively large and have the least hop-count, and then determines a proper route among them, which consumes the minimum network energy compared to any other routes.

$$\alpha = \frac{Min-RE}{k \cdot No-Hops} \dots \dots \dots (1)$$

Here Min-RE is the minimum residual energy on the route and No-Hops is the hop count of the route between source and destination. And k is the weight coefficients for the hop count. The energy consumption of one hop in the network will be little, where one hop means a data transfer from a node to the next node. The weight coefficient k is used to adjust the difference of Min-RE and No-Hops in simulation.

### 3.4. The analysis of routing protocols

To understand the operations of the proposed protocol, we consider three different routing protocols for operational comparison:

- **Case 1:** Choose a route with the minimum hop count between source and destination. (AODV routing protocol).
- **Case 2:** Choose a route with largest minimum residual energy. (Max\_Min Energy (Min-ER) routing protocol)
- **Case 3:** Choose a route with the large minimum residual energy and less hop count. i.e. with the longest network lifetime (our proposed routing protocol).

Consider a network illustrated in Figure 4. Here we consider a simple routing example to setup route from source node S to destination node D. The number

written on a node represents the value of residual node energy. We consider three different cases of routes. Since the Case 1 considers only the minimum hop count, it selects route  $\langle S-B-J-D \rangle$  which has the hop count of 3. In the Case-2, select route  $\langle S-A-K-F-L-H-G-D \rangle$  which has Min-RE 6 is chosen because the route has the largest minimum residual energy among routes. Our proposed model needs to compute the value of  $\alpha$  by using formula (1), and selects a route with largest value of  $\alpha$ . Thus Case 3 selects route  $\langle S-C-E-I-D \rangle$ , which has largest  $\alpha$  value of 1.25.

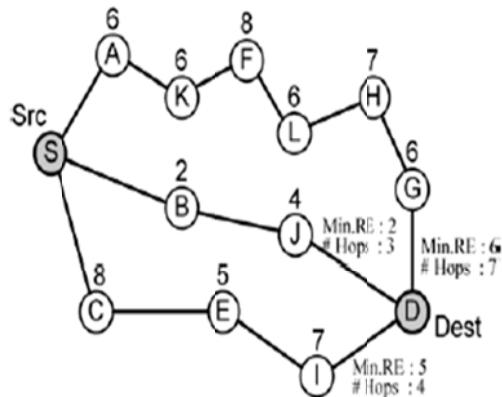


Figure 4. A sample network for establishment of routing paths

Case 1 selects the shortest path without considering residual energy of nodes, which is the same as the AODV routing algorithm. This case does not sustain a long lifetime in the network as described in section 2. Case 2 selects a route with largest minimum residual energy to extend network lifetime but it has serious problem in terms of the hop count. Case-3 improves the drawbacks of Case 1 and Case-2 by considering both residual energy and hop count. It extends network lifetime by arranging almost all nodes to involve in data transfer. The proposed protocol also

selects a route with the longest lifetime in the network without performance degradation such as delay time and node energy consumption.

#### 4. Performance Evaluation

The performance analysis of routing protocols is evaluated with the NS-2 simulator [7]. Then our proposed protocol is compared to other two routing protocol (Case 1 and Case 2) in terms of the average end-to-end delay and the network lifetime.

##### 4.1. Simulation Environment

In this simulation, our experiment model performed on 100 nodes, which were randomly deployed and distributed in a  $500 \times 500$  square meter area. We assume that all nodes have no mobility since the nodes are fixed in applications of most wireless sensor networks. Simulations are performed for 60 seconds. We set the propagation model of wireless sensor network as two-ray ground reflection model and set the maximum transmission range of nodes as 100 meters. The MAC protocol is set to IEEE 802.11 and the bandwidth of channel is set to 1Mbps.

