IJCSI

International Journal of Computer Science Issues

Volume 8, Issue 5, No 2, September 2011
ISSN (Online): 1694-0814

www.IJCSI.org

IJCSI proceedings are currently indexed by:

IJCSI Publicity Board 2011

IJCSI Editorial Board 2011

EDITORIAL

In this fifth 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 its proceedings free of charge for all researchers.

We are pleased to present IJCSI Volume 8, Issue 5, No 2, September 2011 (IJCSI Vol. 8, Issue 5, No 2). The acceptance rate for this issue is 31.7%.

IJCSI Editorial Board
September 2011 Issue
ISSN (Online): 1694-0814

www.IJCSI.org

IJCSI Reviewers Committee 2011

• 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. Supheakmungkol SARIN, Waseda University, Japan
• Mr. Shoukat Ullah, Govt. Post Graduate College Bannu, Pakistan
• Dr. Vivian Augustine, Telecom Zimbabwe, Zimbabwe
• Mrs. Mutalli Vatila, 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

- 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
- Dr. D.I. George Amalarethinam, Jamal Mohamed College, Bharathidasan University, India

• Mr. Khaleel Ahmad, S.V.S. University, India
• Mr. Amin Zehtabian, Babol Noshirvani University of Technology / Tetta Electronic Company, Iran
• Mr. Rahul Katarya, Department of Information Technology , Delhi Technological University, India
• Dr. Vincent Ele Asor, University of Port Harcourt, Nigeria
• Ms. Prayas Kad, Capgemini Australia Ltd, Australia
• Mr. Alireza Jolfaei, Faculty and Research Center of Communication and Information Technology, IHU, Iran
• Mr. Nitish Gupta, GGSIPU, India
• Dr. Mohd Lazim Abdullah, University of Malaysia Terengganu, Malaysia
• Mr. Rupesh Nasre., Indian Institute of Science, Bangalore., India.
• Mrs. Dimpi Srivastava, Dept of Computer science, Information Technology and Computer Application, MIET, Meerut, India
• Prof. Santosh Balkrishna Patil, S.S.G.M. College of Engineering, Shegaon, India
• Mr. Mohd Dilshad Ansari, Jaypee University of Information Technology Solan (HP), India
• Mr. Ashwani Kumar, Jaypee University of Information Technology Solan(HP), India
• Dr. Abbas Karimi, Faculty of Engineering, I.A.U. Arak Branch, Iran
• Mr. Fahimuddin.Shaik, AITS, Rajampet, India
• Mr. Vahid Majid Nezhad, Islamic Azad University, Iran
• Ms. C. Divya, Dr G R Damodaran College of Science, Coimbatore-641014, Tamilnadu, India
• Prof. D. P. Sharma, AMU, Ethiopia
• Dr. Sukumar Senthilkumar, School of Mathematical Sciences, Universiti Sains Malaysia, Malaysia
• Mr. Sanjay Bhargava, Banasthali University, Jaipur, Rajasthan, India
• Prof. Rajesh Deshmukh, Shri Shankaracharya Institute of Professional Management & Technology, India
• Mr. Shervan Fekri Ershad, shiraz international university, Iran
• Dr. Vladimir Urosevic, Ministry of Interior, Republic of Serbia
• Mr. Ajit Singh, MDU Rohtak, India

TABLE OF CONTENTS

Graph Colouring Algorithm for Validating Labelled 2D Line Drawing Objects

M.Z. Matondang[1], A.A.Samah[2] H.Haron[3] and H.A.Majid[4]

[1] **Faculty of Computer and Information Systems, Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor Bahru, Malaysia**

[2] **Faculty of Computer and Information Systems, Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor Bahru, Malaysia**

[3] **Faculty of Computer and Information Systems, Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor Bahru, Malaysia**

[4] **Faculty of Computer and Information Systems, Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor Bahru, Malaysia**

Abstract: Line labelling has been used to determine whether a two-dimensional (2D) line drawing object is a possible or impossible representation of a three-dimensional (3D) solid object. However, the results are not sufficiently robust because the existing line labelling methods do not have any validation method to verify their own result. In this research paper, the concept of graph colouring is applied to a validation technique for a labelled 2D line drawing. As a result, a graph colouring algorithm for validating labelled 2D line drawings is presented. A high-level programming language, MATLAB R2009a, and two primitive 2D line drawing classes, prism and pyramid are used to show how the algorithms can be implemented. The proposed algorithm also shows that the minimum number of colours needed to colour the labelled 2D line drawing object is equal to 3 for prisms and $n-1$ for pyramids, where n is the number of vertices (junctions) in the pyramid objects.

Keywords: *Graph colouring, line labelling, line drawing, validation*.

1. Introduction

Discussions on three-dimensional (3D) solid object reconstruction from two-dimensional (2D) single views reveal that there are several steps that need to be accomplished to transform a 2D drawing into a 3D view [1-3]. As the input of a reconstruction process, a 2D drawing could be a regular or an irregular line drawing. However, when the input is an irregular form or a sketch, then the input should be converted into a regular 2D line drawing first, before proceeding to the interpretation of the drawing. However, it is assumed that the interpretation of a 2D single view into a 3D view is always an easy task. Generally, people assume that the interpretation of a drawing or an image as an object does not need conscious thought. Moreover, human vision itself appears effortless. For this reason, the validation of the drawing is rarely considered with regard to an object's transformation, especially when moving from a 2D view into a 3D view. Some mistakes arise because the drawing could be an impossible drawing or a possible drawing that cannot be represented as a common object, such as a matchbox. In other words, the validity of the 2D drawing (whether it represents an object or not) is important to know. A validation process filters a 2D drawing for a 3D reconstruction process, accepting the possible drawings and rejecting the impossible drawings. In this research paper, this validation process will be the main focus of the discussion.

Several previous studies show that line labelling has been used when representing 2D line drawings and in reconstructing 3D objects [4-8]. These studies have combined line labelling with geometric models to reconstruct 3D objects from 2D line drawings. Line labelling is also useful for identifying impossible objects and for validating 2D line drawings. The many line labelling scenarios used before provide motivation to use line labelling in the validation of 2D line drawings in the present research. However, in this research paper, the concept of graph colouring is applied to colour labelled 2D line drawings. The objective of the colouring process is the validation of the label of the 2D line drawing. We need to increase the speed of the validation process for 2D line drawings, to determine whether a 2D line drawing contains a possible or an impossible 3D object. In addition, the colour labelling process is needed because the line labelling algorithm does not have any validation technique for its own result [9,10]. Readers interested in a detailed discussion about the line labelling algorithm are referred to [6-8,11]. The next section presents a discussion about the concept of graph colouring and some related issues.

2. Graph Colouring

In graph theory, a graph has a different meaning compared to statistical topics that utilise bar, circle and line graphs as representations of mathematical equations. Here, a graph is a model that consists of vertices and edges. Primarily, graphs illustrate real life problems and situations, such as social acquaintances, sports schedules, transportation routes, and computer networks [12].

By definition, a graph $G=(V,E)$ is a mathematical model consisting of a finite set of vertices $V=\{v_1,v_2,\ldots,v_n\}$, which are represented by points, and a finite set of edges $E=\{e_1,e_2,\ldots,e_m\}$, which are represented by line segments, where *n* and *m* are integers. Here, two vertices connected by an edge are said to be adjacent, and an edge is said to be incident to the vertices it connects. The degree (*deg*) of a vertex is the number of adjacent vertices. In this research paper, we discuss a special case in graph theory that is called graph colouring. Graph colouring is a special case of graph labelling, where colour is used as the label, so that there are no two adjacent vertices, edges, or faces that are assigned the same colour. If we are given a graph $G(V,E)$, then the chromatic number $\chi(G)=k$ is defined as the minimum numbers of colours needed to colour $G(V,E)$. Assume that $G(V,E)$ is a graph and for which $G^*=(V,C)$ is a mapping function $f:v\rightarrow c$ with $c\in C$ a finite set of colours, such that if $(v_1v_2)\in E$, then $f(v_1)\neq f(v_2)$. This statement implies that adjacent vertices are not assigned the same colour [12].

There are three types of graph colouring, known as vertex, edge, and face colouring. In this research paper, the discussion is focused on edge colouring. The edge colouring of a graph, $G(V,E)$, is a colour assignment for each edge in $G(V,E)$, such that for each two adjacent edges e_j and e_k to a vertex v_i, those edges do not share the same colour. For a detailed discussion of graph colouring properties, readers can refer to [13] as an authoritative reference on graph colouring.

Some related literature and previous studies show that graph colouring is used to solve problems that may involve conflicts or items that need to be separated [12]. Several applications previously performed include separating chemicals during lab work, separating animals in a zoo, scheduling classes or exams, and (the most common application) colouring maps to separate distinct countries. Iturriaga-Velazquez [14] shows that the original problem involving the four-colour problem is the question of whether four colours are sufficient to colour the countries on a world map, never assigning the same colour to two countries with a common boundary. However, over time, graph colouring has been applied to many various fields of research. Marx [15] explained the applications of graph colouring to scheduling problems in his paper, while Gaceb et al. [16] carried out physical layout segmentation for postal sorting systems using a graph colouring application. Redl [17] used the graph colouring approach for university timetabling at the University of Houston, and Dobrolowski et al. [18] developed the Koala Graph Coloring Library, which is an open graph colouring library for real world applications. However, this research goes beyond previous studies and applications because we attempt to use the concept of graph colouring to develop an algorithm for the validation of labelled 2D line drawings. The proposed algorithm expedites the validation of 2D line drawings (to classify them as possible versus impossible objects), which is accomplished by a line labelling algorithm. This method is expected to provide better validation compared to the previous work of Matondang et al. [10], which is the only line labelling algorithm used to perform such a validation. The line labelling algorithm itself cannot be used as the validation technique to validate its own labelled 2D line drawing; in other words, the line labelling algorithm is not sufficiently robust for the labelling process to determine whether the line drawing is valid or invalid, representing a solid versus an impossible object, respectively. Therefore, the combination of both the line labelling algorithm and the graph colouring application can be useful to speed up and enhance the validation of a 2D line drawing. The graph colouring algorithm has been successful in many fields, which provides motivation to adapt it to the validation of 2D line drawings. The next section presents the proposed algorithm.

3. The Proposed Algorithm

This section presents the proposed algorithm for validating labelled 2D line drawings based on the the concept of graph colouring. Our expected output is a valid, labelled 2D line drawing object with a different colour on each edge. Table 1 shows the algorithm.

Table 1: The proposed algorithm

Edge (line) colouring for labelled 2D line drawing object

- Step 1: input the number of vertices (junctions) in the labelled 2D line drawing given (object given).
  %%start the edge labelling.
- Step 2: input every horizontal edge.
- Step 3: input every vertical edge.
  *%%evaluate the greatest degree *deg in the object given.*
- Step 4: determine the greatest degree **deg* for each vertex (junction) in the object given.
  %%assume h, n as an integer number

```
for h=1:numberofvertex;
   n = 1;
   for i=1:numberofedge;
      if labeledge(i,1) == V(h,1);
         Degree(h,1) = n;
         n = n + 1;
      elseif labeledge(i,2) == V(h,1);
         Degree(h,1) = n;
         n = n + 1;
      end
   end
end
```

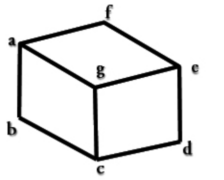
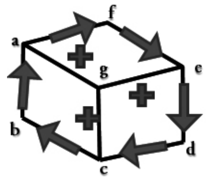
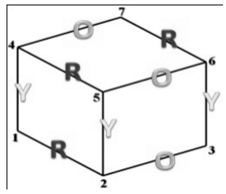
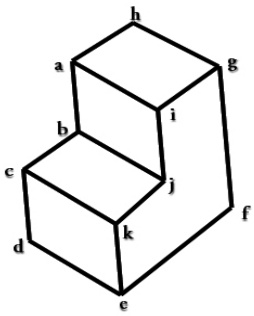
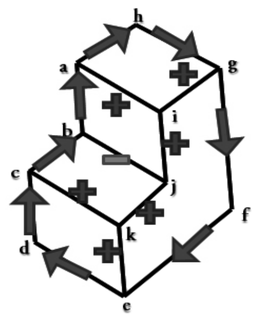
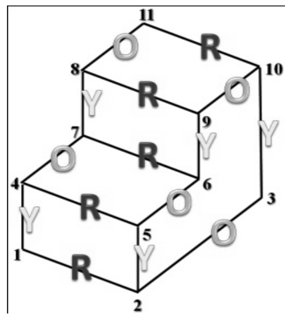
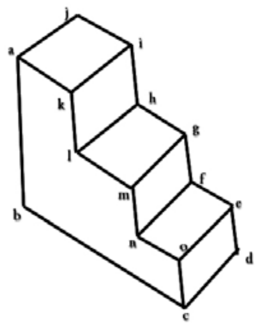
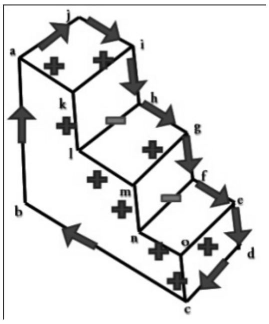
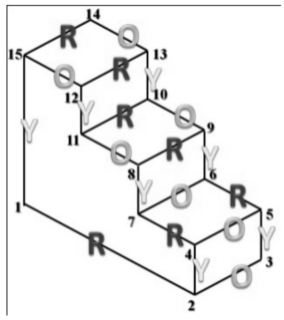
- Step 5: colouring process
  Assume *m* is an integer number

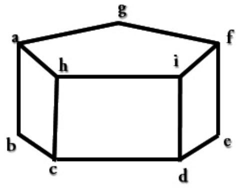
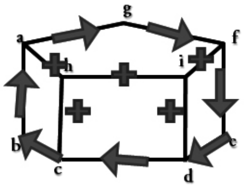
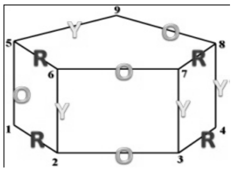
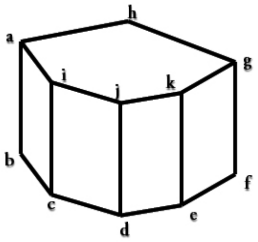
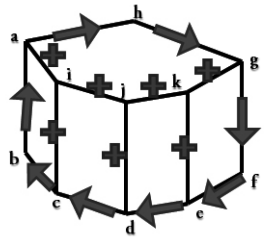
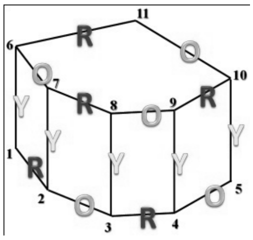
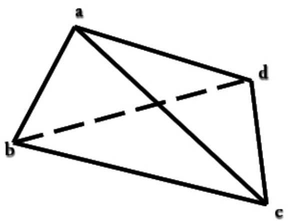
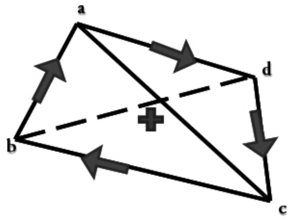
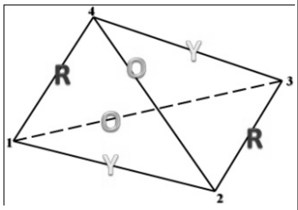
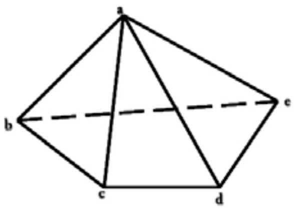
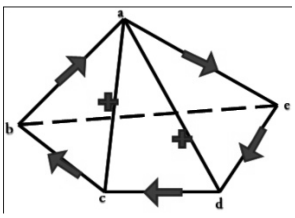
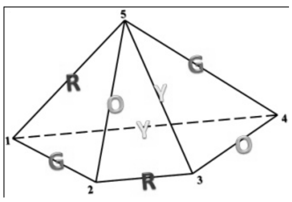
```
L = zeros(numberofedge,1);
c = 0; %%assume c as the starting point for the iteration
w = 1; %%assume w as the starting point for the vertex addressing in S
for m=1:maxdegree
   c = c + 1;
   S = zeros(numberofvertex,1);
         for j=1:numberofedge
                  if L(j,1)==0
                  if  labeledge(j,1)~=S  &
labeledge(j,2)~=S
                  L(j,1)=c;
                           S(w,1)=labeledge(j,1);
                           w = w + 1;
                           S(w,1)=labeledge(j,2);
                           w = w + 1;
                           elseif  labeledge(j,1)==S
          ;
                  L(j,1)=0;
                           else labeledge(j,2)==S ;
                  L(j,1)=0;
                           end
                           end
                  end
      end
```

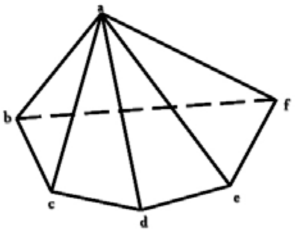
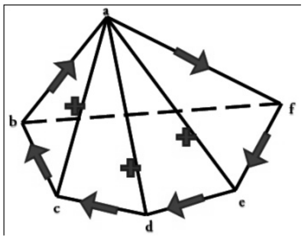
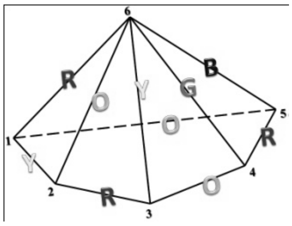
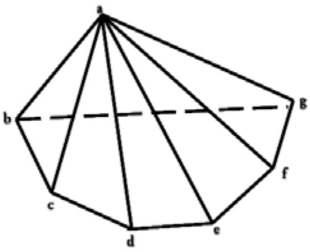
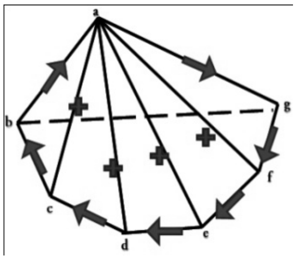
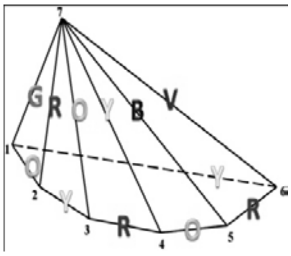
4. Experimental Results

In this section, we present results that show how the proposed algorithm has been implemented in two classes of 2D line drawing objects, namely the prism and pyramid. The results show the validation of the 2D line drawing using both the line labelling and the proposed graph colouring algorithm. However, assumptions have been made to simplify the implementation. The assumptions are as follows: First, the tested 2D line drawings are assumed to be an engineering sketch in the form of a 2D line drawing that represents a solid model. Second, the 2D line drawing is assumed to represent a valid solid model when all unwanted points or lines have been removed, and there are no unconnected points or lines. Third, the solid model is assumed to be a 2D line drawing with all of the informative lines shown. Fourth, there is only one hidden point in the backside of the solid model. These assumptions make the proposed algorithm more logical, or otherwise the engineering sketch is not seen as a solid model because the projection is parallel to the other faces of the object. In this case, it is impossible to interpret, reconstruct and represent the sketch as a solid model, and hence, the analysis of the accuracy of the results will become simpler.

Table 2: The experimental result

2D line drawing	**Labelled 2D line drawing using line labelling algorithm**	**Valid-labelled of 2D line drawing with different colours at each edge**	**Role Model**
 Cube		 • Red (R), Orange (O), Yellow (Y)	• Label and colour (1,2) → red (2,3) → orange (4,5) → red (5,6) → orange (6,7) → red (7,4) → orange (1,4) → yellow (2,5) → yellow (3,6) → yellow • Number of visible vertices = 7 • Number of visible edges = 8 • Number of colours needed $k = 3$
 L-block		 • Red (R), Orange (O), Yellow (Y)	• Label and colour (1,2) → red (2,3) → orange (4,5) → red (5,6) → orange (6,7) → red (7,4) → orange (8,9) → red (9,10) → orange (10,11) → red (11,8) → orange (1,4) → yellow (2,5) → yellow (8,7) → yellow (9,6) → yellow (3,10) → yellow • Number of visible vertices = 11 • Number of visible edges = 15 • Number of colours needed $k = 3$
 Stairs		 • Red (R), Orange (O), Yellow (Y)	• Label and colour (1,2) → red (2,3) → orange (7,4) → red (6,5) → red (4,5) → orange (6,7) → orange (8,9) → red (9,10) → orange (10,11) → red (11,8) → orange (12,13) → red (13,14) → orange (14,15) → red (15,12) → orange (15,1) → yellow (11,22) → yellow (8,7) → yellow (10,13) → yellow (9,6) → yellow (2,4) → yellow (3,5) → yellow • Number of visible vertices = 15 • Number of visible edges = 21 • Number of colours needed $k = 3$

 pentahedral-prism		 • Red (R), Orange (O), Yellow (Y)	• Label and colour (1,2) → red (2,3) → orange (3,4) → red (5,6) → red (6,7) → orange (7,8) → red (8,9) → orange (9,5) → yellow (6,2) → yellow (1,5) → orange (3,7) → yellow (4,8) → yellow • Number of visible vertices = 9 • Number of visible edges = 12 • Number of colours needed $k = 3$
 hexahedral-prism		 • Red (R), Orange (O), Yellow (Y)	• Label and colour (1,2) → red (2,3) → orange (3,4) → red (4,5) → orange (11,6) → red (7,8) → red (6,7) → orange (9,8) → orange (9,10) → red (10,11) → orange (1,6) → yellow (3,8) → yellow (4,9) → yellow (5,10) → yellow (2,7) → yellow • Number of visible vertices = 11 • Number of visible edges = 15 • Numbers of colour needed $k = 3$
 trihedral-pyramid		 • Red (R), Orange (O), Yellow (Y)	• Label and colour (4,1) → red (4,2) → orange (4,3) → yellow (1,2) → yellow (2,3) → red (3,1) → orange • Number of visible vertices = 4 • Number of visible edges = 5 • Number of colours needed $k = 3$
 kwartahedral-pyramid		 • Red (R), Orange (O), Yellow (Y) Green (G)	• Label and colour (5,1) → red (5,2) → orange (5,3) → yellow (5,4) → green (2,3) → red (3,4) → orange (4,1) → yellow (1,2) → green • Number of visible vertices = 5 • Number of visible edges = 7 • Number of colours needed $k = 4$

 pentahedral-pyramid		 • Red (R), Orange (O), Yellow (Y) Green (G), Blue (B)	• Label and colour (6,1) → red (6,2) → orange (6,3) → yellow (6,4) → green (6,5) → blue (2,3) → red (3,4) → orange (4,5) → red (5,1) → orange (1,2) → yellow • Number of visible vertices = 6 • Number of visible edges = 9 • Number of colours needed $k = 5$
 hexahedral-pyramid		 • Red (R), Orange (O), Yellow (Y) Green (G), Blue (B), Violet (V)	• Label and colour (7,2) → red (7,3) → orange (7,4) → yellow (7,1) → green (7,5) → blue (7,6) → violet (1,2) → orange (2,3) → yellow (3,4) → red (4,5) → orange (5,6) → red (6,1) → yellow • Number of visible vertices = 7 • Number of visible edges = 11 • Number of colours needed $k = 6$

Based on the results shown in Table 2 and the assumptions that we made for implementing the proposed algorithm, we found the number of colours needed for a 2D line drawing's classification as a prism or a pyramid. For the prism class with *n* vertices (junctions), there are *n* – 3 vertices with *deg* = 3 and 3 vertices with *deg* = 2. The number of colours needed to colour the prism is equal to the maximum number of the degree **deg* and is valid only for prisms where the number of vertices is equal to $n \geq 6$.
For the pyramid class with *n* vertices (junctions), there are *n* – 1 vertices with *deg* = 3 and only one vertex with *deg* = $n-1$. The number of colours needed to colour the pyramid is equal to $n-1$ and is valid only for pyramids where the number of vertices is equal to $n \geq 4$.