Each sensor node in the experimental network is assumed to have an initial energy level of 7 Joules. A node consumes the energy power of 600mW on packet transmission and consumes the energy power of 300mW on packet reception. The used traffic model is an UDP/CBR traffic model. Size of data packet is set to 512byte and traffic rate varies to 2, 3, 4, 5, 6, 7, 8, 9, 10 packets/sec to compare performance depend on traffic load. In this simulation, the weight coefficient  $k$  is calculated based on traffic model, bandwidth, and energy consumption of a node. Our simulation model uses a

sensor network that has the bandwidth of 1 Mbps, the packet size of 512 bytes. Thus, packet transmission time per link is calculated, as about 0.004096seconds and the node energy consumption for our simulation model is about 0.0037 Joule.

#### 4.2. Simulation Results

The major performance metrics of a wireless sensor network are the end-to-end delays (or throughput) and network lifetime. In order to compare network lifetime of three different routing protocols, we measured the number of exhausted energy nodes every second for 60 seconds. Figure-5 illustrates that number of exhausted node of each model according to simulation time. The vertical axis is represented the number of exhausted energy nodes in the network. The increase of the exhausted energy nodes may cause a network partition that makes network functions impossible. The number of exhausted energy nodes in AODV (Case 1), Min-ER (Case 2), and our protocol start appearing at 35, 42, and 47 seconds, respectively. The number in these protocols is saturated on 80% of nodes at 45, 48, and 55 seconds, respectively. As shown in Figure 5, our proposed protocol has longer lifetime duration than other protocols. In Particular, 60% of nodes in our protocol work normally at the elapsed time of 55 seconds compared to 20 % in other protocols. This result shows that our routing protocol properly leads to balanced energy consumption of sensor nodes.

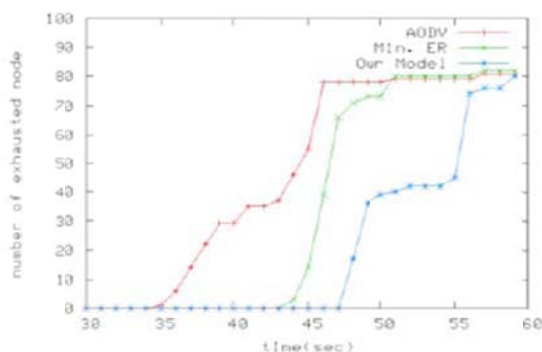


Figure 5. Comparison of the number of exhausted energy nodes

Figure 6 gives the average end-to-end delay of all three protocols in respect with traffic loads. The AODV protocol has minimum delay and Min-ER has maximum delay. Additionally, the delay of our protocol was little higher than that of AODV. Our protocol has a relatively good delay characteristic without degradation of performance compared to AODV.

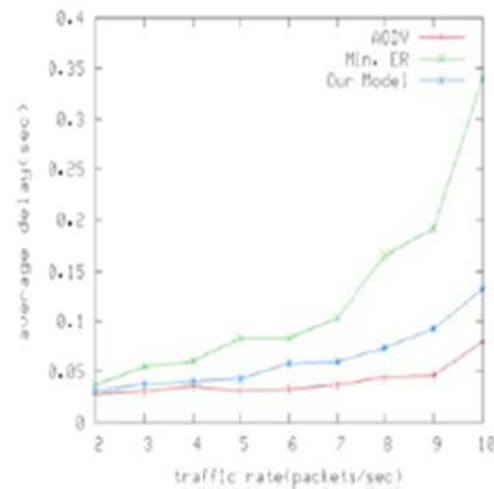


Figure 6. End-to end delay for traffic rate

Based upon the simulation results, we confirmed that our proposed protocol can control the residual node energy and the hop count in a wireless sensor network and effectively extend the network lifetime without performance degradation.

#### 5. Conclusions

In this work, we proposed power aware routing protocol, which improves the lifetime of sensor networks. The protocol considers both hop count and

the residual energy of nodes in the network. Based upon the NS-2 simulation, the protocol has been verified with very good performance in network lifetime and end-to-end delay. If we used a simulation mode of the large number of nodes (or 1000 or more), our protocol make network lifetime much longer compared to AODV and Min-ER protocols. Consequently, our proposed protocol can effectively extend the network lifetime without other performance degradation.