5. Conclusions and Future Work

Line labelling has been used to determine whether a two-dimensional (2D) line drawing represents a possible or an impossible three-dimensional (3D) solid object. However, prior work is not sufficiently robust because the line labelling method does not have any validation method for its own result. In this research paper, a graph colouring method is applied as the validation technique for a labelled 2D line drawing object. As a result, the graph colouring algorithm for validating a labelled 2D line drawing objects is presented. A high-level programming language, MATLAB R2009a, and two primitive 2D line drawing classes, prism and pyramid, are used to show how the proposed algorithm is implemented. Based on the experimental results, it is shown that for the prism class with *n* vertices (junctions), there are *n* – 3 vertices with *deg* = 3 and three vertices with *deg* = 2. The number of colours needed to colour the prism is equal to the maximum number of *deg* and is valid only for prisms having $n \geq 6$ vertices. Meanwhile, for the pyramid class with *n* vertices (junctions), there are *n* – 1 vertices with *deg* = 3 and one vertex with *deg* = $n-1$. The number of colours needed to colour the pyramid is equal to $n-1$ and is valid only for pyramids for which the number of vertices is equal to $n \geq 4$.
A suggestion for future research is to attempt to extend the algorithm for a more complex and general 2D line drawing that is not limited to objects in the prism and pyramid classes.

6. Acknowledgement

The authors would like to thank the Universiti Teknologi Malaysia (UTM) for UTM Short Term grant (Vot-77228), and Research Management Center (RMC) for the support in making this project a success.

References:

[1] M. Z. Matondang, H. Haron and S. Thalib, “Three-dimensional visualization of two-dimensional data: the mathematical modeling,” Proc. of the 2nd IMT-GT Regional Conference on Mathematics, Statistics and Applications, 13 - 15 June 2006, Penang, Malaysia, Computer Sciences and Applications. Ed. by Yahya Abu Hassan, Adli Mustafa and Zarita Zainuddin, Vol. 4.

[2] M. Z. Matondang, Solid Model Reconstruction using Neural Network and the Mathematical Model Representation, Master Thesis, Universiti Teknologi Malaysia, 2009.

[3] S. Mardzuki, Labeling Algorithm For Validation Of 2d Line Drawing, Master Thesis, Universiti Teknologi Malaysia, 2011.

[4] T. Kanade, Recovery of the Three-Dimensional Shape of an Object from a Single View, Artificial Intelligence. 17 (1981) 409-460.

[5] I. J. Grimstead and R. R. Martin, Creating Solid Model From Single 2D Sketches, Proceedings Third Symp. On Solid Modeling Applications, ACM SIGGRAPH. (1995) 233-337.

[6] I. J. Grimstead, Interactive Sketch Input of Boundary Representation Solid Models, PhD Thesis, Univ. of Cardiff, UK, 1997.

[7] P. A. C. Varley and R. R. Martin, Estimating Depth from Line Drawing, In Edt. K.Lee and Patrikalakis, Proc. 7th ACM Symposium on Solid Modeling and Applications, SM’02. ACM Press. (2002) 180-191.

[8] P. A. C. Varley, R. R. Martin and H, Suzuki, Making the Most of using Depth Reasoning to Label Line Drawings of Engineering Objects, *ACM Symposium on Solid Modeling and Application*, (2004) 13-32.

[9] S. Mardzuki, M. Z. Matondang, and H. Haron, Computational Approach in Validating Reconstructed Solid Model Based on Approximate Depth Value, Proceeding the 5th International Conference on Information Technology and Applications (ICITA 2008), CAIRNS-Queensland Australia. (2008) 687 – 694.

[10] M. Z. Matondang, S. Mardzuki and H. Haron, Transformation of engineering sketch to valid solid object, Proc. of Intl. Conf. of The 9th Asia Pacific Industrial Engineering & Management Systems (APIEMS 08) Conference and The 11th Asia Pacific Regional Meeting of International Foundation for Production Research, Bali – Indonesia. (2008) 2707–2715.

[11] D. A. Huffman, Impossible Object as Nonsense Sentences, Machine Intelligence, Newyork: American Elsevier. 6 (1971) 295-323.

[12] L. A. Robinson, Graph Theory for the Middle School, Master Thesis, Faculty of the Department of Mathematics, East Tennessee State University. UMI Number: 1436267. 2006.

[13] T. R. Jensen and B. Toft, Graph Coloring Problems, Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons Inc., New York. A Wiley Interscience Publication, 1995.

[14] C. C. Iturriaga-Velazquez, Map Labeling Problems, PhD Thesis, University of Waterloo, Ontorio: Canada, 1999.

[15] D. Marx, Graph Coloring Problems and Their Applications in Schedulling, Periodica of Polytechnica Ser. El. Eng. 489 (2004) 11-16.

[16] D. Gaceb, V. Eglin, F. Lebourgeois, and H. Emptoz, Application of Graph Coloring in Physical Layout Segmentation, IEEE 19th Intl. Conf. on Pattern Recognition. (2008) 1-4.

[17] T. A. Redl, University Timetabling Via Graph Coloring: an alternative approach, University of Houston, Houston, 2007.

[18] T. Dobrolowski, D. Dereniowski, and L. Kuszner, *Koala Graph Coloring Library: An Open Graph Coloring Library for Real World Applications*, IT, Gdansk, Poland, May 2008.

M. Z. Matondang is a PhD student in computer science Universiti Teknologi Malaysia, Johor – Malaysia. He holds Master degree also in computer science from the same University in 2009 and Bachelor degree in mathematics from the Universitas Sumatera Utara, Medan, Indonesia in 2005. His researches interests are in operational research, soft computing and also in mathematical modeling.

A. A. Samah has received the Diploma and B.Sc. degree from University of Technology Malaysia in 1991and 1993 respectively. In 1996, she obtained her M.Sc. from the University of Southampton, UK and recently in 2010, she received her PhD from Salford university, UK. Currently she is a lecturer in Faculty of Comp. Science and Information System, University of Technology Malaysia. Her research interests encompass Image Processing, Soft Computing Techniques and Operational and Simulation Modeling.

H. Haron has received the Dip, B.Sc. degree and PhD from University of Technology Malaysia in 1987, 1989 and 2004 respectively. He was awarded M.Sc. from the University of Brighton, UK in 1995. Currently he is an Associate Professor in Faculty of Comp. Science and Information System, University of Technology Malaysia. His research interests include Image Processing and Computer Aided Geometric Design.

H. A. Majid has received the Dip and B.Sc. degree from University of Technology Malaysia in 1993 and 1995 respectively. In 1998, he obtained his M.Sc. from the University of Salford, UK. Currently he is a lecturer in Faculty of Comp. Science and Information System, University of Technology Malaysia. His research interests focused on Image Processing, Operations Management and, Warranty and Maintenance.

Attention Driven Face Recognition, Learning from Human Vision System

Fang Fang[1], Laiyun Qing[2], Chongxiu Wang[2], Jun Miao[3], Xilin Chen[3] and Wen Gao[4]

[1] School of Computer Science and Technology, Harbin Institute of Technology
Harbin, 150001, China

[2] Graduate University of Chinese Academy of Science
Beijing, 100049, China

[3] Key Lab of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences
Beijing, 100190, China

[4] Institute of Digital Media, Peking University
Beijing, 100871, China

Abstract

This paper proposes a novel face recognition algorithm inspired by Human Visual System (HVS). Firstly, we learn where people look by recording observers' eye movements when they are viewing face images. We find that the observers are consistent in the regions fixated and such fixated regions are selected as the salient regions. Secondly, we represent the face images by four scales of Local Binary Gabor Patterns (LGBPs) for the salient regions whereas one scale LGBPs for the others, inspired by the fact that fovea of HVS has a higher spatial acuity than the periphery. Thirdly, we integrate the global information of face images in face recognition. The experimental results demonstrate that the proposed method learning from human beings is comparable with those learned with machine learning algorithms, which shows that the characteristics of the HVS provide valuable insights into face recognition.

Keywords: *Face Recognition, Selective Attention, Human Visual System.*

1. Introduction

Face recognition has become a popular area of research in computer vision and one of the most successful applications of image analysis and understanding [1, 2] over the past several decades. Its' wide range of potential uses include security applications, intelligence-computer interaction and so on. The nature and scientific challenges of face recognition decides that not only computer science researchers are interested in it, but also neuroscientists and psychologists.

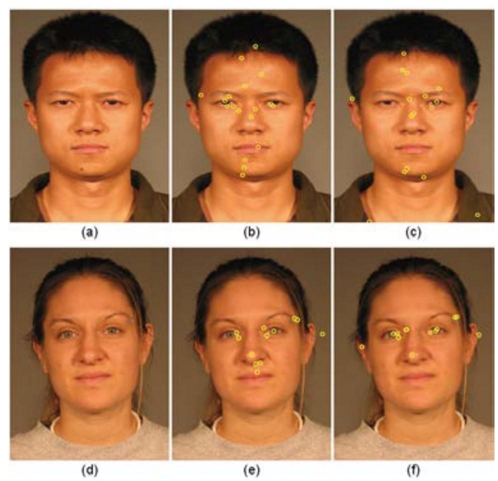


Fig. 1 Some examples of the recorded fixations, of which (b) and (e) are fixations of one observer, and (c) and (f) are fixations of another observer. The yellow circles represent the fixations.

Much progress has been made in the past two decades and numerous face recognition algorithms [3] have been developed. The most popular ones are based on machine learning, such as Eigenfaces [4] and Fisherfaces [5], SVM [6] and AdaBoost [7]. However, most of the statistical methods suffer from the generalizations problem.

It is the general opinion that advances in computer vision research will provide useful insights to neuroscientists and psychologists into how human brain works, and vice versa. Several algorithms [8, 9, 10] are based on Gabor filters, which are known to model the responses of the simple cells in vision cortex.

In this paper, we argue that we can take a further step following such direction as mimicking human vision systems. Though many face recognition systems sample the image in a grid-like fashion, human beings are able to select the most interesting regions to focus on (fixations) and jump (saccade) between them. some examples of the fixations when a observer is viewing a face image is shown in 1. This selection process is an important aspect of attention, and it has a profound impact on our perception [11]. Moreover, the spatial resolution of the human visual system decreases dramatically away from the point of fixation, which is attributed mainly to the fact that the ganglion cells are packed densely at the center of the retina (i.e. the foveola), and the sampling rate drops rapidly [12]. Therefore, the eye gathers most information during the fixations while little information is gathered during the saccades.

Inspired by the fact that HVS has a much higher spatial acuity than the periphery, a face recognition representation based on spatial variant sampling is proposed to simulate such foveated imaging phenomenon. We sample the image in a retinal way. The fixated regions (fovea) are filled with data from more Gabor filters, while less Gabor filters are reserved for the outside of the fovea. The output of the Gabor filters is further encoded by the Local Gabor Binary Patterns (LGBPs) as in [10].

The visual psychophysicists' research also has shown that human observers are able to obtain the layout of a scene within a short glance before any fixations [13]. Within such a glance, the grasped information provides useful information about spatial configuration and scene category. Therefore the global information of face images coded as the low frequency components gotten by Fourier transform is integrated into the face recognition system.

The remaining parts of this paper are organized as follows. Section 2 learns the salient regions of face images from the eye movement data. In section 3, face representation based on attention modeling which integrated both the information of salient regions and glance is presented. Extensive experiments are conducted in Section 4, followed by conclusions and discussions in the last section.

2. Selective Attention on Face Images

Though many popular face representations are based on grid sampling, human visual system works in a different way. In order to build a detailed representation of a scene, the human visual system actively scans the visual environment using discrete fixations linked by ballistic saccadic eye movements [14]. The eye gathers most information during the fixations while little information is gathered during the saccade. Such eye movement has been thought as the explicit form of visual attention.

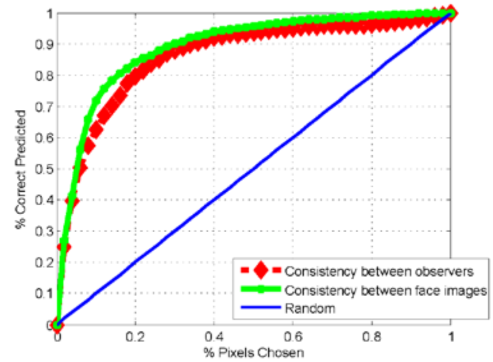


Fig. 2 Consistency of different observers viewing the same face image, and same observers viewing different face images. The consistency is defined as inter-subject consistency. We use fixations of all-except-one observers to generate a "saliency map" to predict fixations of the excluded observer. Y axis is the correct predicting rate and X axis is the selected region proportion. The curve of the consistency is much higher than random.

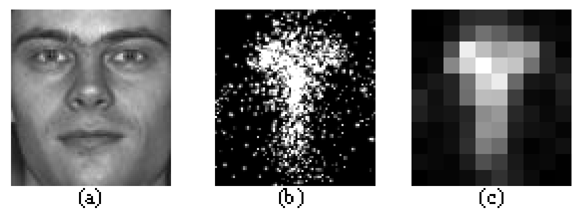


Fig. 3 (a) A cropped face image. (b) The statistical distribution of fixation positions. (c) The probability of each local region.

We use EyeLink II system [20] to record the eye fixations when observers are viewing face images. In the experiments, four male college students who are naive to the purpose of the experiments are selected as subjects. We present each face image to the subjects at the center of a 19 inch high refresh rate CRT monitor which has the resolution of 1600×1200. Each image is presented to the subjects for five seconds. Every subject views 100 frontal view face images from FRGC [2] dataset.

Some examples of the recorded fixations are shown in Fig. 1. It can be seen from Fig. 1 that, though the fixations vary among the observers and images, the regions which are most likely to be fixated are highly consistent.

We examine the inter-observer consistency among subjects' fixations, the operation is similar to "leave-one" method: for each stimulus, we use fixations of all-except-one observers to generate a "saliency map" to predict fixations of the excluded observer. The results averaged over subjects and stimulus are shown in Fig. 2. Similarly, the consistency of one observer for different face images is

shown in Fig. 2. It can be seen from Fig. 2 that the curve of consistency is much higher than random.
We get the distribution of the fixations from the samples of the fixations of the four observers on the 100 frontal view face images, as shown in Fig. 3(b). Because our face representation is based on regions and there are total 10×11 regions, we add the probability in each region and get Fig. 3(c), where greater possibility being fixated is indicated by whiter value. Fig. 3 demonstrates that HVS puts more importance on regions like two eyes, nose and mouth, which is well consistent with our intuition.

3. Face Recognition System Based on attention

According to the researches of the psychophysicists [13], human understand a scene by firstly glancing the scene, followed by fixating some interesting regions in it. We develop a face recognition system based on such selective attention mechanism, which integrate both of the information gathered by fixations and glance.

3.1 Face Recognition Based on Fixations

Despite a large field of view, HVS processes only a tiny central region (the fovea) with very great detail while the resolution decreases rapidly with increasing eccentricity [12]. We employ the multi-scale and multi-orientation Gabor filters to simulate such foveated phenomenon. The fixated regions are convoluted with more scales' Gabor filters to mimic the fovea. While in the non-fixated regions, less scales' Gabor filters are used to imitate the periphery, which forces the periphery contains much less information than the fovea.
We only list the parameters of the Gabor filters as follows:

$$\psi_{\mu,\upsilon} = \frac{\left\|k_{\mu,\upsilon}\right\|}{\sigma^2} e^{\left(-\left\|k_{\mu,\upsilon}\right\|^2 \|z\|^2 / 2\sigma^2\right)} \left[e^{ik_{\mu,\upsilon}z} - e^{-\sigma^2/2}\right] \quad (1)$$

Where μ and υ define the orientation and scale of the Gabor filters, $z = (x, y)$, $\|\cdot\|$ denotes the norm operator, and the wave vector $k_{\mu,\upsilon} = k_\upsilon e^{i\phi_\mu}$,where $k_\upsilon = k_{\max} / \lambda^v$ and $\phi_u = \pi\,\mu/8$. λ is the spacing factor between filters in the frequency domain.
In order to further enhance the information in Gabor filters, the magnitude values of the Gabor filters are further encoded with Local Binary Pattern (LBP) operator [15], namely LGBP [10]. The details of LBP are omitted for the limited space. Please refer to [10] for more details.
The input face images are normalized to 80×88 pixels and each image is divided into 10 × 11 non-overlapping regions. We choose 20 percent of the regions that have larger probability being fixated as the salient regions, as shown in Fig. 4(a), indicated by white value. All the others are non-salient regions.

Table 1: The accuracies of different choices of scales on the FERET datasets

Methods	***fb***	***fc***	***DupI***	***DupII***
LGBP[10]	**0.96**	**0.96**	**0.69**	**0.61**
WLGBP[10]	**0.98**	**0.97**	**0.74**	**0.71**
ν1=0	0.88	0.87	0.39	0.21
V1=1	0.94	0.96	0.56	0.41
V1=2	0.95	0.96	0.66	0.56
V1=3	0.96	0.96	0.70	0.64
V1=4	0.94	0.94	0.67	0.61
V1=3,V2=1,2,4	**0.97**	**0.97**	**0.74**	**0.72**

In the original LGBP [10], five scales $\nu \in \{0, 1, \cdots, 4\}$ and eight orientations $\mu \in \{0, 1, \cdots, 7\}$ Gabor filters are used, which means 5×8×110 = 4, 400 Gabor filters are used. In our experiment, in order to reserve more information in fovea and less information in periphery, four scales $\nu \in \{1, 2, 3, 4\}$ and eight orientations Gabor filters are employed for the 22 salient regions, while for non-salient regions, only one scale $\nu \in \{3\}$ Gabor filters are used. The reasons for choosing such scales can be found from Table 1, which will be discussed in Section 4. Therefore there are totally 32×22+8×(110-22) = 1; 408 Gabor filters are used, only 32 percent of the original LGBP [10].
Histogram intersection as in [10] is used as similarity measure when we compare two face images.

3.2 Face Recognition Based on Glance

As discussed above, human are able to obtain the layout of a scene within a short glance before any fixations. The information gathered during glance are the global information of the scene. We represent such global information by the lower frequencies gotten by 2-D Discrete Fourier Transform (DFT), though some other global features such as GIST can also be used for this purpose. These features are further processed by Fisher's Linear Discriminant Analysis (FLDA) [16] to reduce the dimension.
Finally, the cosine of the angle between two feature vectors based on glance is computed as the similarity measurement.

3.3 Face Recognition by Integration

The information gathered by fixations and glance, both of which are based on attention mechanism, are integrated into the final face recognition system.

Similarity based on glance between two input face images can be represented as $\eta(m_1, m_2)$, where m_1 and m_2 are feature vectors based on glance of two input face images respectively. And the similarity based on fixations between them can be represented as $\psi(H_1, H_2)$, where H_1 and H_2 are feature vectors based on fixations of two input face images respectively. Let ω_G denotes the weight of the similarity based on glance, and then the combined similarity of the two face images can be represented as follows:

$$C = \omega_G \times \eta(m_1, m_2) + (1-\omega_G) \times \psi(H_1, H_2) \quad (2)$$

4. Experimental Results

Experiments are conducted on FERET database [17] to evaluate the proposed method. FERET database includes 1; 196 objects, where the gallery set has 1; 196 face images, and the four probe sets are fb, fc, DupI and DupII, which has 1; 195, 194, 722 and 234 face images respectively. fb and fc were taken the same time as gallery with different expression and lighting respectively. The DupI were obtained anywhere between one minute and 1; 031 days after their respective gallery matches. The DupII were those taken only at least 18 months after their gallery entries.

4.1 Experimental Results of Fixations

Several choices of the scales are used to represent the non-salient regions are tested. Table 1 shows that the experimental results on FERET database, where ν_1 represents the scales used in all regions and ν_2 represents the scales used only in salient regions. It can be seen from Table 1 that, the 3rd scale has the best performance, and then followed by the 4th, 2nd and 1st scales, which is consisted with the studies that middle spatial frequencies provide information that is more useful for face recognition than are other frequency ranges [19].

Therefore we use the 3rd scale at all the regions of the face images and use the 4th, 2nd, and 1st scales in fixated regions additionally. The performance is better than original LGBP [10], which uses five scales in all the regions of the face images. The performance is also comparable with Weighted Local Gabor Binary Pattern (WLGBP) [10], which not only uses five scales in all the regions of the face images, but also different weights of different regions of each Gabor filter learned by Linear Discriminant Analysis.

We perform another two sets of experiments to further illustrate that the salient regions selected by human vision systems are reasonable. One set includes two randomly selected salient regions (see in Fig. 4(b) (c)).The accuracies of the randomly selected regions on FERET database are shown in Table 2. The performance with the regions fixated by human beings is better than those with randomly selected regions, which is consistent with our

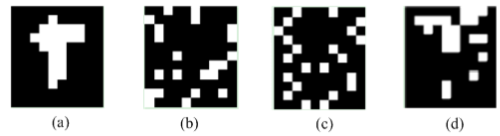


Fig. 4 Different choices of salient regions. (a) Salient regions selected by human vision systems. (b) Randomly selected salient regions set 1. (c) Randomly selected salient regions set 2. (d) Salient regions selected by AdaBoost.

Table 2: Different choices of salient regions random on the FERET datasets

Methods	***fb***	***fc***	***DupI***	***DupII***
Random Regions 1	0.95	0.96	0.68	0.61
Random Regions 2	0.95	0.96	0.69	0.62
AdaBoost	0.97	0.99	0.74	0.70
Adaboost Selected Regions	0.98	0.99	0.74	0.70
Fixations	**0.97**	**0.97**	**0.74**	**0.72**

Table 3: The accuracies of integration on the FERET datasets

Methods	***fb***	***fc***	***DupI***	***DupII***
WLGBP[10]	0.98	0.97	0.74	0.71
Integration	**0.99**	**0.99**	**0.76**	**0.71**

intuition. The other set experiments are performed by AdaBoost[18]. The results of AdaBoost and those of the most largerest weights regions selected by AdaBoost (Fig. 4(d)) are also shown in Table 2, both use the same LGBPs features. It can be seen that the performances of human selected regions are comparable to those of AdaBoost and AdaBoost selected salient regions. Please note that the face images used to collect eye fixations are not included in FERET datasets. Actually we hope that human beings are not so sensitive to datasets compared with machine learning algorithms (please note that AdaBoost are a little worse in the case of DupII, which maybe not been included in training set.)

4.2 Experimental Results of Integration

As mentioned in section 3, classifiers based on fixations and glance are combined to form the integrated classifier.

The weight of similarity based on glance $\omega_G = 0.13$. The choice of ω_G is according to experience, and the experimental results are not sensitive to ω_G range from 0.10 to 0.15. The experimental results on FERET dataset are shown in Table 3.

It can be seen from the results that, by combining the feature based on glance, the performance is better than WLGBP [10]. Experimental results demonstrate that the feature based on fixations and glance are indeed complementary for distinguishing faces, which is consistent with HVS.

5. Conlusions

This paper proposes a novel face recognition algorithm inspired by the selective attention of HVS. The regions are most likely to be fixated by human eyes are selected as salient regions. Motivated by the fact that fovea has a much higher spatial acuity than the periphery, a face representation based on spatial variant sampling is proposed to simulate such foveated imaging phenomenon of human eyes, where more information is reserved for the salient regions. Moreover, the information based on glance which adopts the low spatial frequency components of the image is integrated into the face recognition system to elicit a percept that occurs before any fixations.

The experimental results on FERET database demonstrate that the proposed method achieves comparable performance with those of machine learning algorithm such as AdaBoost, which shows that the characteristics of the HVS provide valuable insights into face recognition.

Though many valuable works have been done in selective attention and many computing models are proposed, there are few works that introduce such attention mechanism into face recognition and other object recognition. This paper is just an attempt to do such things. Please note that some components of the proposed framework can be modified. For example, LGBP [10] can be replaced with other alternative features to mimic fixations.

Acknowledgments

This research is partially sponsored by National Basic Research Program of China (No.2009CB320902), Beijing Natural Science Foundation (No. 4102013), Natural Science Foundation of China (No.60970087, 61070116, 61070149 and 61175115) and President Fund of Graduate University of Chinese Academy of Sciences (No.085102HN00).

References

[1] M. H. Yang, D. J. Kriegman, and N. Ahuja, “Detecting faces in images: A survey,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 1, pp. 34-58, 2002.

[2] P. J. Phillips, P. Grother, R. J. Micheals, D. M. Blackburn, E. Tabassi, and J. M. Bone, “Face recognition vendor test 2002: Evaluation report,” Tech. Rep. NISTIR 6965, National Institute of Standards and Technology, 2003. [Online]. Available: http://www.frvt.org.

[3] W. Zhao, R. Chellappa, J. Phillips, and A. Rosenfeld, “Face recognition: A literature survey,” ACM Comput. Surv., vol. 35, no. 4, pp. 399-458, 2003.

[4] M. Turk, A. Pentland, “Eigenfaces for recognition,” Journal of Cognitive Neuroscience,vol. 3, no. 1, 1991.

[5] P. Belhumer, P. Hespanha, and D. Kriegman, “Eigenfaecs vs. fisherfaces: Recognition using class specific linear projection,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 711-720, 1997.

[6] K. R. M¨uller, S. Mika, G. R¨atsch, K. Tsuda, and B. Sch¨olkopf, “An introduction to Kernel-based learning algorithms,” IEEE Transactions on Neural Networks, vol. 12, no. 2, pp. 181-201, 2001.

[7] P. Yang, S. G. Shan, W. Gao, S. Z. Li, D. Zhang, “Face recognition using Ada-Boosted Gabor features,” Automatic Face and Gesture Recognition, pp 356-361, 2004.

[8] L. Wiskott, J. M. Fellous, N. Kruger, C. Malsburg, “Face Recognition by Elastic Bunch Graph Matching,” IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 775-779, 1997.

[9] C. Liu and H. Wechsler, “Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition”, IEEE Transaction on Image Processing, vol. 11, no. 5, pp. 467-476, 2002.

[10] W. C. Zhang, S. G. Shan, W. Gao, X. L. Chen, H. M. Zhang, “Local Gabor Binary Pattern Histogram Sequence (LGBPHS): a Novel Non-Statistical Model for Face Representation and Recognition,” International Conference on Computer Vision, pp 786-791, 2005.

[11] G. Rizzolatti, L. Riggio, I. Dascola, C. Umilt_a, “Reorienting attention across the horizontal and vertical meridians: evidence in favor of a premotor theory of attention,” Neuropsychologia, vol. 25, pp. 31-40, 1987.

[12] Andrew T. Duchowski, “Eye Tracking Methodology: Theory and Practice,” Springer, 2nd edition, August 3, 2007, ISBN:1846286085.

[13] M. C. Potter, “Meaning in visual search,” Science, vol. 187, no. 4180, pp. 965-966, 1975.

[14] A. Yarbus, “Eye Movements and Vision”, New York: Plenum, 1967.

[15] T. Ojala, M. Pietik¨ainen, and T. M¨aenp¨aa, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp. 971-987, 2002.

[16] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K. R. Mullers, “Fisher Discriminant Analysis with Kernels,” IEEE Conference on Neural Networks for Signal Processing, 1999.

[17] P. J. Phillips, H. M. Syed, A. Rizvi, and P. J. Rauss, “The FERET evaluation methodology for face-recognition algorithms,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 10, pp. 1090-1034, 2000.

[18] Y. Freund, R. E. Schapire. “A Decision-Theoretic Generalization of on-Line Learning and an Application to Boosting,” In Computational Learning Theory: Eurocolt ’95, pp. 23–37, 1995.
[19] I. Boutet, C. Collin, J. Faubert, “Configural face encoding and spatial frequency information,” Perception and Psychophysics, vol. 65, pp. 1078-1093, 2003.
[20] http://.www.sr-research.com/.

Fang Fang received the B.S. degree in computer science from the Harbin Institute of Technology (HIT), Harbin, China, in 2004. She is currently working towards the Ph.D. degree at the School of Computer Science and Technology, HIT. Her research interests mainly include face recognition and image Saliency detection.

Laiyun Qing (S’03–M’09) received the B.Sc. and M.Sc. degrees in computer science from Northeastern University, Shenyang, China, in 1996 and 1999, respectively. She received the Ph.D. degree in computer science from Chinese Academy of Sciences, Beijing, in 2005.
She is currently an Associated Professor at the School of Information Science and Engineering, Graduate University of the Chinese Academy of Sciences, Beijing. Her research interests include pattern recognition, image processing, and statistical learning. Her current research focuses on visual perception and cognition.

Chongxiu Wang received the B.S. degree in Computer Science and Technology from Beijing Institute University (BIT), Beijing, China, in 2009. She is now working towards the M.S. degree at the Graduate University of the Chinese Academy of Sciences, Beijing, China. Her research mainly includes face recognition and object recognition.

Jun Miao (S’00–M’04) received the B.Sc. and M.Sc. degrees in computer science from Beijing University of Technology, Beijing, China, in 1993 and 1999, respectively. He received the Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, in 2005.
He is currently an Associated Professor at the Institute of Computing Technology, Chinese Academy of Sciences, Beijing. His research interests include artificial intelligence, neural networks, neural information processing, image understanding, and biological vision. He has published more than 30 research articles in refereed journals and proceedings on face detection, visual neural networks, visual neural information coding, neural oscillation, image segmentation, visual perception, and cognition. His two main contributions are the technique of Human Face Gravity-Center Template for face detection and the model of Visual Perceiving and Eyeball-Motion Controlling Neural Network for visual search, respectively.
Dr. Miao is a member of the China Computer Federation and a member of the Chinese Society for Neuroscience. He is the recipient of Microsoft Fellowship Award in 2000, a recipient of 2003 Shanghai Science and Technology Progress Awards (the First Award), and a recipient of 2005 National Science and Technology Progress Awards of China (the Second Award).

Xilin Chen (M’00–SM’09) received the B.S., M.S., and Ph.D. degrees in computer science from the Harbin Institute of Technology, Harbin, China, in 1988, 1991, and 1994, respectively.
He was a Professor with the Harbin Institute of Technology from 1999 to 2005. He was a Visiting Scholar with Carnegie Mellon University, Pittsburgh, PA, from 2001 to 2004. Since August 2004, he has been with the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, where he is also with the Key Laboratory of Intelligent Information Processing and the ICT-ISVISION Joint Research and Development Laboratory for Face Recognition. His research interests include image processing, pattern recognition, computer vision, and multimodal interfaces.
Dr. Chen has served as a program committee member for more than 20 international and national conferences. He has received several awards, including the State Scientific and Technological Progress Award in 2000, 2003, and 2005 in China for his research work.

Wen Gao (M’92–SM’05–F’09) received the B.Sc. and M.Sc. degrees in computer science from the Harbin Institute of Technology, Harbin, China, in 1985 and 1988, respectively. He received the Ph.D. degree in electronics engineering from the University of Tokyo, Tokyo, Japan, in 1991.
He is currently a Professor of Computer Science at Peking University, Beijing, China. Before joining Peking University, he was a Professor of Computer Science at Harbin Institute of Technology from 1991 to 1995, and a professor at the Institute of Computing Technology of Chinese Academy of Sciences. He has published extensively, including four books and over 600 technical articles in refereed journals and conference proceedings in the areas of image processing, video coding and communication, pattern recognition, multimedia information retrieval, multimodal interface, and bioinformatics.
Dr. Gao serves on the editorial board for several journals, such as the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FORVIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE TRANSACTIONS ON AUTONOMOUS MENTAL DEVELOPMENT, the EURASIP Journal of Image Communications, and the Journal of Visual Communication and Image Representation. He chaired a number of prestigious international conferences on multimedia and video signal processing, such as IEEE ICME and ACM Multimedia, and also served on the advisory and technical committees of numerous professional organizations.

Host-based Web Anomaly Intrusion Detection System, an Artificial Immune System Approach

Iman Khalkhali[1], Reza Azmi[2], Mozhgan Azimpour-Kivi[1] and Mohammad Khansari[3]

[1] School of Engineering and Science, Sharif University of Technology-International Campus
Kish Island, Iran
[2] Department of Computer Engineering, Alzahra University
Tehran, Iran
[3]Departmen of Network Science and Technology, Faculty of New Science and Technology,
University of Tehran

Abstract

Recently, the shortcomings of current security solutions in protecting web servers and web applications against web-based attacks have encouraged many researchers to work on web intrusion detection systems (WIDSs). In this paper, a host-based web anomaly detection system is presented which analyzes the POST and GET requests processed and logged in web servers' access log files. A special kind of web access log file is introduced which eliminates the shortcomings of common log files for defining legitimate users' sessions boundaries. Different features are extracted from this access log file in order to model the operations of the system. For the detection task, we propose the use of a novel approach inspired by the natural immune system. The capability of the proposed mechanism is evaluated by comparing the results to some well-known neural networks. The results indicate high ability of the immune inspired system in detecting suspicious activities.

Keywords: *Host-based Web Anomaly IDS, Enhanced Custom Log File, Artificial Immune System, Negative Selection Algorithm, Neural Network.*

1. Introduction

Nowadays, the World Wide Web (WWW) plays an important role in human life. Web applications are becoming increasingly popular in all aspects of human activities; ranging from science and business to entertainments. Consequently, web servers and web application are becoming the major targets of many attacks. Due to the growing number of computer crimes, needs for techniques that can secure and protect web servers and web applications against malicious attacks have been highlighted. Unfortunately, current security solutions, operating at network and transport layers, have insufficient capabilities in providing acceptable level of protection against web-based attacks. These issues have given rise to the ever evolving researches on web intrusion detection systems (WIDSs).

A WIDS dynamically monitors the input requests to the web server in order to decide whether a given set of requests is indicative of an attack or represents a normal web surfing activity. As the web servers record all the requests processed by them in access log files, these files could be considered as a major source of information that can be analyzed by WIDS. In order to detect web-based attacks, intrusion detection systems (IDSs) can be prepared with number of patterns of well-known attacks. These systems are called signature detection systems. Signature based IDSs typically require a signature to be defined for every possible attacks that can be performed by an attacker. Moreover, the set of signatures should be updated periodically in order to keep the system reliable. Although these systems can effectively detect known intrusion attempts, they are unable to detect novel attacks. Hence, these systems are vulnerable to zero-days attacks.

To overcome the shortcomings of signature based IDSs, anomaly detection systems are proposed for detecting unknown attacks. The anomaly based WIDSs try to build a profile of the normal states of the system and detect deviation from this normal profile that may indicate a suspicious activity. The main shortcoming of anomaly based WIDSs is that the false positive rate (proportion of events mistakenly detected as attacks) is usually higher compared to signature detection systems. Furthermore, it could be a difficult task to define what exactly a normal behavior is in complex environments.

Generally, anomaly detection approaches consist of two phases: training phase and testing phase. In the training phase the profile of normal behaviors is built. Then, it is applied to new data in order to detect abnormal behaviors in the testing phase. Various techniques have been applied to solve the general problem of intrusion detection. These techniques include statistical based methods [1], data mining methods [2-4], Markov models [5], [6], grammar based methods [7], [8], Artificial Neural Networks (ANNs) [9], [10], and Artificial Immune Systems (AISs) [11], [12].

As mentioned earlier, access log files of web servers are an important source of information for WIDSs. There are some shortcomings to common access log files generated by common web servers such as Apache. The first problem arises when defining web sessions. As the boundaries of sessions are not clearly defined, extraction of web sessions from these log files is not a straightforward process. Although there are some heuristics, such as time out scheme [13] that can lead us to identify web sessions, these methods are not very accurate in practice. On the other hand, common log files do not contain the POST requests processed by the web server. Generally speaking, both GET and POST methods have a key point in interaction of users with web applications. In order to overcome these issues, a log file generator is introduced that eliminates the aforementioned drawbacks in order to generate a special log file called the Enhanced Custom Log file (ECL). The detailed information of ECL is discussed later.

In this paper, an anomaly detection system for detecting web-based attacks is presented. In the training phase, the anomaly detection system takes the ECL as the input and builds a dataset to tries to learn how to distinguish normal behaviors from attack by considering three parameters. These parameters include: the number of values assigned to variables of each request within a session; the number of dual composition of characters, in windows of size two, from the values assigned to variables of each request within the session; and a window of size three from the users' navigational path within the sessions. Finally, we apply an Artificial Immune System (AIS), introduced in our previous work [27], in order to equip the detection system to distinguish normal behaviors from the abnormal ones. In order to evaluate the capability of the system, the system is encountered with new patterns in the testing phase. Also, to prove the ability of the proposed AIS algorithm the results are compared to different well-know ANNs.

The remainder of this paper is organized as follows. In Section 2, a review on some available IDSs is presented. Section 3 introduces the fundamental operations of the proposed anomaly based WIDS including data gathering, feature extraction and detection mechanism. In Section 4, the experimental evaluation of the proposed system is presented. Moreover, the detection ability of the system is compared to some common ANN. Finally in Section 5, we conclude our study.

2. Related Work

Generally speaking, an anomaly detector needs to have a proper definition of what normal behaviors are. To answer this question, there are two possible approaches. An anomaly detector can be provided by a set of rules or specifications of what is regarded as normal behavior based on the human expertise. This approach could be assumed as an extension of misuse detection systems. In the second approach, the anomaly detector automatically learns the behavior of the system under normal operations and then generates an alarm when a deviation is detected from the normal model [14].

Prior to introducing some related work, we should mention that the extendibility of a WIDS is proved when the training and test data are good samples of real world web attacks. Also, the attacks presented in these data need to be representative of variety of attacks that exist in current web. The most important datasets that has been used for evaluating IDSs are provided by DARPA/MIT Lincoln Laboratory in 1998 and 1999 [15]. These data are used by many researchers in intrusion detection researches because massive datasets are rare and also they provide an immediate comparison with the original Lincoln Lab's test. As there are some inherent problems with this datasets [16], [17], some researchers improved their own attack database to supplement the attacks in the Lincoln Lab's data [18].

Various features are proposed to be extracted from the audit data that is available in a particular system. Krugel et al. [19], proposed a service-specific IDS. In this system, the anomaly score of a request is calculated using three properties: type of the request, length of the request and payload distribution. In this approach, if the length of the request is longer than the average length, it is likely to be an attack. This approach is based on the observation that attack codes, which may cause buffer overflow, often contain a large number of NULL characters. In the case of a model for character distribution, the ASCII characters are grouped to 6 categories. Then a single uniform distribution model of these 6 groups for all requests of one service over all possible payloads' length is computed. Finally, a chi-square test is performed against this model to calculate the anomaly score of test requests. In a similar attempt, anomaly detection system proposed by Wang and Stolfo [14] models the normal payloads of the network traffic by profiling the byte frequency distribution of payloads of particular length, flowing a particular host and port. Then, the size of the model is reduced using a clustering method. In the detection phase, the Mahalanobis distance is used to calculate the similarity of new data to the normal profile.

The system proposed by Kruegel and Vigna [20], is an IDS that use similar features to our work. This system utilizes a number of anomaly detection techniques to detect attacks against web servers and web-based applications. The system analyzes the HTTP GET requests that use parameters to pass values to server-side programs. These requests are recorded in access log file of web servers. Multiple models are produced using a wide range of features of the mentioned requests. These features include the presence or absence of attributes (parameters), attributes length, attribute characters distribution, the order of attributes, access frequency, inter-request time delay, etc.

Different algorithms are proposed to generate the normal model of the systems and distinguish the normal behaviors from the abnormal ones. As discussed earlier, some anomaly detection systems produce a statistical model for a system and detect large deviation from this profile [14], [19], [20]. The others use machine learning techniques or Markov chains in order to make a profile of normal operations of system [21]. The session anomaly detection (SAD) system [22] analyzes web access logs in order to detect anomalous web sessions. In this system, the Bayesian parameter estimation is adopted to build profiles of normal webpage request sequences. Then, the likelihood of each event is estimated based on previous probability distribution and an anomaly score is assigned to each event. The anomaly detection system presented in [23], have applied Markov chains in order to model the HTTP traffic. In the training stage, the packet payloads of the training data (normal data) are segmented into a number of continuous blocks. Then, a dictionary of these blocks is produced and a symbol is associated to each entry in the dictionary. Each segmented payload is represented as a sequence of corresponding symbols from the dictionary. The arrays of symbols are then used to train a Markov chain to capture the occurrence and the spatial appearance of each sequence within the data. In the testing stage, the obtained Markov model is used to evaluate the incoming HTTP traffic. In some researches, ANNs have been applied in anomaly detection, mainly due to their adaptability to changes in environments. Furthermore, ANN can model complex relationships between inputs and outputs and also find patterns in data by using non-linear statistical data modeling. Ingham et al. [18] also developed their framework for comparing different anomaly detection techniques based on their gathered data.

In recent years, a growing number of computer scientists have applied immunological models to several domains of computer and network security [24], [25]. These systems are characterized by some interesting features such as adaptability, self-organizing and distribution. Among the AIS models, the Negative Selection (NS) algorithm has a high potential use in intrusion detection. The basic idea of this algorithm is to generate a set of detectors that are able to distinguish self (normal) from the non-self (abnormal) behaviors. Gonzales et al. [26], [25] proposed a real-valued negative selection to generated non-self samples using the self data. A conventional classifier is then applied to detect abnormal patterns.

3. Proposed Method

In this section, different parts of the proposed Web Host-based Intrusion Detection System (WHIDS) are described. The general view of the proposed system is depicted in Fig. 1. As illustrated in this figure, the proposed WHIDS is composed of five major stages. In the first stage, with the assistance of the PHP log file generator, the HTTP streams recorded with their sessions in ECL format are fed to the system as input. In the second stage called the feature extraction stage, multiple features are extracted from the input data. In the training stage, a learning algorithm is employed in order to learn the states of the system regarding the features that are extracted in previous stage. The fourth stage involves analyzing the incoming sessions and comparing them to learned states models, built in training stage, in order to detect malicious activities. The last stage handles the output of the WHIDS. The output of the proposed system is an alarm signal for the system security officer when a suspicious activity is detected. The following sections describe the detailed information of each stage.

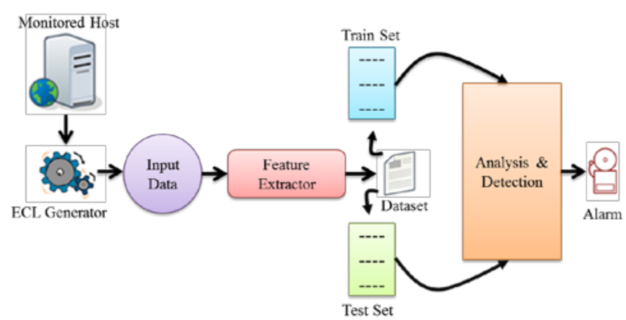


Fig. 1 The general view of the propose WHIDS.

3.1 Data Gathering

Generally, in order to fully examine an IDS a suitable dataset is necessary. The key point is to accurately separate the normal data from the abnormal (attack) data. To achieve this goal, in the data gathering phase, the web applications and web server should be provided by a fully protected condition to ensure that we can gather the pure normal user behaviors. On the other hand, in order to prove the capability of the system in detecting various attacks, the attack data should cover large range of today's web attacks.

As mentioned earlier, our anomaly detector analyzes a special kind of access log in order to build a normal profile of web sessions and also detect malicious activates. The reasons that the common log files (Apache log files) are not employed are as follows. First, the proposed WHIDS analyses both GET and POST HTTP requests received by the web server. Generally speaking, most PHP applications work with GET or POST methods and both methods are needed to be used alternatively for interacting with users. As the POST requests are not recorded in common log file of web servers, a new log generation mechanism is needed to record these requests in log files. It should be noted that there are other methods for interacting with web application such as HEAD and PUT, which are ignored because of their low importance. Second, the boundaries of sessions are not clearly defined in common log files. As a result, extracting web sessions from these log files is not a straightforward process. Although there are some heuristics, such as time out scheme that can lead us to identify web sessions, these methods are not very accurate in practice.

Considering the aforementioned problems, we develop a module called PHP log generator, written in PHP language, in order to generate a log file that satisfies the needs of our anomaly detector. The output of the PHP log generator would be a special kind of log file called Enhanced Custom Log file (ECL). An ECL entry represents a request to the web server which belongs to a specific session. The fields of an ECL entry include: ID (sequence number), session identification (ID), client IP address, time of the request, date of the request, method of the request (POST or GET), host name, requested URL, query string (the list of variables and values which are passed to web application), network address translation (NAT) IP address, forward IP address, user agent, protocol, server port number, and client port number.

In order to clarify the functionality of the PHP log generator, consider the process of generating common log file in Apache web servers. This process consists of 5 steps (see Fig. 2). In the first step, a user writes his/her request in the URL and sends the request to the web server. In the second step, PHP interpreter receives the request and fetches the PHP file and interpret it to HTML. Then, it sends HTML file for the user. Prior to this step, the request is recorded in common log file in Apache web server.
In the case of the method employed in this research, PHP log generator captures users' requests before allowing the PHP interpreter to fetch the PHP file (see Fig. 3). Furthermore, a cookie is created which contains the session ID of the visitor. This cookie is stored in the user's machine and helps the log file generator to recall the user in future connections; since each time the user sends a request, the session ID is enclosed in that request. It is worth bearing in mind that, expiration date of the cookies are settable by the log file generator when it aims to set session ID of the cookie.

In the method used in this research, the session is generated by PHP log generator and is then recorded in ECL. Since the session is a unique identity which is assigned to users, it allows us to track clients. Session generation has some difficulties when the client closes the browser and re-opens it again. The process of restarting the session with closing and opening the browser can lead to mistake in tracking users. In other words, when users open and close the web browser in different times, more than one session are created and logged for them wrongly. To avoid this problem, the cookies have been used. By employing cookies, not only sessions are generated, but also they can be stored for a specific period of time.

3.2 Feature Extraction

In this section, the process of generating a dataset containing different features that are extracted from the ECL is explained. The dataset, which is generated by the system, can be used by other researchers as a benchmark. As mentioned earlier, three different features are extracted by the feature extraction module in proposed WHIDS. These features include: the number of values assigned to variables of each request within a session (Histogram); the number of dual composition of characters in windows of size two from the values assigned to variables of each request within the session (Double windows); and a window of size 3 from the users' navigational path within the sessions (Markov windows). Consider that, in feature extraction process, input data are actually web sessions that can consist of multiple requests to the web server.

Histogram: Generally, a request to a web application may contain some variables and values. The length of a value assigned to a variable can be helpful in detecting anomalous requests. Generally, values can be either fixed-size tokens (such as session identifiers that are entitled in the request to resume a session) or short strings representing human inputs (e.g. fields in an HTML form). As a result, the length of the values assigned to variables of a particular web application does not very much. The

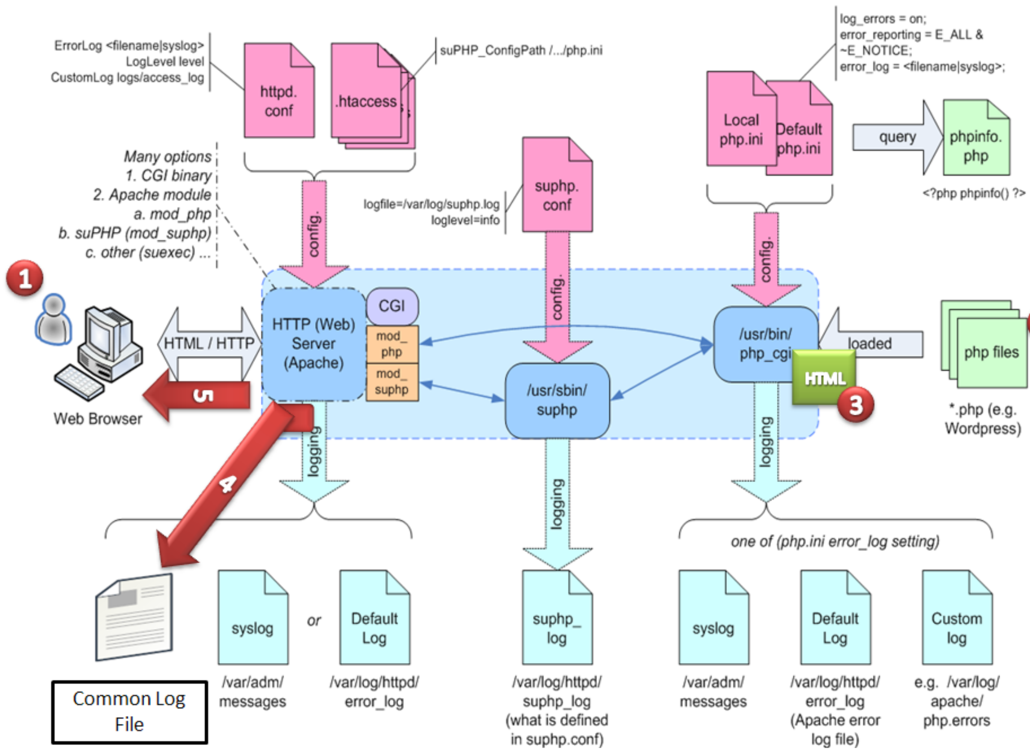


Fig. 2 The processes of common log file generation in an Apache web server.