The applications in wireless sensor networks may require different performance metrics. Some applications are focused on the lifetime of network and the others on delay. Some efficient routing mechanisms in respect with applications may be needed for further studies.

## References

[1] Ian F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," IEEE Communications Magazine, volume 40, Issue 8, pp.102-114, Aug. 2002.

[2] K. Akkaya and M. Younis, "A Survey of Routing Protocols in Wireless Sensor Networks, " in the Elsevier Ad Hoc Network Journal, Vol 3/3, pp.325-349, 2005.

[3] Q. Jiang and D. Manivannan, "Routing protocols for sensor networks," Proceedings of CCNC 2004, pp.93-98, Jan. 2004.

[4] Suresh Singh and Mike Woo, "Power-aware routing in mobile ad hoc networks", Proceedings of the 4th annual ACM/IEEE international conference on

Mobile computing and networking, Dallas, Texas, pp. 181 -190, 1998.

[5] Charles E. Perkins and Elizabeth M. Royer. "Ad hoc On-demand Distance Vector Routing." Proceedings of the 2nd IEEE Workshop on Mobile Computing Systems and Applications, New Orleans, LA, pp. 90-100, February 1999.

[6] Charles E. Perkins, "Ad hoc On-demand Distance Vector (AODV) Routing." RFC 3561,IETF MANET Working Group, July 2003.

[7] Information Sciences Institute, "The Network Simulator ns-2" <http://www.isi.edu/nanam/ns/>, University of Southern California.

[8] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam and E. Cayirci, "Wireless Sensor Networks: A Survey," Computer Networks, Vol. 38, No. 4, March 2002, pp. 393-422. doi:10.1016/S1389-1286(01)00302-4

[9] X. Zhao, K. Mao, S. Cai and Q. Chen, "Data-Centric Routing Mechanism Using Hash-Value in Wireless Sensor Network," Wireless Sensor Network, Vol. 2, No. 9, 2010, pp. 703-709. doi:10.4236/wsn.2010.29086

## BIOGRAPHY

### First Author :

Rajesh Kumar Sahoo is presently working as Assistant Professor in Ajay Binay Institute Of Technology,Cuttack,Orissa,India. He has acquired his M.Tech degree fromKIIT University, KIIT,Bhubaneswar, Orissa, India. He is a research student of Berhampur University,Berhampur. He has contributed more than two papers to

Journals and Proceedings. He has written two books on “Computer Architecture and Organization” and “Computer Architecture and organization-II”. His areas of interests are in Software Engineering, Object Oriented Systems, Sensor Network, Computer Architecture and Compiler Design etc.

University. He is a life member of CSI, ISTE & OITS, and a Fellow of ACEEE. His special field of interests are Intelligent Agents, Service Oriented System Modeling, Data mining, Network Intrusion Detection.

### **Second Author :**

Satyabrata Das received the degree in Computer Sc & engineering from Utkal University, in 1996. He received the M.Tech. degree in CSE from ITER, Bhubaneswar. He is a research student of Fakir Mohan University, Balasore in the dept. of I&CT Currently, he is an Asst. Professor at College of Engineering Bhubaneswar, Orissa. His interests are in AI, Soft Computing, Data Mining, DSP, Neural Network.

### **Third Author:**

Dr. Durga Prasad Mohapatra studied his M.Tech at National Institute of Technology, Rourkela, India. He has received his Ph. D from Indian Institute of Technology, Kharagpur, India. Currently, he is working as Associate Professor at National Institute of Technology, Rourkela. His special fields of interest include Software Engineering, Discrete Mathematical Structure, slicing Object-Oriented Programming. Real-time Systems and distributed computing.