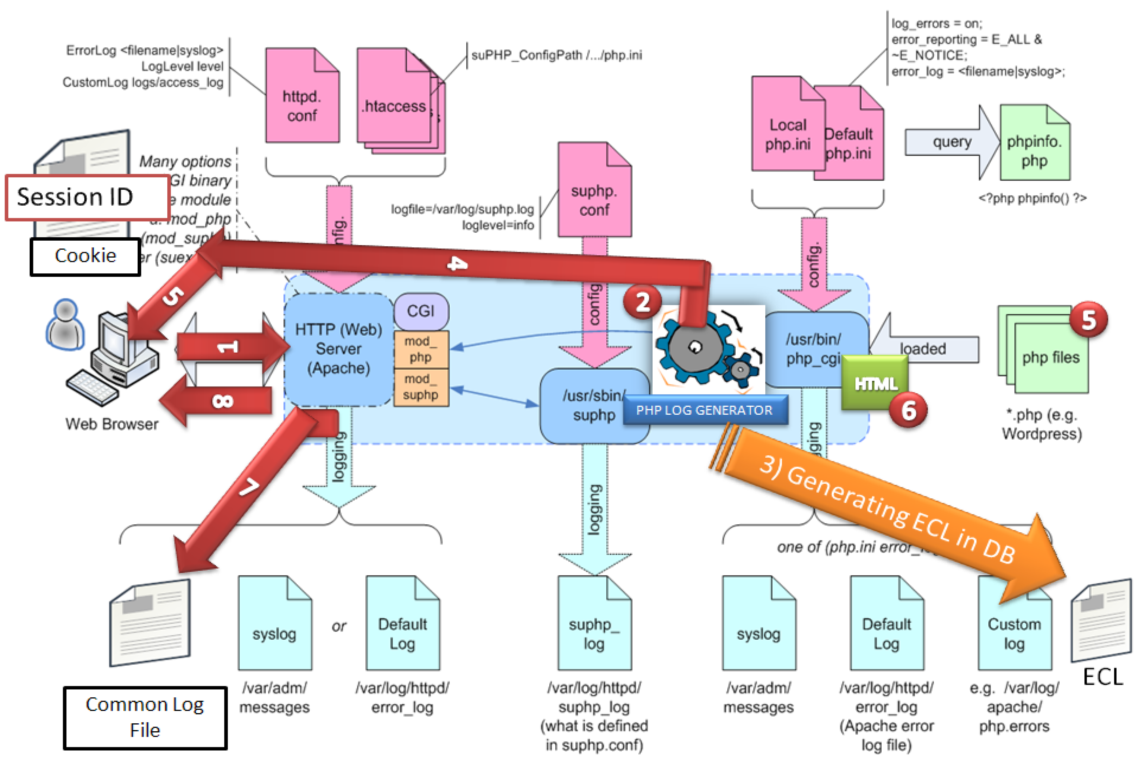


Fig. 3 The process of ECL generation in Apache web server.

situation can be different when a malicious activity is occurred. As an example, in the buffer overflow attack, the length of the input value which is passed to the web application may vary from the normal range.

Based on the above discussion, for the first feature that is proposed to be extracted from the ECL, called Histogram, we extract the variables’ values and count their character. These values are stored in a database whose columns are defined as particular intervals with the length of four. Intervals are assumed for the number of characters in each variable. For each variable in the requests of a session, we count the number of its characters. Then, the corresponding field’s value in the table of the database is incremented by one.

As an example, imagine a login request in a session which contains a password variable assigned with the value of ‘XXX’. As the length of the value for this example is 3, the value for the corresponding column, $1 \leq x < 5$, is increased by one in the row that is correspond to the session of the request. Consider that, after counting the length of each variable in a session, the vector that is corresponded to that session is normalized. That is why the range of output is between 0 and 1. The normalized value for each field in a vector of a session is calculated by dividing the value of that field by the sum of values over all the fields in that vector.

Double Windows: Generally, the parameters that are passed to a web application have regular characteristic. As an example, they mostly contain printable characters. In the case of a malicious activity, the structure of the characters which are passed to the web application could be quite different. Moreover, many attacks would send a series of a single character. As a result, modeling how different characters can appear in a normal query can be useful in detecting abnormal occurrence of them.

Based on the above discussion, the second feature that is extracted from the ECL is the number of dual composition of characters, in windows of size two, from the values assigned to variables of each request within the session. Generally, the characters that are allowed to be used as the value of a variable include ‘~’, ‘!’, ‘@’, ‘#’, ‘$’, ‘%’, ‘^’, ‘&’, ‘*’, ‘(‘, ‘)’, ‘_’, ‘+’, ‘`’, ‘-‘, ‘=’, ‘[‘, ‘]’, ‘\’, ‘;’, ‘"’, ‘,’, ‘.’, ‘/’, ‘{‘, ‘}’, ‘|’, ‘:’, ‘<’, ‘>’, ‘?’, ‘"’, and the letters ‘a’ to ‘z’. Here, we attempted to create all possible double composition of the aforementioned characters. In this case, 3364 doubles can be created. As in the case of histogram, we have one column for each double and increment the counter of each column when the corresponding double is found in the value assigned to a variable of a request in a session.

The high dimensionality of the produced table leads us to add columns to the table on demand. In other words, it is not necessary to add all possible doubles as a column to the table, but they are added only if they are seen at least in a window of values assigned to variables of a request. For example, if the word ‘admin’ is assigned to a value of a variable, the doubles that are added to the system will be ‘ad’, ‘dm’,’mi’ and ‘in’. Fig. 4 (a) shows the table when all doubles are put into the table, whereas Fig. 4 (b) shows the table when doubles are added when they occur in the variables’ values. The advantage of the second approach is that many columns will be discarded and the produced table would be less sparse. Furthermore, this reduction decreases memory usage.

	ab	ac	ad	ae	...
SESSION ID 1	0	10	0	0	..
SESSION ID 2	0	20	0	0	..

(a)

	ac	bd	%a	<>	...
SESSION ID 1	3	3	82	7	..
SESSION ID 2	7	5	6	0	..

(b)

Fig. 4 The schematic view of the doubles window table. (a) When all possible doubles are put into the table. (b) When doubles are added to the table on demand.

Markov Windows: Markov models are well-known as a method to capture the sequences of web navigation in anomaly detection systems. This model can detect unreasonable transition of user and mitigate authentication bypass attack. In addition to the sequence of web navigations, we consider the variables which are passed in the request by defining a new concept named treat.

We consider various windows of the user’s navigational path within the sessions. Here, for the sake of efficiency the size of the windows are set to three. Assume that a website is included these pages: a, b, c, d, e, f, and g (the page a represent index.php or something similar). Accordingly, a user can have different navigational patterns for visiting pages (as an example a-b-c >> b-c-d >> c-d-e >> d-e-f). As mentioned, we also consider the variables which are passed through the query in conjunction with the navigational path of the user and call these a treat. By introducing the treat concept in previous example, the page “index.php” will be changed to “index.php? variable1=”. Advantage of using treats instead of pages is that we can model the navigational path of a user in conjunction with the variables which are passing during this navigation. As an example, three queries: “index.php?var1=”, “index.php?var2=”, and “index.php?var1&var2=” are considered as different treat although the visited web page is identical in them (see Fig. 5). In other words, it is possible to go from one page to another by passing different kind of variables without considering their values.

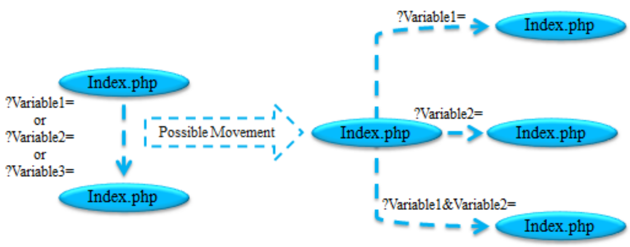


Fig. 5 The concept of treat against the movement.

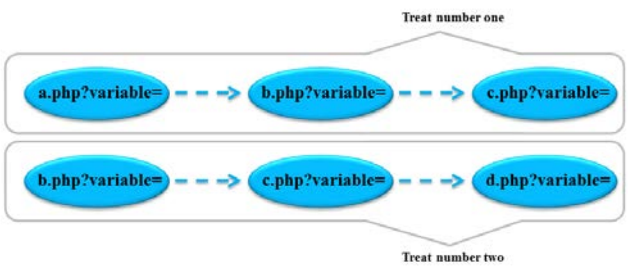


Fig. 6 The treat model.

Eventually, we make a table of treats which are observed in sessions. Similar to previous features, all possible triples are not added as a column to the table, but they are just added when they have been seen in a users' sessions. Hence, the table contains triples that have been generated at least one time by one user. It should be mentioned that, repeated triples will not be taken into account. Fig. 6 represents two treats stored in the system. After specifying all triples in users' sessions, we try to count the number of occurrence of each treats in each session. Hence, if we observe a particular treat in a session, the corresponding column, in the row related to that session, should be increased by one. Finally, the model is provided by normalizing these counts as explained previously.

3.3 Analysis and Detection

As it was mentioned before, the WHIDS system consists of an analysis and detection module. This module receives the dataset generated in previous stage as input and analyzes it to learn the normal behaviors and abnormal behaviors. In other word, we have divided the dataset to training and testing data. The training data is used to train the system to distinguish the normal behaviors from the abnormal ones. Finally, the test data is fed to the trained algorithm in order to evaluate the capability of the algorithm in detecting malicious activities.

The analysis and detection module is completely independent from other parts, so it allows us to be able to alternate the learning algorithms. Previously, we proposed a real-valued negative selection (RNS) algorithm for binary classification [27]. We use this AIS algorithm as the analysis and detection module in the proposed system. Briefly, this algorithm tries to generate some real-valued detectors. We inspired by the universal gravitational law to spread the detectors in the problem space in order to cover the non-self space properly. The detailed information of this algorithm is presented in [27]. After generating the detectors from the training samples, the testing samples are presented to the detector set and are compared to them based on the Euclidean distance. The samples which are closed to detectors sufficiently are considered as a potential suspicious activity.

In order to prove the capability of the immune inspired algorithm, this algorithm is replaced by some well-known ANNs. The results of the effectiveness of these algorithms in detecting suspicious activities are compared to each other in following sections.

3. Experimental Evaluation

In this section, we discuss our approach to evaluate the effectiveness of the negative selection algorithm in detecting malicious activities in web server log files. Moreover, we examine the capability of the features that are extracted from log files to capture the properties of attributes that have been collected in ECL files.

Beside the proposed RNS algorithm, three different algorithms including Multilayer Perceptron (MLP), Radial Basis Function Network (RBFN), and Naïve Bayes (NB) algorithm are chosen to learn the dataset which is generated from the ECL log files. Both MLP and RBFN have been employed in network based IDSs. The reason that they are chosen is that, they proved to have high ability in detecting intrusion in the network based systems. Consider that, due to the independency of the detector system's modules from the analysis and detection module, it is possible to change the algorithms employed in this part or use the combination of that algorithms as a new one.

As described earlier, due to the special kind of the attributes that we defined to model the behavior of the system, we used ECL. Also, for producing ECL we designed a social website and collect access log files in ECL format. For evaluating the effectiveness of the detection system, we need both normal and attack data. Therefore in the first step, we provided the web server with a fully protected condition to ensure that we can gather the pure normal users' behaviors. In the second step, we produced the variety of attack such as SQL injection, path traversal, command injection, XSS, etc. to a web server which in not under any external traffic for collecting attack data. After data gathering, we analyzed the log files to extract features that are comprehensively described in previous sections. Finally, a dataset is generated which can be analyzed by different algorithms to learn and detect the system behavior. Table 1 provides details information of important properties of the generated dataset from the ECL files. The table shows the time interval during which the ECL files were recorded, the log file size and the total number of HTTP requests in the log file.

Table 1: The details information of the dataset used for evaluation.

Dataset	***Time Interval***	***Size (MB)***	***HTTP Requests***
Normal Data	7 days	5	30,000
Attack Data	7 days	21	156,700

As it was mentioned earlier, three different categories of features are extracted from the ECL. The structure of the final dataset is as follows. There are 10 columns for the histogram. In addition, there are 1160 columns for doubles which 506 columns are share between attack, and normal and 654 columns are not share between attack, and normal. Finally, in the case of extracted treats, 30 fields or columns are shared as a normal and attack treat and 3474 fields are not shared. Among these non-shared fields 2486 columns are normal treat and 988 columns are attack.

4.1 Detection Effectiveness

An important key point in evaluating the system is choosing evaluation measures. Suitable measures can show real nature of the system and performance of algorithms. In this section, we compare the algorithms used in analysis and detection module from three different points of view including: the time they spend for learning the normal behaviors in training phase and detecting intrusions in testing data; the error of predicted value compared to the real value; and the measures revealing how much the system can detect anomalous behaviors.

The testing technique, employed in all the tests performed in this section, is 10 times 10-folds cross validation, which can guarantee the generality of the results. In other words, we run each algorithm 10 times with 10-folds cross validation and the final values for evaluation measures is the average of these 10 runs. Also, in the case of the proposed RNS algorithm, the number of generated detectors in the evaluations presented in this section is set to 50.

Execution Time: One of the important factors in the field of WHIDS is time the system spends to model the monitored system. In WHIDS systems, the faster we are able to detect the intrusion, the sooner we can find the weakness points of the system and patch the holes. The time measure in WHIDS is the time that it takes for the algorithm to build a model for the training samples and predicting the status of the test samples. However, it is worth mentioning that, execution time alone, is not enough to evaluate an algorithm and the time should always be seen with the ability of the system in detecting intrusions. For example, if an algorithm is fast and needs a very short time to build a model but has a low ability in detecting anomalous behaviors, it cannot be a useful algorithm. On the other hand, an algorithm that takes an enormous amount of time to properly build the normal model and detect intrusions is unsuitable; since, by the time the intrusions are detected, the system may have been already damaged too much.

Table 2 shows the execution time of each four algorithms, employed in analysis and detection module, over the dataset. As we can see in Table 2, among the four algorithms that were mentioned (proposed RNS, Naïve Bayes, RBFN and MLP), Naïve Bayes has the shortest execution time, followed by RBFN, NS, and finally MLP. In other words, MLP takes the longest time while the NB and RBFN spend a short time and proposed RNS is in the middle. It should be noted that, due to the very long time that MLP takes to respond, even if it performs very well on other measures, it cannot be an efficient algorithm for some environments.

Table 2: The detailed information of the dataset.

Algorithm	***Time taken to build model (sec)***
Proposed RNS	297.95
NB	0.20
MLP	2973.40
RBFN	9.30

Error: In this part different kind of error metrics are measured to evaluate the error in the prediction values compared to the real value of the test set. The results are compared for each algorithm in analysis and detection module. Consider that, in the case of the proposed RNS algorithm, the calculation of errors is not a common task. For this reason we ignore the calculation of error measures for proposed RNS algorithm.

Therefore, three different error metrics are measured for each learning algorithm. The first metric is the Mean Absolute Error (MAE). As the name suggests, the mean absolute error is an average of the absolute errors and shows how the predicted values differs from the true values. The closer the prediction value to the true value, the smaller mean absolute error would be.

The second error metric employed in our evaluation is the Root Mean Squared Error (RMSE). RMSE is a measure of error that is used frequently for evaluating algorithms. It is similar to MAE since like MAE, it measures the average difference between prediction value and true value. However, there is one important difference between MAE and RMSE. RMSE is more sensitive to outliers than MAE. The RMSE is calculated such that the effect of the large differences between predicted value and true value are amplified. In other words, larger differences between predicted value and the true value have a greater power to affect the obtained value of error. On the other hand in MAE, errors of different sizes are all treated in the same way. Both RMSE and MAE are useful measures of error. However, in some situations using RMSE is more meaningful while in other situations using MAE is more appropriate.

The last measure for reflecting the error corresponds to each algorithm in this evaluation is the Kappa Statistics. This metric is a statistical measure which shows the consistency between predicted and true value in a dataset. Also, it corrects the consistencies that occur by chance. High value of Kappa shows high consistency between predicted and measured values and indicates better performance of algorithm.

Table 3 shows the results for the aforementioned error measures for each three learning algorithms. As suggested by this table, the predicted values by MLP algorithm have lower error in comparison with the other two algorithms. Moreover, the RBFN is more successful than Naïve Bayes. However, it should be mentioned again that the error measurements, like all other measures, cannot be used alone to evaluate an algorithm. As we discussed in previous section, the MLP spends an enormous amount of time to construct the normal model and gets the results.

Table 3: The error measures for each three learning algorithms.

Algorithm	*MAE*	*RMSE*	*Kappa Statistics*
NB	0.27502	0.52325	0.34114
MLP	0.04959	0.18309	0.90605
RBFN	0.05329	0.20916	0.89822

Efficiency Measures: In this section different kind of metrics are measured to evaluate the ability of the algorithms to learn the properties of the features of the data and also detecting the malicious activities. The results are presented for each algorithm in analysis and detection module.

Generally, four situations can be assumed corresponding the relation between the result of an analysis for a sample event and its actual nature in an IDS. These situations include: false positive (FP), if the analyzed event is not an attack, but it is classified as a malicious activity; true positive (TP), if the analyzed event is correctly classified as intrusion; false negative (FN), if the analyzed event is malicious, but it is classified as a normal activity in the system; and true negative (TN), if the analyzed event is correctly classified as a normal activity.
Considering these situations, two measures are defined that can represent the effectiveness of a detection tool. These two measures are the TP rate (recall) and FP rate (false alarm rate). The recall measure can be defined as the fraction of correctly classified instances to the total number of inputs that should have been classified as anomalous. In other words, recall is the fraction of true positive rates to the number of all cases that should have been classified as positive. On the other hand, the false alarm rate can be defined as the proportion of actually normal cases that were incorrectly classified as anomalous.
There are some other measures that can reveal the effectiveness of the detection system. As instance, we can consider the Precision. The precision can be defined as the number of correct predictions divided by all the results that have been specified by the algorithm. A measure that can show the combination of the precision and recall is F-measure, which is provided by taking the harmonic mean of these two values.

Table 4 represents the discussed measures for each of four algorithms when 10 times 10-fold cross validation test method was performed over the dataset. As suggested by this table, the MLP and RBFN are performing quite the same and they can excel the NB algorithm in detecting attacks correctly and not producing incorrect alarms. However, we can see that the false alarm rate generated by the MLP algorithm is slightly smaller compared to that of RFBN. Moreover, the recall and the precision, and hence the F-measure corresponding to MLP exceed that of RBFN. In additions, the proposed RNS algorithm has the highest values for recall and precision while producing the least false alarm. As a result, we can claim that the proposed RNS algorithm is performing more effectively in detecting malicious activities compared to other three algorithms.

Table 4: The results of the efficiency measures for each four learning algorithms.

Algorithm	***Recall***	***Precision***	***F-measure***	***False Alarm Rate***
Proposed RNS	0.9679	0.9795	0.9834	0.0007
NB	0.7449	0.7607	0.6862	0.4217
MLP	0.9561	0.9571	0.9557	0.0610
RBFN	0.9524	0.9533	0.9523	0.0656

4.1 Discussion

In this section we summarize the obtained results from the evaluation mechanism presented in previous sections. As discussed earlier, the execution time of the MLP is too higher than that of the NB algorithm and RBFN. Furthermore, we can see that the NB is the fastest algorithm in building the normal model and detecting anomalies. Also, the proposed RNS algorithm has a mediocre execution time.

The errors of the predicted values for the NB algorithms are the highest among all three algorithms, indicating how NB is performing poorly in predicting the values for sample data. MLP is performing more accurately in predicting the values corresponding to data samples. Finally, the RBFN, with a slight difference from MLP, is performing quite powerfully in predicting the values for data samples. As discussed, the error measures are not calculated for the proposed RNS algorithm.

On the other hand, from the efficiency point of view, the proposed RNS algorithm is performing more powerfully in detecting anomalous behaviors with generating fewer mistaken alarms compared to other three learning algorithms. As the results of Table 4 suggest, the ability of the MLP and RBFN are quite the same. Furthermore, we can conclude that the NB algorithm is performing poorly in detecting attacks, although it is the fastest algorithm.

To sum up, we can claim that, from both the cost (execution time) and the efficiency points of view, the proposed RNS algorithm can be selected as the best choice for the analysis and detection module of the proposed WHID among the other three algorithms discussed in this paper.

4. Conclusions and Future Work

The main goal of this research was designing a host-based WIDS. We proposed to employ the enhanced custom log file in order to eliminate the inherent problems of common log files in defining web sessions boundaries. Moreover, ECL provides us with the POST requests along with the GET requests from the HTTP protocol. Different features were extracted from the ECL file which can represent the operations of the monitored web server. In this research, a dataset of normal and attack data were produced which can be used by other researchers in the field of WIDSs. Finally, we proposed the use of a novel RNS algorithm, inspired by the natural immune system, in order to produce a set of detectors that can cover the space of non-self (attack) properly and match to the non-self data and detect them. The results presented in this paper, proved the high ability of the proposed algorithm in detecting abnormal activities compared to some well-know and classical learning algorithms.

The most significant improvement that can lead to fruitful this research in future is to prepare the system to perform in an online state. As discussed earlier, the AISs have some interesting characteristics, such as adaptability and self-organizing, that make them attractive for working on data streams, such as HTTP streams, instead of stored data in an offline manner. This can also be a beginning for designing host-based web intrusion prevention systems.

References

[1] D. Qu et al., “Statistical anomaly detection for link-state routing protocols”, in Proceedings of the 6th International Conference on Network Protocols, 1998, pp. 62-70.

[2] W. Lee, S. J. Stolfo, and K. W. Mok, “A data mining framework for building intrusion detection models”, in Proceedings of the 1999 IEEE Symposium on Security and Privacy, 1999, pp. 120–132.

[3] A. Nalluri and D. C. Kar, “A web-based system for Intrusion Detection”, Journal of Computing Sciences in Colleges, vol. 20, no. 4, 2005, pp. 274–281.

[4] K. Sequeira and M. Zaki, “ADMIT: anomaly-based data mining for intrusions,” in Proceedings of the 8th International Conference on Knowledge Discovery and Data Mining, ACM SIGKDD, 2002, pp. 386–395.

[5] Y. Song, A. D. Keromytis, and S. J. Stolfo, “Spectrogram: A mixture-of-markov-chains model for anomaly detection in web traffic,” in Proceedings of the 16th Annual Network and Distributed System Security Symposium (NDSS), 2009.

[6] J. Hu, X. Yu, D. Qiu, and H. H. Chen, “A simple and efficient hidden Markov model scheme for host-based anomaly intrusion detection,” IEEE Network, vol. 23, no. 1, 2009, pp. 42–47.

[7] Z. Su and G. Wassermann, “The essence of command injection attacks in web applications,” in Proceedings of

33rd Symposium on Principles of programming languages, ACM SIGPLAN, 2006, vol. 41, pp. 372–382.

[8] V. Gorodetski and I. Kotenko, "Attacks against computer network: Formal grammar-based framework and simulation tool," in Recent Advances in Intrusion Detection, 2002, pp. 219–238.

[9] G. Wang, J. Hao, J. Ma, and L. Huang, "A new approach to intrusion detection using Artificial Neural Networks and fuzzy clustering," Expert Systems with Applications, vol. 37, no. 9, 2010, pp. 6225–6232.

[10] E. Corchado and Á. Herrero, "Neural visualization of network traffic data for intrusion detection," Applied Soft Computing, vol. 11, no. 2, 2011, pp. 2042–2056.

[11] U. Aickelin, J. Greensmith, and J. Twycross, "Immune system approaches to intrusion detection–a review," Artificial Immune Systems, 2004, pp. 316–329.

[12] F. S. de Paula, L. N. de Castro, and P. L. de Geus, "An intrusion detection system using ideas from the immune system," in Congress on Evolutionary Computation, CEC2004, vol. 1, 2004, pp. 1059–1066.

[13] S. Sprenkle, E. Gibson, S. Sampath, and L. Pollock, "A case study of automatically creating test suites from web application field data," in Proceedings of the 2006 workshop on Testing, Analysis, and Verification of Web Services and Applications, 2006, pp. 1–9.

[14] K. Wang and S. J. Stolfo, "Anomalous payload-based network intrusion detection," in Recent Advances in Intrusion Detection, 2004, pp. 203–222.

[15] "MIT Lincoln Laboratory: Information Systems Technology." [Online]. Available: http://www.ll.mit.edu/mission/communications/ist/corpora/ideval/data/index.html. [Accessed: 27-Aug-2011].

[16] J. McHugh, "The 1998 Lincoln Laboratory IDS evaluation," in Recent Advances in Intrusion Detection, 2000, pp. 145–161.

[17] J. McHugh, "Testing intrusion detection systems: A critique of the 1998 and 1999 DARPA intrusion detection system evaluations as performed by Lincoln Laboratory," ACM Transactions on Information and System Security, vol. 3, no. 4, 2000, pp. 262–294.

[18] K. Ingham and H. Inoue, "Comparing anomaly detection techniques for http," in Recent Advances in Intrusion Detection, 2007, pp. 42–62.

[19] C. Krügel, T. Toth, and E. Kirda, "Service specific anomaly detection for network intrusion detection," in Proceedings of the 2002 ACM Symposium on Applied Computing, 2002, pp. 201–208.

[20] C. Kruegel and G. Vigna, "Anomaly detection of web-based attacks," in Proceedings of the 10th ACM Conference on Computer and Communications Security, 2003, pp. 251–261.

[21] N. Ye, "A markov chain model of temporal behavior for anomaly detection," in Proceedings of the 2000 IEEE Systems, Man, and Cybernetics Information Assurance and Security Workshop, 2000, vol. 166, p. 169.

[22] S. Cho and S. Cha, "SAD: web session anomaly detection based on parameter estimation," Computers & Security, vol. 23, no. 4, 2004, pp. 312–319.

[23] J. M. Estévez-Tapiador, P. García-Teodoro, and J. E. Díaz-Verdejo, "Measuring normality in HTTP traffic for anomaly-based intrusion detection," Computer Networks, vol. 45, no. 2, 2004, pp. 175–193.

[24] S. A. Hofmeyr and S. Forrest, "Immunity by design: An artificial immune system," in Proceedings of the Genetic and Evolutionary Computation Conference, 1999, vol. 2, pp. 1289–1296.

[25] F. González, D. Dasgupta, and R. Kozma, "Combining negative selection and classification techniques for anomaly detection," Computational Intelligence, vol. 1, 2002.

[26] F. A. González and D. Dasgupta, "Anomaly detection using real-valued negative selection," Genetic Programming and Evolvable Machines, vol. 4, no. 4, 2003, pp. 383–403.

[27] M. Azimpour-Kivi, R. Azmi, and S. Ghorbani-Faal, "Artificial immune systems (AIS) for classification and its application to anomaly detection," in Proceedings of The 4th Iran Data Mining Conference, IDMC 10, Iran, 2010.

An agent-based and semantic service oriented approach for service discovery in Network Centric Warfare (NCW)

Masoud Barati[1] , Mohammad Sajjad Khaksar Fasaei[2]

[1] Young Researchers Club, Kermanshah Branch, Islamic Azad University, Kermanshah, Iran

[2] Islamic Azad University - Songhor Branch
Songhor,Kermanshah, Iran

Abstract

In the new age one of the most challenging aspects of the information age is discovering the desired services with most accuracy and speed. In this paper a new semantic model is presented for the service discovering process in network centric warfare (NCW). The model is based on predefined ontology which is determined for the most important aspects of NCW in the application level. Services is based on a service oriented infrastructure and implemented as a grid which is developed as independent services. In the model, service discovery mechanism allows each of service agents to interact with its agent neighbors semantically and form a service chain for a specific task dynamically. The introduced model considered the semantic SOA which can overcome the undesirable time-delays and overhead corresponding to complexity of service discovery at heterogeneous environment such as NCW.

Keywords: *network centric warfare, brokering, service agent, ontology, semantic similarity function.*

1. Introduction

If we analyze the recent worlds, we come to this understanding that traditional war approaches cannot help us achieve the goals of the information era. The consequences of information era and the change of the notion of the power from technology toward informational concepts as well as transmission of power toward the edges of the system have caused profound changes in the management of organizations(specially in military fields). In the information era, the process of power establishment consist of getting row data, combining and processing those data for the purpose of producing information and extracting knowledge from them, and making appropriate decisions and finally reconsidering the results for correcting the future decisions[1]. Considering the nonlinear and undetermined nature of the war environment, it should be noted that absence of sufficient accuracy and speed in each of the processes can reduce the efficiency and authenticity of the final decisions to a considerable degree. Network centric war (NCW) attempts to inject the concepts of information era into the war environment. It must be regarded that the purpose of NCW is not to exploit computer networks or to utilize automatic weapons; it is intended to exploit the smallest available weapons capitals and gaining benefits from such instruments by synergetic processes to reach the possible usage in other to satisfy the organization purposes. NCW consider the war environment a distributed, heterogeneous, and interactive environment that consist of cognitive and responsive agents in which using a common understanding and a situational understanding move toward the determined purposes. This society interacts in an environment based on reliance for an acceptable level of sensation and finally convergence toward the intention of commanding[2]. According to the great importance of providing information in NCW, it should be noted that providing and distributing incorrect information could be destructive to the same extent.

One of the most important assumptions of modern war is sharing and sending information to existences of the system, without considering geographical and logical intervals. So, every legal existence should be able to access necessary information in everywhere in the network in any given time. As transparency, agility, weak connections, high accessibility and conformity are of the features of environments related to the modern wars, it is visible that architectures and common ancient approaches are not to satisfy such needs[3].

For increasing the capacity of flexibility and defining pre-conventions before hand service description and their relationship will be expressed by added semantic to the under analysis environment. The emphasis of semantic service on ‘meaning and content’ for adding semantic to the service is for creating understandable machines. Deducing semantic cause concurring on unsuitable time delays lead from environmental complexities like NCW.

In this paper at first we propose a semantic service brokering model for NCW, then for service discovery in the service layer of the model we will introduce the semantic similarity of domain ontology and use a heuristic algorithm for finding neighboring service agents in a decentralization manner for the discovery of services, without the overhead of negotiation, a service agent only needs to know neighbor nodes he is connecting with and what the semantic descriptions of neighboring service agents are. The method of using semantic similarity provides a flexibility to allow individual service agents to perform flexible semantic matches.

Because the algorithm allows individual service agents to only query neighbor nodes without the need of global knowledge, the strategy of the algorithm can let individual service agents have a self-adapting capability to different network structures. With low diameter and low average degree, small world networks have been found important in many fields such as road maps, electric power grids, metabolite processing networks, neural networks, and social influence networks etc. In this study, the description of the proposed algorithm in finding neighbors is within a small world network structure. The algorithm can adapt its connectivity within the small world network structure. Similarly, the algorithm can also be applied to other network structures by the self adapting policy.

2. Related works

various works have been done in the realm of service brokering and recourse discovery in the service-oriented systems that some of them will be discussed here.

In the presented architecture in [4] a method suggested whose quality service parameters are analyzed by a function for discovering service suitability rate and the broking operation will be performed by this function. In suggested architecture of this article service requests will be presented to the broker and according to the suitability functions decide a service that maximize the amount of this function.

Article [5] devoted to the suggestion of a system whose broker according to the quality of service insurability dynamically monitor the network services and determine the best services for presenting to the those who request. suggested broker in this article is able to replace users connections with services when access is not achievable and provide an environment highly fault tolerance. In this article a light software broker for benefiting in the distributed environments is presented. In the presented strategy service quality of software (Qos) is centric that set on distributed service suppliers.

Authors of article[6] by classifying system services have considered a dynamic method based on service quality parameters that considered variant features of Qos in the system .on the other hand various works have been done in the realm of working with service-oriented environments considering mental models that are. two software broker of reliability and performance. reliability software broker can discover a software service dynamically that can provide the best efficacy according the effective and operational needs of clients of services. Software performance broker can supply service error tolerance for clients of the services by migration of service supplies of clients of the cervices between multiple software services.

In article [7] a service-oriented environment for distributed multimedia operations is presented those cervices composition in this environment will be performed considering predetermined anthology according to the current system capabilities.

In the presented methods meaning features for discovering services are considered .also service brokers from centralized methods to discover the service are used. we propose a flexible approach to discover services semantically by using a decentralized fashion. We employ ontology driven semantic description to describe services. The algorithm defines service agents as service carriers , allows individual agents to manage their connectivity based on local knowledge by using semantic similarity (e.g., who is connecting with him and what service the neighbor is carrying, conveyed by the semantic description?), and dynamically forms service chains to complete tasks or jobs.

3. The proposed Service Brokering Model for NCW

Our proposed service brokering model is an ontology-based five layered model as shown in Fig.3.1. As shown in the figure, system ontology has a pivot role from bottom to the top. From the NCW point of view, the application layer will be the NCW platform through which NCW is presented to the user. This layer is based on the NCW ontology to understand the conceptual structure on how to generate a common operational picture. User requirements and requests are injected into the other layers through the application layer.

Goal based layer is responsible for translating the received requests into goals[8]. This process is done using the corresponding ontology from the ontology pyramid. In this process each goal may also be broken down into smaller goals for further distributions through the brokering layer.

The brokering layer receives the generated goals from the upper layer and distributes the goals (some or all of them) among the brokering services. The brokering layer may be implemented as a brokering grid involving different services each one for a particular attribute. Our proposed brokering grid takes advantage of five independent services according to five important attributes of availability, maintainability, performance, policy, and reliability[3].

The service layer contains the floating services on the overall system infrastructure. Web service is an appropriate technology for exploiting the SOA tenets, so this layer may also be called as web service layer. This layer includes all the web service containers and the web service cores, negotiating according to web service standards as WSDL, and SOAP protocols. Brokering services in the brokering grid conceptually can also be considered as a part of the service layer.

Communication layer refers to the underlying communication network. It is implemented as an IP network with varieties of protocols. The ontology pyramid also interferes the function of this layer to adjust communication protocols, routing strategies, etc.

Although the paper focuses on service discovery, the proposed model shown in Fig.3.1 can be applied for other NCW aspects as well.

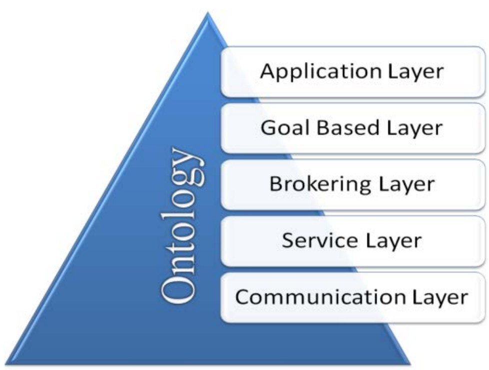


Fig.3.1 The proposed service brokering model for NCW.

4. Ontology and semantic similarity

Ontology is a formal structure that contains information about semantic description data. This structure contains a set of concepts and their connections. It can be used in the recovery of information that deals with users inquiries [9, 10]. Ontology O is defined as

$$O = \{C, \leq_C, R, \leq_R, A\},$$

Where C is a set of concepts, R is a set of relationships, $\leq_C$ is a partial order on C and $\leq_R$ is a partial order on R. Also, A is defined as a set of axioms [11, 12].

Semantic similarity function is used for computing similarity between two concepts. Similarity between concepts shows degree of their common. Similarity function is defined as sim(x, y): C*C → [0, 1]. The output of this function is a real number in interval [0, 1] that shows the degree of similarity between two concepts x and y. The zero output means lack of sharing and output one shows complete similarity between two concepts [13-17]. The semantic similarity as follows:

$$\mathrm{sim}(x,y) = \rho\frac{|\alpha(x)\cap\alpha(y)|}{|\alpha(x)|} + (1-\rho)\frac{|\alpha(x)\cap\alpha(y)|}{|\alpha(y)|} \qquad (4.1)$$

In this function $p \in [0,1]$ determines the degree of influence of generalizations. α(x) is the set of nodes upwards reachable from x, and α(x) ∩ α(y) as the reachable nodes shared by x and y. For more information refer to [18,19].

For example, an ontology with hierarchical structure graph, as shown in Fig. 4.1, has 5 concepts and has 'is a' relationships. we define Weapon as a root node, and which has sub nodes including Air_weapon, Sea-weapon and Ground-weapon. Air_weapon also includes sub-nodes Bomber_weapon, Helicopter_weapon and Missile_weapon.

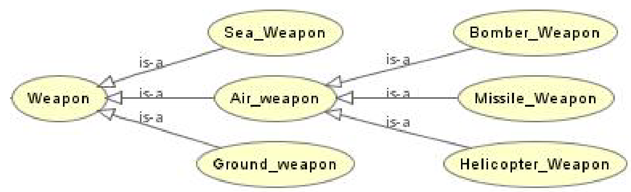


Fig.4.1 An ontology example graph of weapon

In terms of Eq. (4.1), the concepts Bomber_weapon, Helicopter_weapon have 3 reachable nodes from themselves, namely,

α(Bomber_weapon) = 3, α(Missile _weapon) = 3, α(Air_weapon) = 2, α(Helicopter_weapon) =3, α(sea_weapon) = 2, α(Bomber_weapon) ∩ α(Air_weapon) = 2, α(Bomber_weapon) ∩ α(Helicopter_weapon) =2 α(Ground-weapon) ∩ α(Sea-weapon) = 1.

5. Description model for service discovery in service layer

In this present article, NCW is modeled as a graph each of the nodes of which defines a service agent and each of the agents carries a specific service type. We assume k different service types existing in the system, which are displayed as $\{S_1,\ldots,S_K\}$. It is noteworthy that service agents have the capability to find their neighboring nodes to form service chains needed for tasks. Tasks are distributed with a certain rate μ in the network and each of them needs m different types of services ($m \leq k$). Supposing that p tasks are distributed in the network, the tasks can be displayed as a p×m matrix, in which every row represents a task and every column the type of the required service. As stated in [20], each node (service agent) may be in one of the three states "inactive", "committed" and "active", which can represented as ST_k = 0, 1 or 2. Inactive state means the node is a free node and not working on any task. Committed state indicates that the node is working on a task but the service chain has not been completed. Active state means that the node is working on a task and the service chain has been completed. Here, we can define semantic similarity matrix as a n×n matrix whose elements represent the degree of similarity between service type needed for the task and service type of the service agent. Regarding the assumption that semantic similarity threshold is 1, the elements of semantic similarity matrix are 0 and 1, in which 1 represents total similarity of service types and 0 indicates dissimilarity of the service type required from that of the service agent. The above model corresponds to the model presented in [20].

6. Proposed algorithm

In this method at first a node will be chosen randomly. If this node be free and has one of the different types of services and also have a free neighbor capable providing another type of service in the task ,then this node will be set in the state of committed (ST=1) therefore this node send a "ready" massage to its free neighbor. until fully satisfaction services required for the task, every neighbor with receiving the message of "ready" from its neighbor make itself in the state of committed, and if having another free neighbor that can provide other types of services required for the task, it will send the "ready" message to that neighbor .in the follow, the pseudo code related to the algorithm is represented.

In this part we give an example for better understanding of the algorithm. suppose that we have a network with 8 nodes that are numbered from 1 to 8. also 4 types of services by numbers from 1 to 4 are distributed in this network. Now we suppose that a task that require four type of services {1,2,3,4}is distributed in this network. as the fig.5.2(a) shows at first the node number 5 that has the service type 2 has chosen randomly ,because in the neighbor of this node (3,7)at least there is one node that can supply another type of task, therefore this node will be set in the state of committed (nodes that are set in the state of committed are shown by red color)then the node 5 send ready message to one of its neighbors (node7) and node 7 after receiving this message set itself in the state of committed (fig.5.2(b)). regarding the services required for remained task {1,4}node number 5 to its neighbor (3) that has the service type 4 , send the ready message to this node and node number 3 will be set in the state of committed with receiving the message (fig.5.2(c)). node number 7 based on the reminded service of task {1} analyze its neighbor and because the node 2 is free and has the service type 1, thus node 7 send the node 2 the ready message. This node also with receiving this message set in the state of committed and a chain with nodes of 2, 3,7,and 5 will be formed for the task(fig.5.2(d)).

INPUT task
OUTPUT services chain for task requires
ASSUME N_K is a node that selected at first randomly
ASSUME ST_i is state of node ith and ST_i can be 0, 1 or 2
ASSUME S_i is the service type of node ith
ASSUME L is size of task requires(length of task)
ASSUME A is a threshold for similarity function
ASSUME TR_m is a service type in task requires
Begin Algorithm
Select randomly a free node (N_K)
If sim(S_k , TR_K)> A **Then**
If N_K has a free neighbor (N_j) and sim (S_j , TR_j) > A **Then**
ST_k = 1
N_K send message "ready" to N_j
End If
End If
While L > 1
If N_j receive the message "ready" from it's neighbor **Then**
ST_j = 1
L = L -1
End If
If N_j has free neighbor (N_i) and sim(S_i , TR_i)> A **Then**
N_j send message "ready" to N_i
End If
End While
End Algorithm

Fig.5.1 The pseudo code of algorithm for service discovery

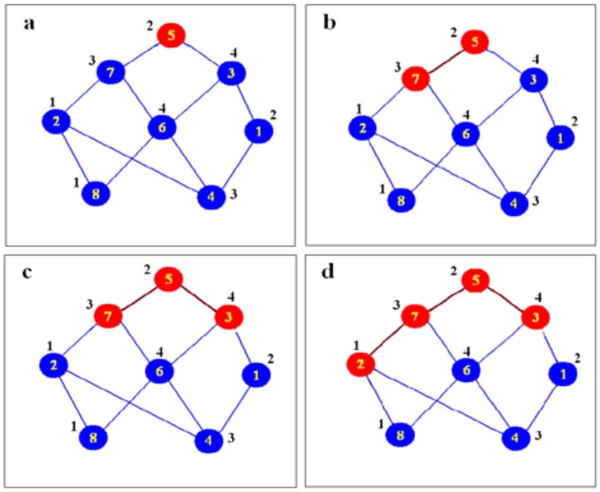


Fig.5.2 The scheme of the algorithm for forming a service chain

6. The Results of Simulation

In this section we will show the simulation results of represented model. In this simulation, first we consider a network with 64 nodes and 16 various service types that are uniformly distributed over the network. Also we assumed that each of the tasks needs 16 service types, the represented algorithm is applicable on every network structure and here we used small world network for applying our algorithm we create small world network graph with 64 nodes based on available methods in [21,22].in every experiment 10 different rate of task creation that is between intervals of [0.05,1] and load in every rate, 1000 tasks in the time step that is two times of required services(32).

We analyze the relation of task success probability under different task loading rate along with network scale with consideration of semantic similarity threshold one.

We define task success probability as the ratio of number of tasks completion and total tasks published.

We consider the task complexity as the ratio of required services for the task and the variety of services are in the network. Here we assume that task complexity is 0.5.

6.1 The relationship between task success probability and task load rates

Here we analyze the task in different lengths of 4, 8, 12, 16 and evaluate the success probability in similarity threshold of one. The results of these experiments are shown in the fig 6.1.

As you see X-axis is task loading rate and Y-axis show the success probability. From the figures we conclude that with increasing task load rate, success probability will decrease.

6.2 the relationship between success probability and increasing the number of nodes

Here we model nodes with the number of 64,128 and 256 that contains 16 various service types that are uniformly distributed and in every time step we send 1000 tasks to them.

We considered success probability of task with regard to three different network scales and suppose task complexity has the fixed amount of 0.5 .the experiment results in the figure 6.1 show that X-axis is task loading rate and Y-axis show task success probability. As observed, when network scale increase, success probability will increase.

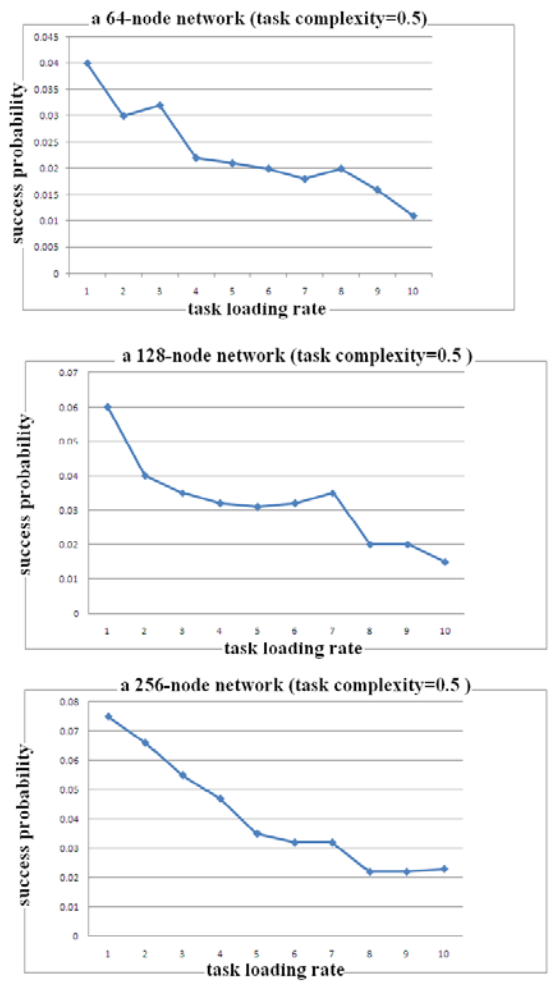


Fig.6.1 The relationship between success probability and different network scale

7. Conclusions and future work

service discovery is one of the challenges in service oriented environments. We obviated this challenge with using a perfect semantic model and a heuristic service discovery algorithm for forming a service chain for a specific task. The simulation results shown that with increasing the nodes number or decreasing task loading rates, the success probability of task completing is increasing.

The future work will focus on how to increase the security aspects in the proposed model and how to increase fault tolerance when a service in the service chain confronted with failure.

References

[1] D. Alberts, R. Hayes, Power to the Edge: Command and Control in Information Age, CCRP Publications(2003).

[2] NCW Roadmap, Australian Department of Defence, Report (2005).

[3] P. Lalbakhsh, S.K. Fasaei, M.N. Fesharaki, "An Ontology-based SOA Enabled Brokering Model for Common Operational Picture, Regarding Network Centric Warfare Services", The 4th National Conference of Iran's Scientific Society on Command, Control, Communications, Computers & Intelligence , Tehran, Iran(2010).

[4] M. P. Papazoglou, W. J. Heuvel, "Service Oriented Architecture: Approaches, Technologies, and Research Issues", The International Journal on Very Large Data Bases. 16(3)(2007),pp. 389-415.

[5] D. A. Menasce, V. Dubey, "Utility-based QoS Brokering in Service Oriented Architecture", IEEE International Conference on Web Services, USA(2007),pp. 422-430.

[6] Jiangyun Xu and Weichang Du, "Software Brokers for Quality of Services in Service-Oriented Distributed Systems", Proceedings of the Second Annual Conference on Communication Networks and Services Research (CNSR'04)(2004).

[7] R. Tabein, A. Nourollah, "Dynamic Broker-based Service Selection with QoS-driven Recurrent Counter Classes", 2008 International Conference on Service Systems and Service Managements, Australia(2008),pp. 1-6.

[8] P. Lalbakhsh, M. N. Fesharaki, "Towards Autonomic Computing using Goal-based Translation Strategy in Heterogeneous Distributed Environments", Proc. 9th WSEAS Int. Conf. Automatic Control, Modeling & Simulation, Turkey(2007) ,pp.65-71.

[9] H. Stuckenschmidt, Ontology-based information sharing in weekly structured environments, Ph.D. thesis, AI Department, Vrije University Amsterdam (2002).

[10] T.R. Gruber, Toward principles for the design of ontologies used for knowledge sharing, KSL-93-04, Knowledge Systems Laboratory, Stanford University (1993).

[11] R. Rada, H. Mili, E. Bicknell and M. Blettner, Development and application of a metric on semantic nets, IEEE Transactions on Systems, Man, and Cybernetics 19 (1989),pp. 17–30.

[12] T. Andreasen, H. Bulskov and R. Knappe, From ontology over similarity to query evaluation, in: R. Bernardi and M. Moortgat (Eds.): 2nd CoLogNET-ElsNET Symposium - Questions and Answers: Theoretical and Applied Perspectives, Amsterdam, Holland (2003),pp. 39–50.

[13] O. Resnik, Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity and natural language, Journal of Artificial Intelligence Research 11 (1999),pp. 95–130.

[14] R. Richardson, A. Smeaton and J. Murphy, Using WordNet as a knowledge base for measuring semantic similarity between words, Tech. Report Working paper CA-1294, School of Computer Applications, Dublin City University, Dublin, Ireland (1994).

[15] M.A. Rodriguez and M.J. Egenhofer, Determining semantic similarity among entity classes from different ontologies, IEEE Transactions on Knowledge and Data Engineering 15 (2003),pp. 442–456.