### **Fourth Author:**

Dr. Manas Ranjan Patra holds a Ph.D. degree in computer Science from the Central University of Hyderabad. Currently, he heads the Department of Computer Science, Berhampur

# **IJCSI CALL FOR PAPERS SEPTEMBER 2011 ISSUE**

**Volume 8, Issue 5**

The topics suggested by this issue can be discussed in term of concepts, surveys, state of the art, research, standards, implementations, running experiments, applications, and industrial case studies. Authors are invited to submit complete unpublished papers, which are not under review in any other conference or journal in the following, but not limited to, topic areas. See authors guide for manuscript preparation and submission guidelines.

**Accepted papers will be published online and indexed by Google Scholar, Cornell's University Library, DBLP, ScientificCommons, CiteSeerX, Bielefeld Academic Search Engine (BASE), SCIRUS, EBSCO, ProQuest and more.**

**Deadline: 31<sup>st</sup> July 2011**

**Notification: 31<sup>st</sup> August 2011**

**Revision: 10<sup>th</sup> September 2011**

**Online Publication: 30<sup>th</sup> September 2011**

- Evolutionary computation
- Industrial systems
- Evolutionary computation
- Autonomic and autonomous systems
- Bio-technologies
- Knowledge data systems
- Mobile and distance education
- Intelligent techniques, logics, and systems
- Knowledge processing
- Information technologies
- Internet and web technologies
- Digital information processing
- Cognitive science and knowledge agent-based systems
- Mobility and multimedia systems
- Systems performance
- Networking and telecommunications
- Software development and deployment
- Knowledge virtualization
- Systems and networks on the chip
- Context-aware systems
- Networking technologies
- Security in network, systems, and applications
- Knowledge for global defense
- Information Systems [IS]
- IPv6 Today - Technology and deployment
- Modeling
- Optimization
- Complexity
- Natural Language Processing
- Speech Synthesis
- Data Mining

**For more topics, please see <http://www.ijcsi.org/call-for-papers.php>**

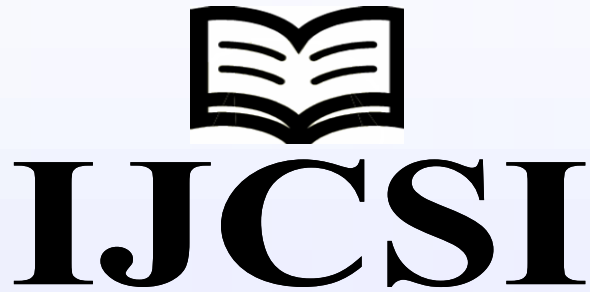
All submitted papers will be judged based on their quality by the technical committee and reviewers. Papers that describe on-going research and experimentation are encouraged. All paper submissions will be handled electronically and detailed instructions on submission procedure are available on IJCSI website ([www.IJCSI.org](http://www.IJCSI.org)).

For more information, please visit the journal website ([www.IJCSI.org](http://www.IJCSI.org))



**© IJCSI PUBLICATION 2011**

**[www.IJCSI.org](http://www.IJCSI.org)**



# IJCSI

The International Journal of Computer Science Issues (IJCSI) is a well-established and notable venue for publishing high quality research papers as recognized by various universities and international professional bodies. IJCSI is a refereed open access international journal for publishing scientific papers in all areas of computer science research. The purpose of establishing IJCSI is to provide assistance in the development of science, fast operative publication and storage of materials and results of scientific researches and representation of the scientific conception of the society.

It also provides a venue for researchers, students and professionals to submit ongoing research and developments in these areas. Authors are encouraged to contribute to the journal by submitting articles that illustrate new research results, projects, surveying works and industrial experiences that describe significant advances in field of computer science.

## **Indexing of IJCSI**

1. Google Scholar
2. Bielefeld Academic Search Engine (BASE)
3. CiteSeerX
4. SCIRUS
5. Docstoc
6. Scribd
7. Cornell's University Library
8. SciRate
9. ScientificCommons
10. DBLP
11. EBSCO
12. ProQuest