[16] N. Seco, T. Veale and J. Hayes, An intrinsic information content metric for semantic similarity in WordNet, Tech. report, University College Dublin, Ireland (2004).

[17] A. Tversky, Features of similarity, Psycological Review 84 (1977),pp. 327–352.

[18] M.M. Richter, Classification and learning of similarity measures, Technical Report SR-92-18 (1992).

[19] R. Knappe, H. Bulskov and T. Andreasen, Similarity graphs, LNAI 2871, in: N.Zhong, Z.W. Ras, S. Tsumoto, E. Suzuki (Eds.): 14th International Symposium on Methodologies for Intelligent Systems, ISMIS 2003, Maebashi, Japan (2003),pp, 668-672.

[20] H. Liangxiu and B. Dave, Semantic-supported and agent-based decentralized grid resource discovery, Future Gener. Comput. Syst. 24 (2008) 806-812.

[21] D.J. Watts and S.H. Strogatz, Collective dynamics of small-world networks, Nature 393 (1998) 440–442.

[22] D.J. Watts, The dynamics of Networks between order and randomness, in: Small Worlds, Princeton University Press, Princeton, New Jersey (1999).

Masoud Barati received the B.S. and M.S. degrees in computer engineering in 2008 and 2010, respectively. He is member of young researchers club in Islamic Azad University of Kermanshah,Iran. He also joined the faculty of computer engineering department at Kangavar Azad University, Kermanshah, Iran, in 2010. His research interests include service oriented architecture, grid computing, ontology, multi agent systems and wireless sensor network.

Mohammad Sajjad Khaksar Fasaei received the B.S. and M.S. degrees in computer engineering in 2008 and 2011, respectively. He joined the faculty of computer engineering department at Songhor Azad University, Kermanshah, Iran, in 2011. His research interests include service oriented architecture, network centric warfare, ontology, rule mining and ant colony.

Obstacle based Range-Free Localization-Error Estimation for WSN

S. Swapna Kumar[1], Dr M. Nanda Kumar[2], Dr V.S Sheeba[3]

[1] Department of Electronics & Communication Engineering,
Anna University Coimbatore, T.N, India.

[2] Department of Electrical Engineering,
Calicut University, Kerala, India.

[3] Department of Electronics & Communication Engineering,
Calicut University, Kerala, India.

Abstract

The projected paper considers a range-free localization protocols for the wireless sensor networks that highlights the localization irrationality problem in presence of obstruction. Here in our proposed research work the mobile anchor node is assumed to have the evidence of their position where an interaction protocol is considered. The proposed enables the presence of relay nodes with the beacon signals for the better optimization techniques by using the coordination ordinary sensor nodes and mobile anchor nodes simultaneously. The impediment consequence in the form of obstacle were studied with the transmission irregularities with the radical change algorithm for computation the localization errors in the range free localization. The significance of the current work considers the diversification of Sensor network topology. The performance analysis executed in various scenarios and our model was compared with DV-Hop, APIT, ROCRSSI, and Amorphous techniques. A simulation result in comparison with the existing system projected outperforms for minimizing localization error even in presence of obstacle.

Keywords: *Distance Vector-Hop, Localization, Mobile Anchor Node, Optimization, Range-based, Range-Free based, Wireless Sensor Networks.*

1. Introduction

Estimating the node localization is an extremely significant task in this field. Node localization consider as one of the complicated issues in wireless sensor network. Localization is consider as the fundamental services that is analogous to various operations such as cluster creation, routing, communication, coverage of network etc. With the aid of Cooperation [1], localization can be achieved with the help of sensor nodes itself without any involvement of humans. The critical issues in WSN operation is to determine the substantial locus of the sensor nodes as the position information will be deployed to find the locus at which the sensor reading originates as well as in any energy aware geographic routing. Various researches [2] towards application of sensor network also found that position estimation is information of interest.

There were many algorithms on localization which has been discussed in past to provide localization information for every node. The protocols used in localization is classified into two categories e.g. range-based and range-free method, in relation to the methodology deployed for estimation of the sensor nodes position information. The range based method is defined by protocols that deploy the absolute point-to-point range estimates or estimation for position. The range-free based method constructs no assumption about the availability or validity of such information. The range-free localization is being considered as a cost-effective alternative to range-based methods because of hardware limitation of deployment of WSN devices. Irregularity in transmission propagation as well as stringent restriction on cost of hardware has rendered localization a very challenging. The range free localization is more capable and promising to achieve higher localization accuracy without introducing any extra hardware in comparison to range-based technique of localization which depends on received signal strength to calculate absolute point-to-point distance.

Range free localization technique deploys information related to network topology as well as connectivity status for evaluating location. Low cost, no extra hardware, little communication traffic as well as flexible precision in position estimation is some of the advantageous features of range-free methods. Therefore range-free technique is considered to be most effective solution for the localization issues in wireless sensor network.

In comparison to range-based approach, the range-free techniques facilitates sensor nodes to evaluate their position without depending on parameters like distance or angles [19] [3]. Such methodology normally requires various anchor nodes, that enable position unknown sensor nodes to estimate their position by using the radio connectivity data among the nodes, or by comparing their RSS feeds with those supplied by neighbour nodes.

Various research techniques related to arrival time difference, received signal strength, arrival time etc, has been already proposed [4] which also discussed about trilateration, maximum likelihood measurements methods. DV-Hop localization, APIT, Centroid localization, amorphous positioning etc is some typical algorithms. Based on the algorithm of DV-HOP, sensor nodes estimate their position based on the anchor positions, number of hops from anchor, and also the average distance per hop [4]. Amorphous positioning algorithm uses offline hop-distance estimations, improving location through a neighbor-information exchange [5].

Many localization schemes have proposed solutions which are based on assumptions may not valid in certain scenarios. Some of such assumptions observed as symmetric radio connectivity, circular radio range, absence of any obstacle, no line-of-sight, poor multipath and flat terrain [6]. Another research loophole in this area is lack of consideration of important parameters like deployment method of WSN, presence of reference points or anchor nodes, cost of localization, energy consumptions etc. As the sensor network is normally of static nature, so obtaining location information by each sensor node is often a challenging task.

The problem of localization in wireless sensor networks has been studied and evaluated predominantly in simulators. Due to the severe hardware constraints imposed on wireless sensor nodes, real system implementations of the proposed simulated solutions have not produced encouraging results. Solutions that use the most tempting means of evaluating relative distances between sensor nodes - RF signal strength, have largely failed in practice, due to the unreliable nature and irregular pattern of the radio communication. Localization schemes that are based on the receive signal strength indicator (RSSI) have been, however, intensively studied in simulators [7]. Analysing from the above stated points about previous research results, therefore in this research journal, a framework for analyzing the mobility of mobile anchor assisted range-free algorithm for WSNs is proposed in presence of obstacles. Hence with deploying relay node, our proposed model can efficiently reduce the effects of obstacles on estimation of node localization.

Furthermore, our proposed model can compute the positions of infeasible points caused by a complex radio transmission environment that is accepted as a problem when the localization inequalities are empty for the feasible set. The rest of this paper is prepared as follows. Section 2 presents related work followed by proposed model description in Section 3. The method is discussed in Section 4. In the Section 5 highlight the algorithms deployed in this research work followed by Simulation results in Section 6 and finally Section 7 will conclude the research proposal.

2. Related Work

Tian He [8] present APIT, a novel localization algorithm that is range-free and revealed that proposed APIT scheme of ours performs excellent under an irregular radio pattern and random node placement, and this result in low communication overhead. The work is compared using the state of the art via extensive simulation; range-free localization schemes recognize the most suitable system configurations of each. In addition, the effect in the case location error of routing and tracking performance is also studied.

Huang [9] presented a complete description of standard DV-Hop and clarified some gaps in previous papers. The major source of errors in standard DV-Hop is identified and two enhancements are proposed: the anchor placement strategy and Weighted DV-Hop. With the anchor placement strategy, the research work had achieved an optimum result with less number of anchors. This will result in a cost-effective implementation. Chong Liu [10] propose a ring-overlapping, range-free approach using based on relationship of Received Signal Strength Indicator (ROCRSSI) which achieves more accurate location estimation than existing high performance Approximate Point in Triangle (APIT) method. Chong Liu [10] has performed thorough performance evaluation on two novel range-free localization methods, APIT and Ring Overlapping based on comparison of ROCRSSI. Evaluation results show that ROCRSSI outperforms APIT in terms of estimation accuracy and communication overhead under the same configuration, and it also greatly alleviates the inherent .uncertain node problem of APIT.

Chia [11] demonstrated that the range-free localization mechanism without using distance or angle information was also able to achieve fine-grained accuracy. The average location error (less than 1 meter) was also competitive to other range based approaches that typically require extra hardware for the deployed sensor node. Gideon Stupp [12] propose an estimate for the protocol based on arrangement which does not require any preliminary steps and prove that its expected accuracy converges protocol improves as the number of anchors increases. Keshtgary et. al. [13] review range-free localization methods to assess the performance of two important methods: “amorphous” and “DV-hop”.

In the proposed method we reflect some parameters like energy consumption, localization accuracy, and network overhead. In the recent papers localization methods is mostly concerted on localization accuracy where a consideration of a group of evaluation parameters, energy consuming, and network overhead in addition to the location accuracy is considered. Andrija e.t. al. [14] tackle the problem of RSS uncertainty, proposed a localization method based on fuzzy set with an improvement of the ring overlapping scheme. The deployed fuzzy set theory is to model relationship between localization regions and the RSS information available to the sensor node. The research paper has described a novel fuzzy set-based range-free localization scheme, which is termed as Fuzzy-Ring. Fuzzy-Ring requires a heterogeneous wireless sensor network composed of two sets of distributed static nodes across a planar sensing field: the position of anchors, i.e. the nodes whose locations are known, and the set of sensor nodes, whose locations are to be determined. The results obtained from simulations demonstrate that our solution improve localization accuracy in the presence of radio irregularity, but even for the case without radio irregularity. Wen-Yuan Liu [15] proposes an enhanced DV-Hop algorithm based on the selection of beacon nodes.

In this paper the proposed range-free and convex optimization approach ensures the minimum localization error. The result of simulation shows that this method can choose a better combination of beacon nodes to locate unknown nodes in the network, and can greatly improve the localization accuracy of the unknown nodes. The Table 1 highlights the comparative analysis of previous research work in localization areas carried out the wireless sensor networks.

TABLE 1: COMPARATIVE ANALYSIS OF PREVIOUS WORK

Authors	Considerations	Target of Research	Results Obtained
Tian He [8]-2003	RF Localization, Irregular radio patterns, random node placement	Effect of localization on routing & tracking [Simulation]	Good Accuracy but slight performance degradation observed
Huang [9]-2008	RB localization, Probabilistic Technique	Security issues in localization process [Simulation]	Location of node is estimated on Beacon instead of sensor, obtain less than 50% localization error, 80% coverage on very sparse network of density 4
Chong Liu [10]-2007	RF localization, RSSI, overlapping rings and their intersection, irregular radio patterns	Real time implementation using Mote Sensor [Real-Time]	Enhance performance than Tien He [5] work.
Chia [11]-2008	RF localization, anchor broadcasting its location information in its movement, Mobile sensor network	Localization for mobile sensor network [Simulation]	No dependency from Hardware, interaction between nodes, and network densities.
Gideon Stupp [12]-2005	RF-localization, random sensor distribution, Anchor broadcasting position info, positioning uncertainty	Localization improvement of basic intersection protocol [Simulation]	Less Network overhead in localization process
Keshtgary [13]-2011	Hop based RF-localization, localization accuracy, energy consumption, N/W overhead	Comparative analysis of localization w.r.t Amorphous and DV-Hop [Simulation]	Amorphous is more accurate than DV-Hop in non-uniform and high diffusion network, Amorphous consumes less energy; DV-hop has better accuracy than amorphous if Anchor nodes are increased.
Andrija [14]-2010	RSS-based RF localization, fuzzy set theory	Enhancement of Chong Liu [8] [Simulation]	Fuzzy rings perform better than ROCRSSI, no consideration of level of fuzzification.
Wen-Yuan Liu [15]-2010	Physical location relationship between beacon nodes, relative position relationship between unknown nodes and beacon triangle	Improved DV hop Algorithm [Simulation]	Improved localization accuracy of unknown nodes.
Lutful Karim e.t. al [18]-2010	RF energy efficient localization technique	Implement Range-free Energy efficient, Localization technique using Mobile Anchor Comparative analysis with Neighbour-info-based Localization System [Simulation]	RELMA is more energy aware and accurate than that of NBLS - an existing Neighbouring information based localization approach.

3. Proposed Model

The proposed model presents a unique localization framework considering the presence of obstruction in environment of wireless sensor network with random node placement and irregular radio patterns for minimizing estimates of range free localization errors. The mobile anchor nodes [19] are those nodes which are in frequent movement in the wireless network and periodically broadcast beacon message, including their current location approach.

The sensor node of different anchors, which may show various patterns [20], is considered to be disturbed by various anisotropic factors. The framework will include mobile anchor evaluation, distance evaluation as well as location estimation. One of the significance of the current work may be formulated considering diversification of Sensor network topology which was not considered in the previous research work (See Table 1). The network topology might be isotropic where the properties of the proximity measurements are identical in all direction where as Anisotropic network condition is just opposite of isotropic network. When the mobile anchor node is in listening mode from the normal sensor node based on this scenario of classification of sensor, the framework will estimate the distance of mobile anchor for each types of classification.

Assumptions made for this proposed networks is:

i. Anchor nodes are mobile in the network over complete runtime.
ii. The nodes are having omnidirectional communication range.
iii. Obstruction can be deployed in any position in the networks
iv. All sensors are deployed randomly.

3.1 Aim of the Proposed Work

To design a new approach towards localization scheme that localizes the randomly deployed sensor nodes and evaluates the performance for minimizing the localization error. Ultimately, the proposed model will analyse the location of sensors by estimating the position of a target object by evaluating the temporal difference of advent of a signal released from that target object to multiple receivers. The proposed framework consists of two different types of input. The first input will be estimation of location or hop counts resulting from mobile anchor initiating flooding and the second input will be broadcasting of first input to its entire respective neighbourhood giving the final output which is estimation of location. Nodes are evaluated with respect to the location of boundary nodes for the estimation of location error. The proposed localization scheme can also be mapped as distributed elucidation as both flooding and local broadcast is exceptional cases of restricted flooding.

4. Methodology

The proposed approach discusses about issues in localization which is based on mobile anchor nodes with diversified transmission energy in presence of any obstructions. Therefore this issue can be effectively altered to problem of solving a set of quadratic inequalities. In such previous research work [16] [17], majority of the methodology considers that set of quadratic inequalities must have solutions, which is not always feasible while majority of the consideration are in range-based localization. Another significance of the current research work is that the previous researches have not considered the analysis with existence of obstruction in wireless sensor network.

Let us assume that the network of n non-anchors and m anchors nodes are present, where for every pair of dual nodes, the framework has introduced (based on measurements) the upper bound $d^{kj}_{\max}$ and lower bound $d^{kj}_{\min}$ to the Euclidean distance between a_k and x_j, and upper bound $d^{ij}_{\max}$ and lower bound $d^{ij}_{\min}$ to the Euclidean distance between xi and xj. Then, the model of the localization problem can be defined as per equation no. 1, 2 & 3.

$$\min_{x}\{J=\sum_{k=1}^{m}\sum_{j\in N_k}e_{kj}+\sum_{i=1}^{n}\sum_{j\in N_g}e_{ij}\} \quad (1)$$

With a condition that

$$(d_{ij}^{\min})^2-e_{ij}\le\|x_i-\hat{x}_j\|^2\le(\hat{d}_{ij}^{\max})^2+e_{ij}, \quad (2)$$

$$\forall i\neq j,\qquad j\in N_i$$

$$(d_{kj}^{\min})^2-e_{kj}\le\|a_k-\hat{x}_j\|^2\le(d_{kj}^{\max})^2+e_{kj}, \quad (3)$$

$$\forall j, j,\qquad j\in N_k$$

Where $e_{kj}\geq 0$ and $e_{ij}\geq 0$ represents localization errors in sensor position estimations, ˆxi and ˆxj are estimated positions of nodes i and j, respectively, and Ni, N_k are groups of neighboring nodes. Let X = [x1, x2 . . . xn] be the 2 × n matrix that needs to be evaluated where the issue of localization may be transformed and formulated in matrix form.

5. Algorithm Description

The proposed framework assumes that the sensor node can identify the boundary node as only such nodes can relay beacons from mobile anchor nodes based on boundary location algorithm [18]. Using distributed contention process, the boundary node will tend to retrieve the coordinates of the position of the node. Node density as well as transmission energy will play a vital role in localization.

Node density as well as transmission energy will play a vital role in localization. Boundary node sets a stay timer after receiving a beacon from mobile anchor node, which defines the temporal factor where the node must stay before retransmitting the current position co-ordinates. The stay timer T_S will be estimated by following equation no 4:

$$T_s = \{C_1\left(\frac{consumed\ energy}{initial\ energy}\right) + \left(\frac{C_2}{neighbor\ node\ density}\right) x\ maximum\ holdup\ period\} \quad (4)$$

C_1 and C_2 are coefficients which provide diverse cost for different parameters. The specific values of C_1 and C_2 can be configured based on which characteristics are more significant for users: power equilibrium or coverage effectiveness. In total, $C_1 + C_2 = 1$ which shows that greater remaining transmission energy and a neighbor node density will result in a shorter stay time. The candidate boundary node will broadcast beacon signals in case they don't receive any beacon signals from the other sensor nodes during its stay timer. Not only this, the contention will be terminated by the other boundary nodes in case they hear retransmitted of the beacon. Therefore, node with highest priority will retransmit first and serve as a relay for mobile anchor node's beacon signal. This technique ensures that guaranteed delivery of mobile anchor node's location co-ordinates to certain areas that cannot directly receive mobile anchor communication. So, the unidentified-location sensor node in these specific areas can obtain a set of inequality constraints on x:

$$r_i < \| x - a_i \| \leq R_i + R_{relay}, \text{ i=1, 2, 3. . . n} \quad (5)$$

Where, R-relay is the current transmission radius for the relay node as per equation no 5. Depending upon this methodology, the application can efficiently decrease the impact of the obstruction on node localization and thereby enhance the location accuracy and minimize errors in its estimation. The algorithm can be given as:

5.1 Pseudo code for finding intersection points

```
for (every node)
        initialize creation of Sensor Network topology (ST);
        declare total nodes, length, width;
        simulation area (SA) = length x width;
for (every ST)
        initialize Transmission radii (RT) and Transmission Area (AT);
        if (RT<SA)
                RT = 2 x AT;
 while (inside AT)
        plot Sensor Node;
Declare obstacle size;
        initialize obstruction position coordinates [Obx and Oby];
                plot network topology (ST`);
for (every ST`)
        estimate Mobile anchor path;
        for (every mobile anchor path)
                compute one-by-one mobile anchor position;
        while (inside AT)
                plot dynamic readings for positions captured by interface
        plot network trajectory set of each node which receives beacon from mobile anchor node.
                dx,y=cal_dist[X, Y] for both anchor and normal node.
 if (dx,y<AT && d>RT)
        plot intersection point
```

5.2 Pseudo code for Estimating Localization Error

```
if (dx,y<AT && d>RT)
dx,y= sqrt {(x-aix(i))^2+(y-aiy(i))^2}
        sum= (dx,y- r)^2+(dx-R)^2
        sum= sqrt {sum}
for (every ST)
        for (every ST`)
                while (inside AT)
                        x1= newBoundaryNode.x;
                        y1=newBoundaryNode.y;
        plot nodes which receive beacon from anchor nodes;
                relay the co-ordinates from the neighbor nodes;
        identify obstacle and estimate location error LE=1/N {Actual Position-estimated position} x 1/RT;
        compute Error for localization;
```

6. Performance Evaluation

In this section, we evaluate the proposed model in terms of localization accuracy and the error estimate in the presence of obstacle and without obstacle considering neighborhood size and compared with the DV-Hop [9], APIT [8], ROCRSSI [10], and Amorphous techniques [13]. We first present the simulation model along with the performance result as follow.

6.1 Simulation Model and Performance Measures

In this section the proposed framework simulation has been carried on Intel Pentium Dual Core E2160 CPU with 1.8 GHz and 2 GB RAM. The architecture is designed in Matlab 7.2 considering, where 100 sensor nodes will be distributed randomly in a two dimensional simulation area of size 100 m x 100 m with specific transmission radii (R) and transmission area (r) deployed for mobile anchor nodes. We evaluated the performance of the proposed framework with many empirical tests in various scenarios by changing the sensing transmission range radius and obstacle orientation. The simulation is considered both with obstacle as well as without obstacle.

- Network initialization size:-
- Node dimension size 100 X 100 m
- Node size: 100 X 100;
- Obstacle size: 20 to 90;
- R= 2 X r; where r is the maximum transmission radius

$$\text{Error} = \frac{1}{N}\sum_{i=1}^{N} ||x_i\text{-}\bar{x}_j||*1/r \tag{6}$$

Figure (1) and Figure (2) shows the simulation results in both scenarios of without and with obstacle respectively.

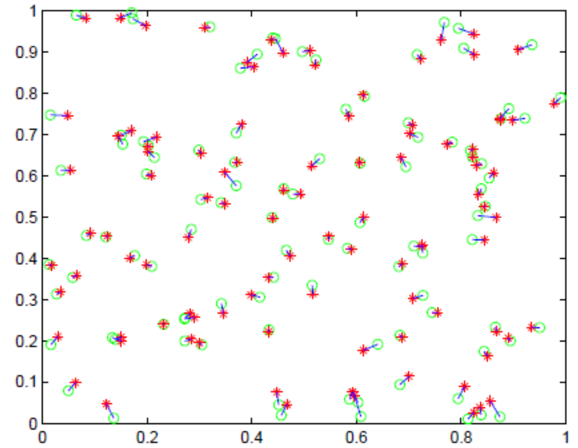


Fig 1. Simulation Results considering without Obstruction

The actual node is shown by circle and the estimated node is shown by asterisk. The line joining the actual and the real node position represent the estimation error as per equation no 6.

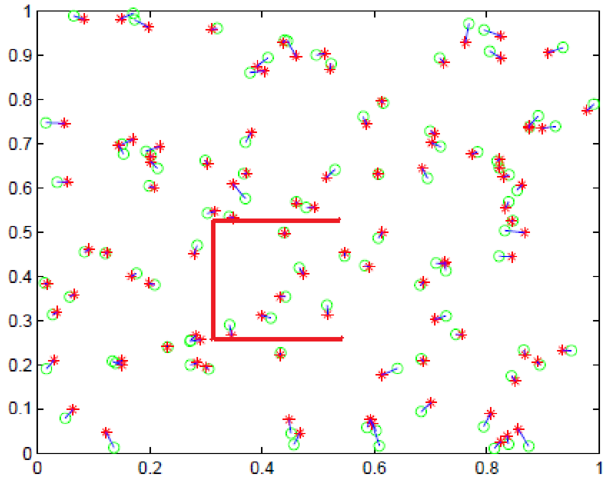


Fig 2. Simulation Results considering in presence of Obstruction

6.2 Simulation Results

In this section simulation results are presented and analyzed. The simulation was then analyzed for both mobile anchor nodes as well as normal nodes. The motive behind this is to evaluate the correctness of our approach of deployment of nodes in wireless sensor network where localization accuracy strongly depends on location of the node in the network. The error percentage indicates the percentage of maximum transmission radius.

TABLE 2: COMPUTING TIME AND LOCALIZATION ERROR—VARIOUS DEPLOYMENT OF ANCHOR NODES

Mode	Average Localization Error [%]	Average Computation Time [s]
Even distribution	2.76	3.04
Random distribution	14.996	5.26

The above Table (2) represents average localization error estimation (%) along with reading of Average Computation time (s) for various passes of simulations. It can be easily observed that the results are quite varying for which it will not be recommended to deploy distance based localization methods to networks with randomly distributed mode of deployment of mobile anchor and normal nodes. Estimating along with computation time will be required as in case of very large scale of sensor network area, computation time can restrict the application for centralized localization techniques.

The performance of the designed application is evaluated using different numbers of mobile anchor nodes which gives various result sets for location error.

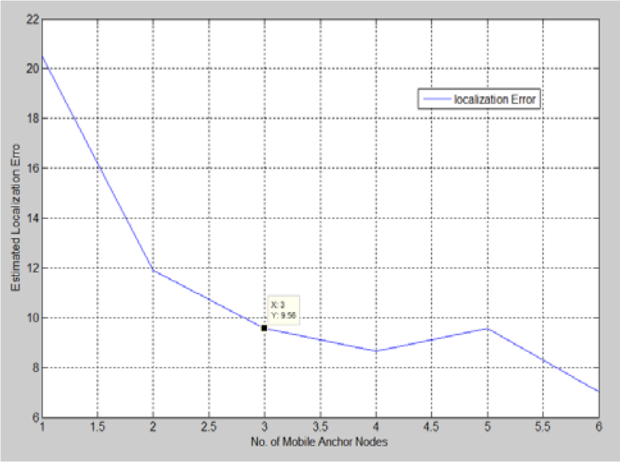


Fig. 3. Impact of the network size on localization accuracy

This evaluation as shown in Figure (3) shows that the when the quantity of mobile anchor nodes are increased, the feasibility of location errors decreases. This represents that increase in mobile anchor node-density will make the localization more accurate.

The localization accuracy is analyzed with varying the values of transmission range to evaluate impact of transmission value on estimating localization errors on range free based techniques.

The variation of the estimated localization error is shown in Figure (4) when the transmission range increases with various obstacles or noises.

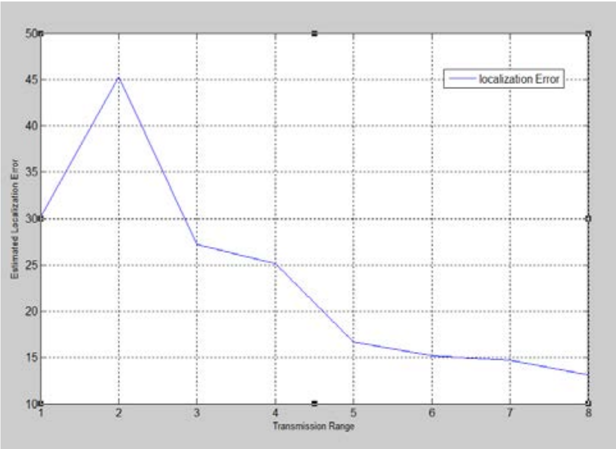


Fig. 4. Impact of the radio range on localization accuracy

The higher transmission level benefits will reduce beyond a specific point. The graphs explain that localization error estimation is minimized or to some extent render it uniform with the increase of transmission range.

The next phase of evaluation is done by considering magnitude of nodes in presence of obstructions which affects the accuracy of localization techniques as shown in Figure (5).

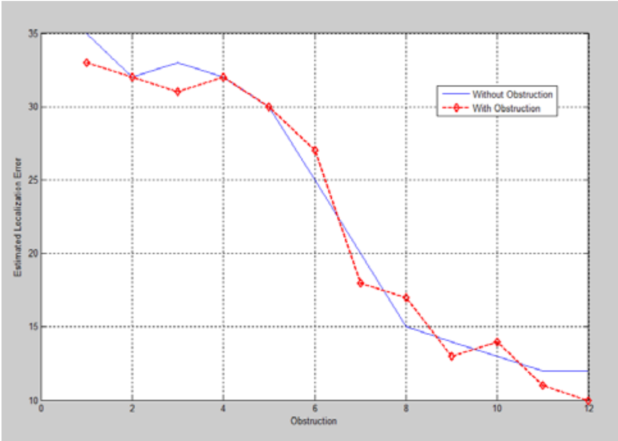


Fig. 5 Impact of Obstacles on localization accuracy

The next performance analysis is done by comparing the proposed technique with previous DV-Hop [9], APIT [8], ROCRSSI [10], and Amorphous techniques [13].

The Figure (6) represents the performance of the proposed technique with the previous research technique of localization. For this evaluation, the APIT scheme is modified for which each mobile anchor and node requires broadcasting once to see that APIT performance shows higher peak.

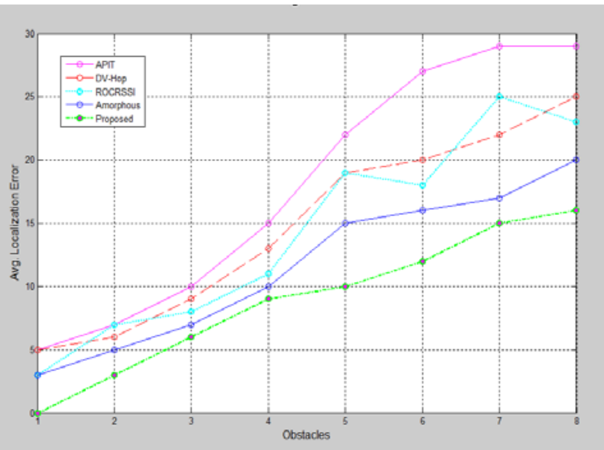


Fig 6. Average Localization Error Vs Obstacle

When the same evaluation is repeated with DV-Hop, ROCRSSI, and Amorphous, it can be observed that performance of Amorphous is comparatively better for estimating localization in presence of obstruction. In order to convert hop counts to Euclidean's distance; the program will compute average distance per hop considering range free techniques. Mobile anchor conduct the location and hop count information inside the network topology. So it can easily seen that performance for reducing localization errors is relatively reduced in our proposed system in comparison to tradition range free localization techniques.

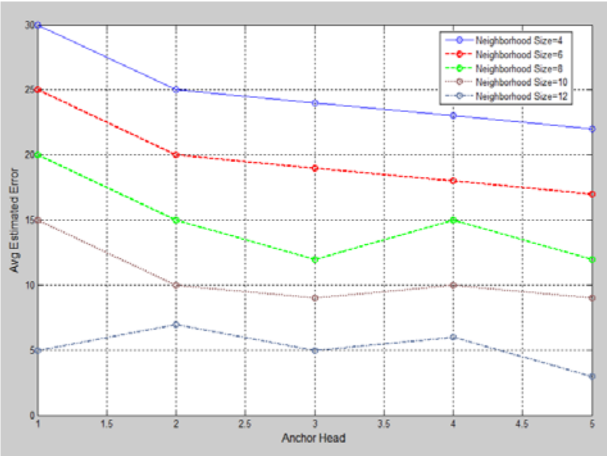


Fig 7. Average Estimated Error Vs Anchor Head

The final performance analysis we shown in the Figure (7) an idea about the consequence of neighbor nodes in estimating localization under various mobile anchor head. In this operation, once the mobile anchors quantity is computed, the hop displacement will be estimated using local averaging.

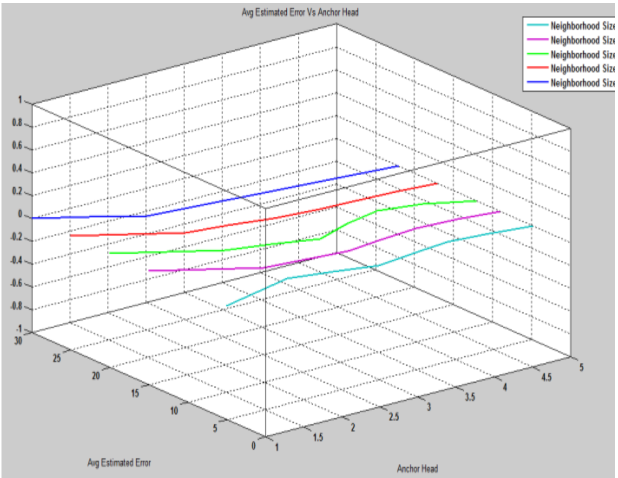


Fig 8. 3D Plot of Average Estimated Error Vs Anchor Head

From the 3D plot of the Figure (8) represents the 3D plot of average estimation for error is increased when the numbers of neighborhood increases. But it can be stated that mobile anchor nodes should not be incremented exponentially or else it will increase the network overheads much for seeking much better localization estimates result. Our result is better than the average localization error computation method.

The idea behind this localization performance evaluation is to find the best possibilities range free localization technique. The existing range free schemer fails in anisotropic wireless sensor networks. By consideration of geometrical feature of the network with the neighboring nodes hop the estimated position could able to compute which can minimize the localization error. The aim of this localization methodology is to create novel energy efficient architecture for the future wireless sensor network that can include the optimum performance factor with respect to proposed localization method and the other research characteristics that is about to propose.

7. Conclusion and Future Works

In this paper, the proposed model analyzed localization errors for wireless sensor nodes in the presence as well as absence of obstacles in anisotropic networks. The model was initially simulated in sensor network environment in absence of any obstacle, where records show that estimation of exact position is very satisfactory. The efficiency of the proposed system is experimented in various scenarios of previous range free localization techniques. In this paper we compare the result with the APIT, DV-Hop, ROCRSSI, and Amorphous methods. The performance analysis shows that our proposed technique has better results for localization estimation even in less node density which is a contrast version of amorphous techniques. The interesting part of the simulation is that when the test is conducted in various pattern of obstacle orientation within the network, the proposed approach shown minimum localization error. We believe that our design will make the range free scheme more practical for large-scale WSNs. The future work will consist of modeling the same with multiple obstructions under the same simulation.

Acknowledgment

The author would like to thank the Mr. Ranjit Patro-Project Team coordinator of R&D-Samsong India and Dr. Vineet Paleri, Professor, NIT, Calicut, India for their valuable support. I appreciate the guidance taken by Mr. Ranjit Patro-Project Team coordinator of R&D-Samsung India for providing their valuable opinion during this study. Also would like to thank the anonymous and efficient reviewer for the constructive feedback and valuable input to shape this paper for the research product. This works was carried towards the academic research of Anna University Coimbatore.

References

[1] Xue Wang, Dao-wei Bi , Liang Ding, and Sheng Wang, ‘Agent Collaborative Target Localization and Classification, Wireless Sensor Networks’, Sensors 2007, 7, 1359-1386.
[2] Jennifer Yick, Biswanath Mukherjee, and Dipak Ghosal, ‘Wireless sensor network survey’, Computer Networks Volume 52, Issue 12, 22 August 2008, Pages 2292-2330.

[3] Mo Li, Yunhao Liu, 'Rendered path: range-free localization in anisotropic sensor networks with holes', Journal IEEE/ACM Transactions on Networking (TON) archive, Volume 18 Issue 1, February 2010.
[4] Avinash Srinivasan and Jie Wu, 'A Survey on Secure Localization in Wireless Sensor Networks', Florida Atlantic University, 2007.
[5] Tian He, Chengdu Huang, Brian M. Blum, John A. Stankovic, Tarek F. Abdelzaher, 'Range-free localization and its impact on large scale sensor networks', ACM Transactions on Embedded Computing Systems (TECS), Volume 4 Issue 4, November 2005.
[6] Praveen Rentala, Ravi Musunuri, Shashidhar Gandham, Udit Saxena, 'Survey on Sensor Networks', Mobile Computing, 2001.
[7] Radu Stoleru, Tian He, and John A. Stankovic. Range-Free Localization. Secure Localization and Time Synchronization for Wireless Sensor and Ad Hoc Networks, Advances in Information Security series, Vol. 30, Springer, 2007
[8] Tian He, Chengdu Huang, Brian M. Blum, John A. Stankovic, Tarek Abdelzaher, 'Range-Free Localization Schemes for Large Scale Sensor Networks', The 9th Annual International Conference on Mobile Computing and Networking (Mobicom), San Diego, CA, September 2003.
[9] Rui Huang, Gergely V. Záruba, 'Monte Carlo localization of wireless sensor networks with a single mobile beacon', Wireless Networks, Volume 15 Issue 8, November 2009.
[10] Chong Liu, Tereus Scott, Kui Wu, Daniel Hoffman, 'Range-free sensor localisation with ring overlapping based on comparison of received signal strength indicator', International Journal of Sensor Networks, Issue: Volume 2, Number 5-6 / 2007, Pages: 399 – 413.
[11] Chia-Ho Ou, 'Range-free node localization for mobile wireless sensor networks', Wireless Pervasive Computing, 2008. ISWPC 2008. 3rd International Symposium on Issue Date: 7-9 May 2008.
[12] Gideon Stupp and Moshe Sidi, 'The expected uncertainty of range-free localization protocols in sensor networks', Theoretical Computer Science, Volume 344, Issue 1, 11 November 2005, Pages 86-99.
[13] M. Keshtgary, M. Fasihy, and Z. Ronaghi, Performance Evaluation of Hop-Based Range-Free Localization Methods in Wireless Sensor Networks', International Scholarly Research Network ISRN Communications and Networking, Volume 2011, Article ID 485486, 6 pages.
[14] Andrija S. Velimirovic, Goran Lj. Djordjevic, Maja M. Velimirovic, and Milica D. Jovanovic, 'A Fuzzy Set-Based Approach to Range-Free Localization in Wireless Sensor Networks', ELEC. ENERG., vol. 23, no. 2, August 2010, 227-244.
[15] Wen-yuan Liu, En-shuang Wang, Zi-jun Chen, Lin Wang, 'An Improved DV-Hop Localization Algorithm based on The Selection of Beacon Nodes', Journal of Convergence Information Technology, Volume 5, Number 9. November 2010.
[16] V. Vivekanandan and V. W. S. Wong, "Concentric anchor-beacons localization algorithm for wireless sensor networks," IEEE Trans. Veh. Technol., vol. 56, no. 5, pp. 2733-2745, Sept. 2007.
[17] B. Xiao, H. Chen, and S. Zhou, 'Distributed localization using a moving beacon in wireless sensor networks', IEEE Trans. Parallel Distributed System., vol. 19, no. 5, pp. 587-600, May 2008.
[18] D. Dong, Y. Liu and X. Liao. Fine –grained boundary recognition in wireless ad-hoc and sensor networks by topological methods. In MobiHoc, pages 135-144, 2009.
[19] Amitangshu Pal, 'Localization Algorithms in Wireless Sensor Networks: Current Approaches and Future Challenges', Network Protocols and Algorithms, ISSN 1943-3581, Vol. 2, No. 1, pp. 45-74, 2010.
[20] Liu Y, Yang Z, Wang X et al., 'Location, localization, and localizability'. Journal of Computer Science and Technology 25(2): 274-297, Mar. 2010
[21] Garth V. Crosby et. al., 'Location-aware, Trust-based Detection and Isolation of Compromised Nodes in Wireless Sensor Networks', International Journal of Network Security, Vol.12, No.2, PP.107 {117, Mar. 2011
[22] Kyungmi Kim et. al., A Self Localization Scheme with Relay Nodes for Mobile Wireless Sensor Networks Journal of Next Generation Information Technology, volume 2, Number 1, February, 2011

PROFILE

Mr. S. Swapna Kumar is a Professor working in the department for Electronics & Communication Engineering branch in AXIS College of Engineering & Technology. He is doing his research study in Anna University Coimbatore in the area of wireless sensor network. He attended various workshops and presented paper in the international conference. He is the member of IEEE.

Dr. M. Nanda Kumar is an Associate Professor in Electrical Engineering, in Government Engineering College, India. He completed his doctorate degree from IIT Madras. He is having several years of teaching experience and several publications.

Dr. Sheeba V.S is a Professor in Electronics & Communication Engineering, in Government Engineering College. She completed her doctorate degree from NIT Calicut. She is having several years of teaching experience. She published papers in national & international journals.

An auto detection system for Electrocardiogram of Foetal heart Issues

Mahmoud Ahmed Suliman Ali [1] and XiaoPing Zeng [2]

[1] **Department of Communication, College of Communications Engineering, Chongqing University, Postal Code 400044, 174shazheng St. Shapingba District, Chongqing, P.R. China**

[2] **Department of Communication, College of Communications Engineering, Chongqing University, Postal Code 400044, 174shazheng St. Shapingba District, Chongqing, P.R. China**

Abstract

Knowledge of the foetal heart signal prevents Foetal problems in the earlier stage. Recently, there has been a growing interest study on extraction of FECG using noninvasive method rather than the old invasive method which was more risky for the mother's health. The problems of extraction of the Foetal signals are the problems that plagued researchers in the field of signal processing. The Objective of this paper is to develop a technique for auto extracting FECG signals based on adaptive filter and simple Genetic algorithm. Practical method for extraction using computer simulations is proposed. A program for carrying out the calculations was developed in MATLAB. The testing of the algorithms was done by using real data from SISTA/DAISY and Physionet. The proposed technique for extraction of FECG was useful and the results appear to agree with the mean values of FECG.

Keywords: *FECG, AECG), TECG, surface potentials, Genetic Algorithm (GA), processing communities, and cutoff frequency.*

1. Introduction

Foetal Electrocardiogram (FECG) [4] extraction is an interesting but a difficult problem in the field of biomedical signal processing. It's a technique for obtaining important information about the condition of the foetal during pregnancy by measuring the electrical signals generated by the foetal heart as measured from multi-channel electrodes placed on the mother's body surface.

Perhaps the question that comes to mind is, "what is the importance of studying the FECG?". This is because the heart is the most important organ in an animal's the body and if it stops for a few minutes, it may lose its life. Thus it is imperative to detect the importance of heart problems before it's too late. Heart disease is classified as the most dangerous to human life, of all diseases in the world [2]. The importance of understanding cardiac electrophysiology is basic to all clinicians whether in postnatal (pediatric or adult) or prenatal (foetal) medicine. The development one in the latter in terms of screening and monitoring purposes is not only an act of preventive medicine, but also to allow further progress in understanding the foetus as a patient [5].

The problem is how to diagnoses the foetal life in the mother's abdomen The most common symptom of foetal death in most of the cases is the decrease in foetal movement which is only diagnosed by examining the absence of cardiac activity in foetal's heart.

In clinics today, two techniques are generally being used to detect the foetal heart beat; these include Ultrasound and Foetal Electrocardiography (FECG). Ultrasound provides only the images but does not confirm whether the foetus is alive or dead, especially if the foetus falls asleep for long periods [3]. Furthermore, ultrasound techniques require a trained technician/physician that can frequently reposition the transducer. Therefore, it cannot be done in a home environment which could be beneficial for problematic pregnancies. Foetal Electrocardiography (FECG) can be an attractive candidate to measure heart conduction signals by means of ordinary electrodes placed on the mother's abdomen.

2. Recording of the foetal ECG

There are two methods used to of record Foetal ECG (FECG). The first one relies on placing an electrode in direct contact with the scalp of the foetal. This is named an invasive technique that can only be used during labor. The second one is non-invasive which involves attaching electrodes on the maternal abdomen. The signals recorded by invasive methods have better quality as compared with noninvasive methods; but the procedure is rather inconvenient and it's limited to recordings during labor [18]. For this reasons the procedure for obtaining the

FECG should be non-invasive. The foetal heart is small so the electrical current it generates is very low. In order to record the FECG, electrodes are placed on the maternal abdomen as close as possible to the foetal heart. The FECG may be acquired by placing a number of electrodes around the general area of the foetal and hoping that at least one of the electrodes will have the FECG with high enough SNR. Beside the problem of electrode placement, noise from electromyography activity affects the signal due to the foetal low voltage signal. Another interfering signal is the maternal ECG (MECG) which can be 5-1000 times higher in its intensity and ability to induce surface potentials [1]. The MECG affects all the electrodes placed on the chest (thoracic electrodes) and also affects the abdominal electrodes, while FECG affects only the abdominal electrodes.

2.1. Foetal signals

The Foetal Electrocardiogram (FECG) is a time-varying signal reflecting the ionic current flow which causes the cardiac fibers to contract and subsequently relax [14]. The surface FECG is obtained by recording the potential difference between two electrodes placed on the surface of the skin [12]. The standard FECG signal consists of six peak signals each defined with a different letter, the P, Q, R, S, T and U peaks. Where the P peak results from the depolarization of the atrial, the P-R interval is the time between the depolarization of the atria and the depolarization of the ventricles. The QRS-complex results from the depolarization of the ventricles, The T wave displays the depolarization of the ventricles and the U wave is usually not present or not important resulting from a rest potential. The origin of the U wave is not clear but it probably represents “after depolarization’s” in the ventricles [11]. The FECG may be divided into the following sections.

P-wave: A small low-voltage deflection away from the Baseline caused by the depolarization of the atria prior to atrial contraction as the activation (depolarization) wave front propagates from the SA node through the atria.

PQ-interval: The time between the beginning of atria depolarization and the beginning of ventricular depolarization.

QRS-complex: The largest-amplitude portion of the FECG caused by currents generated when the ventricles depolarize prior to their contraction. Although atrial repolarization occurs before ventricular depolarization, the latter waveform (i.e., the QRS-complex) is of much greater amplitude and atrial repolarization is therefore not seen on the FECG.

QT-interval: The time between the onset of ventricular depolarization and the end of ventricular repolarization. Clinical studies have demonstrated that the QT-interval increases linearly as the RR-interval increases [14]. Prolonged QT-interval may be associated with delayed ventricular repolarization which may cause ventricular tachyarrhythmia’s leading to sudden cardiac death.

ST-interval: The time between the end of S-wave and the beginning of T-wave. Significantly elevated or depressed amplitudes away from the baseline are often associated with cardiac illness.

T-wave: Ventricular repolarization, whereby the cardiac muscle is prepared for the next cycle of the ECG.

The signals shape in FECG are the same as at maternal ECG (MECG), but completely different in values. Table 1 shows the variation of the ECG signals value between the maternal and foetal. Figure (1) shows the standard P, Q, R, S, T and U complex signals, the entire non-invasive FECG signal should be processed to be like this shape and around the values shown in Table 1.

Also can compare the maternal normal heart beat rate, during pregnancy the heart rate goes around 80-90 beats per min when was the normal adult heart beat bout 72 rate and the mean foetal heart beat rate along gestational period varying from 120 to 160 beats per min. There is a variation in heart beat rate because more QRS signal in FECG than QRS signal in MECG. As mentioned above a human heart can be described as an electrical dipole, unfortunately till now no finding information on the exact voltage value coming from the heart and, may any human body has a unique voltage value.

Table 1: Amplitude-time relations in maternal and foetal electrocardiography signal calculated as mean values from 20 traces recorded between week 36 and 41 of gestation (620 averaged QRS complexes from maternal and 760 from foetal electrocardiogram were analyzed)[16].

	QRS amplitude (µv)	***QRS Width (ms)***
Mother	150	100
Foetal	30	50

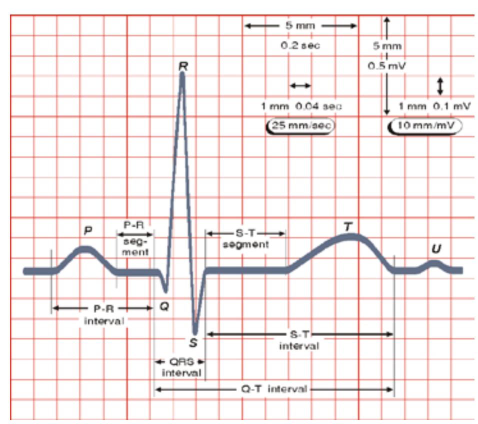


Fig. 1: Standard P, Q, R, S, T and U wave form of human heart; adopted from [1].

3. MATERIALS AND METHOD

3.1. Model Extraction

In this study the signal picked up by the Thoracic Electrodes (TECG) is represented by $X_T(t)$ and the one taken by Abdominal Electrodes (AECG) is represented by $X_A(t)$. These two signal models can be written as:

$$x_A(t) = M_a(t) + F_a(t) + N_a(t) + \eta_a(t) \quad (1)$$

$$x_T(t) = M_b(t) + N_b(t) + \eta_b(t) \quad (2)$$

Where:

$M_a(t)$ and $M_b(t)$ are Pure maternal ECG,

F_a (t) is the pure foetal ECG

η_a (t) $=\eta_b$ are the low-rank or structured noise representing other biological sources that contaminate the ECG,

N_a (t) and N_b (t) are the full-rank observation noise that always exists in physiological measurements.

The foetal ECG is very weak compared to the maternal ECG signal and noise level, so it is shownonly in Eq.(1) and neglected in Eq.(2).

In these equations η_a (t) and η_b can be eliminated by a high pass filter with a cutoff frequency of 2Hz. Therefore Eq. (1) and Eq. (2) can be simplified to:

$$x_A(t) = M_a(t) + F_a(t) + N_a(t) \quad (3)$$

$$x_T(t) = M_b(t) + N_b(t) \quad (4)$$

To extract a pure F_a (t), Eq. (4) and Eq (3) can be used as follows:

Since M_a (t) $\neq M_b(t)$ and $N_a(t) \neq N_b(t)$, use a factor K to equalize (4) to have

$$Kx_T(t) = KM_b(t) + KN_b(t) \quad (5)$$

Then subtract Eq. (5) from Eq. (3) to get

$$[x_A(t) - Kx_T(t)] = [M_a(t) - KM_b(t)] + F_a(t) + [N_a(t) - KN_b(t)] \quad (6)$$

Thus, the pure FECG can be obtained from the following equation:

$$F_a(t) = [x_A(t) - Kx_T(t)] - [M_a(t) - KM_b(t)] - [N_a(t) - KN_b(t)] \quad (7)$$

After cancelling the maternal signal and noise we get:

$$F_a(t) = [x_A(t) - Kx_T(t)] \quad (8)$$

The output of this equation is used as input to simple Genetic Algorithms (GA) to remove any undesired signals.

3.2. Genetic Algorithm and Adaptive Filtering Model

Fig.2 shows the form of the technique used to extract the desired FECG signal. The figure contains the steps of simple Genetic Algorithms (GAs); The GA used here is adaptive heuristic search algorithm premised on the evolutionary ideas of natural selection and genetic.

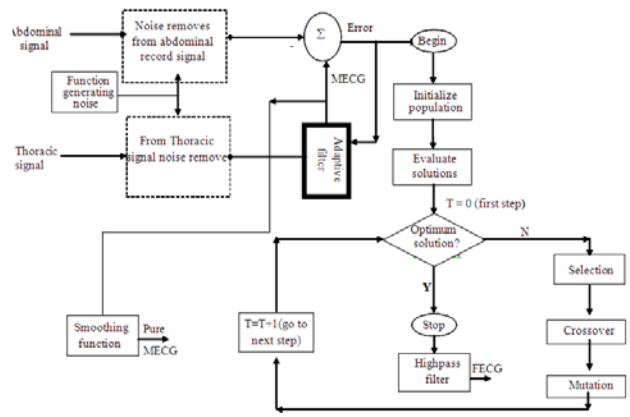


Fig 1: model for auto extraction FECG

3.3. Analysis of Maternal Abdominal Signals:

It is believed that the signals acquired from the wall of abdomen motherland (Non-invasive record) is in fact FECG signal and MECG signal with several overlapping noises .The analysis of these signals is shown in Fig.3 where the amplitude and frequency range of foetal ECG have been compared with other noises. The labels in this figure is as follows: (mECG) stand for the maternal electrocardiogram, (mEEG) for electroencephalogram, (mEHG) for electrohystrogram, (mEOG) for electrooculogram, (mEMG) for electromyogram, (mEHG) for electrohystrogram, and (fECG) stand for the foetal ECG

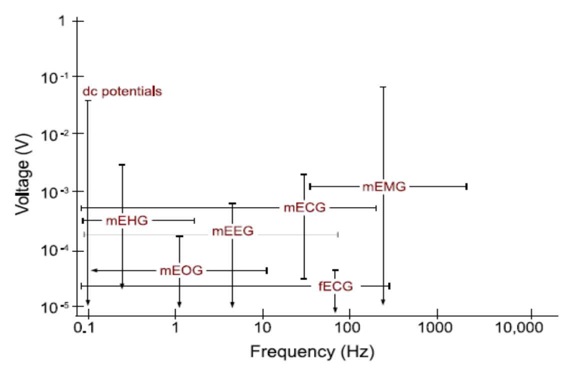


Fig.3: The amplitude and frequency range of biosignals that can interfere with foetal cardiac signals [19].

3.4. Adaptive filtering

An adaptive filter is a self-adjusting filter [8]. Its transfer function according to an optimization algorithm is driven by an error signal. The purpose of the adaptive filter in noise cancellation is to remove the noise from a signal adaptively to improve the signal to noise ratio. The usage of adaptive filters is one of the most popular proposed solutions to reduce the signal corruption caused by predictable and unpredictable noise[17]. Adaptive filters are required for some applications because some parameters of the desired processing operation are not known in advance. Adaptive filters have been successfully applied in diverse fields such as communications, radar, seismology, and biomedical engineering. Figure. (4) shows the diagram of a typical Adaptive Noise Cancellation (ANC) System used for removing noise from human ECG. This paper uses the same idea to extract the signal of foetal heart in a better modified way as contained in Figure.(2).

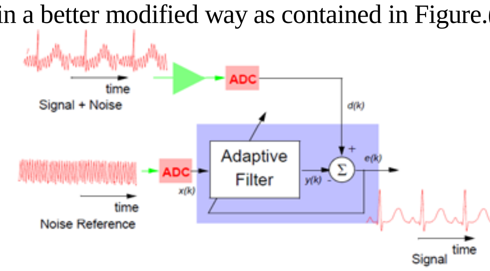


Fig .4: Adaptive Noise Cancellation (ANC)[21].

3.5. Genetic algorithm(GA)

GA is a powerful & broadly applicable stochastic research techniques that is practically being used to solve optimization problems on the basis of natural genetics. Genetic algorithms, are the most widely known type of evolutionary computation method today [7]. It is also applicable in problems where traditional estimation and optimization methods are not appropriate [13]. Genetic algorithms generally start with a population of randomly generated design vectors, test the fitness of those vectors, select the best ones, and recombine the parameter values (i.e., exchange some elements) of the best designs. Recently, the genetic algorithms technique was applied to biomedical engineering especially in foetal electrocardiogram signal [9]. The architecture Figer.(2) used for this paper is a combination of an adaptive filter and genetic algorithm (GA), where the GA is recruited whenever the first step adaptive filter is suspected of reaching local minima. The second step is an independent GA search without the adaptive filter. The process of exchanging elements among successful designs of GA also has a biological analog, which is referred to as "crossover and mutation." In this paper the case is Scheduling problem, so the method used here is string crossover. During crossover step of the algorithm, segments are cut and spliced between strings. The general framework and basic step of GA can be viewed as the flowchart at the right side of Figer.(2).

3.6. Frequency Response &Cutoff Frequency

It is essential that the ambient noise should be kept as low as possible and this is carried out with the help of an active low pass filter having a cutoff frequency of 70 Hz, due to the fact that the foetal heart beats lies in the frequency range of 20 to 70 Hz [20].
The filter used to compute the frequency response is using discrete Fourier transforms (DFT) with transfer function. This can be written as:

$$H(z) = \frac{b(1) + b(2)z^{(-1)} + \cdots + b(M+1)z^{(-m)}}{a(1) + a(2)z^{(-1)} + \cdots + a(N+1)z^{(-n)}} \qquad (9)$$

Where
B; is coefficient vector of numerator polynomial
a; is coefficient vector of denominator polynomial
After optimizing the foetus signal, a high pass filter with normalized digital cutoff frequency, $\Omega c=\omega c\ T/\pi$ equal to 0.1 H(z) and number of poles equal 3is designed,

3.7. FECG DATABASE

To test the algorithm, public databases widely used by the signal processing community known as SISTA/DAISY dataset [10] is used. It consists of a single dataset of cutaneous potential recording of a pregnant woman. A total of 8 channels (4 abdominal and 3 thoracic) are available, sampled at 500 Hz , lasting 10 seconds and The lengths of the data were 2500 point. The heart rate of the mother's signal is approximately 84 beats per minute and the Foetus corresponding to a heart rate of 132 beats per minute.
The locations of leads on maternal body for an 8-channel maternal ECG acquisition system are shown in Figure.(5). Maternal thorax ECG (TECG) signals are sampled from thorax leads while maternal abdominal ECG (AECG) signals are obtained from abdominal leads.

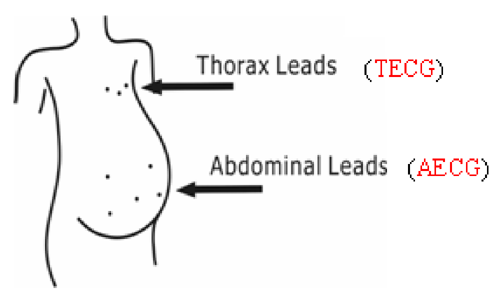
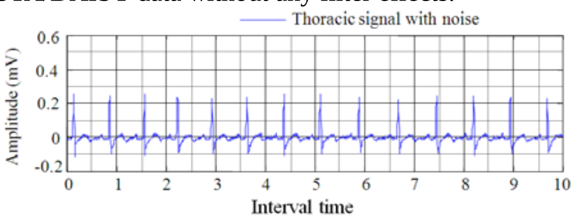


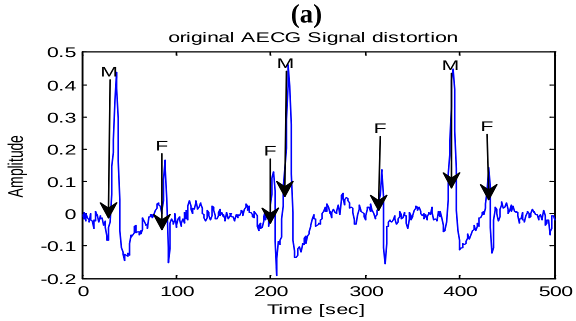
Fig.5: Positions electrodes lead on the body of mother.These methods can be practically applied from the 11th week of gestation, but the signal quality depends on recording period between 26 and 41st week of gestation.

3.8. Original signal distortion

Figer.(6) shows in section (a) real thoracic signal (TECG) of maternal which contain MECG + noise and shows in section (b) real sample of abdominal signal (AECG) which contains MECG + FECG + noise. Where M denotes maternal QRS amplitude region and F denotes foetal QRS amplitude region. These graphs show direct plot using SISTA/DAISY data without any filter effects.



(a)



(b)

Fig. 6 original graph for (a) TECG, (b) AECG

RESULTS AND DISCUSSION

The proposed scheme shown in Fig.2 is modeled under MATLAB domain, thereafter, the proposed model is trained with the SISTA/DAISY dataset. A combined signal shown in the upper part of Figer.(9) is then applied to the model and successively being extracted as shown in the lower part of Figer.(9). Thus the maternal ECG signal interference was canceled from the foetal heart ECG signal. Figer.(7) and Fig.8 represent noise that was already contaminating with signal taken from thoracic and abdomen of the mother. Also the noise figures show the evidence of an effective noise removal tool system.

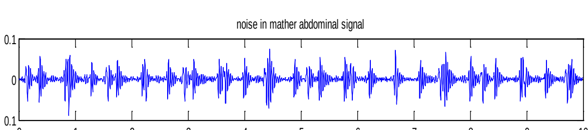


Fig. 7: Noise in abdominal signal

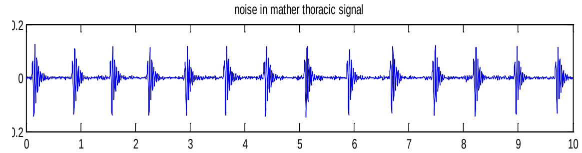


Fig .8: Noise in thoracic signal

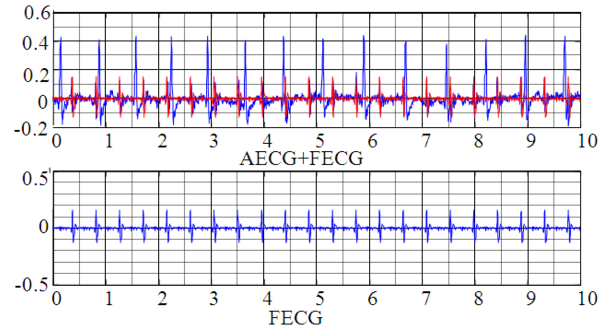


Fig. 9:A graph for MECG and FECG recombined after extracted, B graph for FECG

The results shown in Figer.(9) contain two graphs .the upper graph shows AECG and FECG. The amplitude of QRS of FECG is about 30 microvolt while that of MECG is about 150 microvolt. These values are almost in agreement to the values of QRS amplitude appeared in Table.1. The obtained results show the effectiveness of the proposed algorithm.

CONCLUSION

In this study we presented Hybrid adaptive filtering with simple genetic algorithm for removing undesired signals which are difficult to be removed by normal filters. The performance and validity of the proposed algorithm have been confirmed by computer simulations and experiments in real-world ECG data. The results obtained almost agree with the standard Foetal ECG signals.

ACKNOWLEDGEMENT

This work was supported in part by the national sciences foundation of china (NO.60971016). I wish to thank Chinese scholarship council (CSC) and Chongqing University for offering me scholarship to pursue my Ph.D study.

REFERENCES

[1] Adam, D. and D. Shavit,. Complete foetal ECG morphology recording by synchronized adaptive filtration. Med. Biological Eng. Comp., 1990, 28: 287-292.

[2] Atef M. Al-Attar, Physiological Study on the Effect of Acalypha wilkesiana Leaves Extract on Streptozotocin-Induced Experimental Diabetes in Male Mice , American Medical Journal 1 (1): 51-58, 2010, ISSN 1949-0070.

[3] A.Poblano, R. H, and Carmina. A,Neurophysiologic Measurement of Continuity in the Sleep of Foetuses during the Last Week of Pregnancy and in Newborns, International

Journal of Biological Sciences,2008 ISSN 1449-2288, 4(1):23-28.

[4] Cremer, M.,. Uber die direkte ableitung der aktionstrome des menschlichen herzens vom oesophagus und ¨uber das elektrokardiogramm des Foetall. M¨unchener Medizinische Wochenschrift, 1906, 53: 811-813.

[5] C. Velay , N. S, T.I, H.C, N. Y, K..O, and Yoshitaka Kimura, Understanding congenital heart defects through abdominal foetal electrocardiography: Case reports and clinical implications, Journal of Obstetrics and Gynaecology Research, Japan, 2011. doi:10.1111/j.1447-0756.2010.01367.x.

[7] Er.Rajiv Kumar et. al. Genetic Algorithm approach to Operating system process scheduling problem, International Journal of Engineering Science and Technology, Vol. 2(9), 2010,4247-4252

[8] Farvet, A.G., Computer matched filter location of foetal R-waves. Med. Biological Eng., 1968, 6: 467-475.

[9] Kianoush.N, S.Ebadi, and S.Sanei., Fetal Electrocardiogram Signal Modelling Using Genetic Algorithm, International Workshop on Medical Measurements and Applications - MeMeA Warsaw, Poland, May 2007, 4-5.

[10] Lathauwer, L.,Database for the identification of systems: FECG data East/Sista K.U. Leuven, Belgium ,2000 [Online].Available: http://www.esat.kuleuven.ac.be/sista/daisy/

[11] Lenssen, A.M.J., From cardiac electrical activity to the ECG: A finite element model, Master thesis, Eindhoven University of Technology,2008.

[12] Lindsley, D.B., 1942. Heart and brain potentials of human foetalles in utero. Am. J. Psychology, 55: 412-416,1942

[13] Mahmud. K and M. D. Ardakani, Inclusion Identification by Inverse Application of Boundary Element Method, Genetic Algorithm and Conjugate Gradient Method, American Journal of Applied Sciences 5 (9): 1158-1166, 2008,ISSN 1546-9239.

[14] Malmivuo, J. and R. Plonsey, Bioelectromagnetism-Principles and Applications of Bioelectric and Biomagnetic Fields, Oxford University Press, New York, 1995. [online] available: http://www.bem.fi/book/index.htm.

[15] Mallika. K, M. P. and Birsen .S.M, Impact of Contrast Functions in Fast-ICA on Twin ECG Separation , IAENG International Journal of Computer Science, 2011. 38:1, IJCS_38_1_05

[16] Matonia, A., J. Jeewski, T. Kupka, Janusz and K. Horoba, 2005. Modelling of Non-invasively Recorded Maternal and Foetal Electrocardiographic Signals. Biocybernetics Biomedical Eng., 25: 27-39.

[17] Raj Kumar Thenua, SIMULATION AND PERFORMANCE ANALAYASIS OF ADAPTIVE FILTER IN NOISE CANCELLATION , International Journal of Engineering Science and Technology ,Vol. 2(9), 2010, 4373-4378

[18] Reza Sameni, M.B. Shamsollahi, C. Jutten and G.D. Cliord, A nonlinear bayesian filtering framework for ecg denoising. IEEE Trans. Biomed. Eng. 2007, 54: 2172-2185.

[19] Reza Sameni, and Gari D. Clifford , A Review of Fetal ECG Signal Processing Issues and Promising Directions, The Open Pacing, Electrophysiology & Therapy Journal, 2010, 3, 4-20

[20] V. Chourasia and A. Mittra . Passive Acoustic Signal Acquisition System for Non-Invasive Fetal Heart Sound Monitoring Applications.. The Internet Journal of Medical Technology.Volume 5,2010, Number 1.

[21] Xilinx,Adaptive Digital Filters, available : http://www.xilinx.com/univ/teaching_materials/dsp_primer/sample/lecture_notes/AdaptiveSignalProcessing.pdf.

Mahmoud Ahmed Suliman ali is a graduate from Omdurman Islamic University (sudan). He received his B.Sc (Electrical & computer Eng) in 1997 and MSc (Telecommunication and information system)in 2002 from the University of Khartoum (Sudan). He is currently pursuing the Ph.D degree in College of communication engineering –Chongqing University- China . His research areas are Biomedical signal processing.

Xiaoping Zeng is a professor, a supervisor of postgraduate and leader of Chongqing's Electronic Association. He is a member of editorial committee for many journals in china .his main research areas include modern circuit theory and application, signal processing and biomedical signal processing.

IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 2, September 2011
ISSN (Online): 1694-0814
www.IJCSI.org

Developing Spring-Roll Dielectric Elastomer Actuator System Based on Optimal Design Parameters

Medhat Awadalla[1] and Besada Anees[2]

[1] Communication, Electronics and Computers Department, Faculty of Engineering, Helwan University, Helwan, Egypt

[2] Communication, Electronics and Computers Department, Faculty of Engineering, Helwan University, Helwan, Egypt

Abstract

To develop a well designed practical spring-roll dielectric elastomer actuator system, we have to optimize its design parameters. In our pervious work, we have achieved the design parameters without taking the applied voltage required to activate the actuator into consideration because the applied voltage might change the optimal design parameters. In the paper, optimal design parameters have been achieved and decreased the applied voltage and the problems of high voltage are avoided. Furthermore, a voltage supply that is able to pump enough controlled charges to the actuator system is developed. A switched based multistage charge pump driven by a controlled low voltage switching power supply and a voltage driver is also proposed. The recently developed voltage supply is characterized by a wide range of controlled voltages. The achieved results based on the proposed approaches show that the design of the actuator is getting simpler and outperforms compared with the other related approaches.

Keywords: *Spring roll dielectric elastomer actuator, modes of failure, optimal design parameters, switching power supply, charge pump.*

1. Introduction

Dielectric elastomer actuators have been intensely studied in the recent decade. To explore some of the basic issues in the design, one particular type of actuators is studied, the spring-roll actuators [1-3]. The construction of a spring-roll actuator is sketched in Fig. 1. Two membranes of a dielectric elastomer are alternated with two electrodes. The laminate is prestretched in two directions in the plane, and then rolled around a spring [4, 5].

Two membranes of a dielectric elastomer are alternated with two electrodes. The laminate is first prestretched and then rolled around a relaxed spring. When the spring roll is subject to a voltage and an axial force, the length of the spring couples the electrical and mechanical actions.

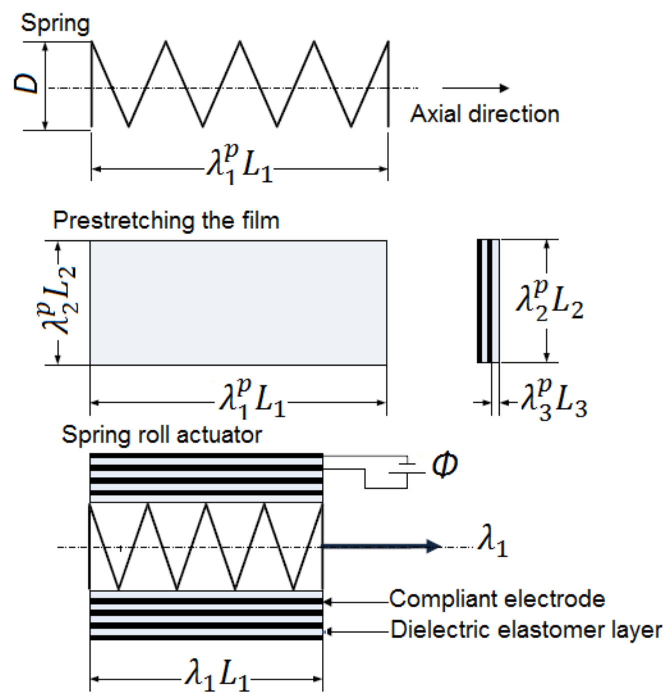


Fig. 1 The construction of a spring-roll dielectric elastomer actuator

When the actuator is subjected to an applied voltage and an applied axial force, the axial elongation couples the electrical and mechanical actions. The parameters of design include prestretches of the elastomer and the stiffness of the spring.

The continuation of the paper is as follows: Equations of state are derived in section 2. Modes of failure are illustrated in section 3. A new concept of actuation range is introduced in section 4. Optimal design parameters, design approaches and samples of actuator design are presented in section 5. Actuator system voltage supply is proposed in section 6. Section 7 gives the conclusion and future work

2. Equations of State

As shown in Fig. 1, the electrodes are compliant and bear no mechanical load. The elastomer has thickness L_3 and sides L_1 and L_2. The relaxed spring has a length $\lambda_1^p L_1$. The elastomer is prestretched to $\lambda_2^p L_2$ and $\lambda_1^p L_1$, and then the elastomer is rolled around the relaxed spring. When the actuator is subjected to an applied voltage Φ and an axial force *P*, the thickness of the laminate changes to $\lambda_3 L_3$, and the length of the spring changes to $\lambda_1 L_1$. However side 2 of the laminate, $\lambda_2^p L_2$, is constrained by the diameter of the spring and remains unchanged. The elastomer is taken to be incompressible, so that $\lambda_1 \lambda_2^p \lambda_3 = 1$.

During the operation, the actuator varies its state in two ways, as specified by two generalized coordinates: the stretch λ_1 in the axial direction, and the charge *Q* on one of the electrodes. Helmholtz free energy *A* of the actuator is prescribed as a function of the two generalized coordinates:

$$A(\lambda_1, Q) = \frac{\mu}{2}\left(\lambda_1^2 + \left(\lambda_2^p\right)^2 + \left(\lambda_1 \lambda_2^p\right)^{-2} - 3\right) L_1 L_2 L_3$$

$$+\frac{1}{2\varepsilon}\left(\frac{Q}{\lambda_1 L_1 \lambda_2^p L_2}\right)^2 L_1 L_2 L_3 + \frac{1}{2} K\left(\lambda_1 L_1 - \lambda_1^p L_1\right)^2 \qquad (1)$$

The free energy of the elastomer is the sum of the elastic energy, with μ being the shear modulus of the elastomer and the dielectric energy, with ε being the permittivity of the elastomer [6, 7]. The spring is taken to obey Hooke's law, with *k* being the stiffness of the spring.

When the actuator is in a state (λ_1, Q), in equilibrium with the applied axial force *P* and the applied voltage Φ, for any small change in the stretch and charge, $d\lambda_1$ and dQ, the change in the Helmholtz free energy equals the work done by the applied force and the voltage, namely [8].

$$dA = PL_1 + \Phi dQ \qquad (2)$$

Consequently, the force and the voltage are the partial differential coefficients of the free-energy function $A(\lambda_1, Q)$. The axial force is work-conjugate to the elongation:

$$P = \frac{\partial A(\lambda_1, Q)}{L_1 \partial \lambda_1} \qquad (3)$$

The voltage is work-conjugate to the charge:

$$\Phi = \frac{\partial A(\lambda_1, Q)}{\partial Q} \qquad (4)$$

Inserting (1) into (3), we can get:

$$\frac{P}{\mu L_2 L_3} = \left(\lambda_1 - \lambda_1^{-3}\left(\lambda_2^p\right)^{-2}\right) - \frac{1}{\lambda_1^3\left(\lambda_2^p\right)^2}\left(\frac{Q}{\sqrt{\mu\varepsilon} L_1 L_2}\right)^2 + \alpha\left(\lambda_1 - \lambda_1^p\right) \qquad (5)$$

where $\alpha = \frac{KL_1}{\mu L_2 L_3}$ is a dimensionless ratio between the stiffness of the spring and that of the elastomer. Equation (5) shows that the axial force is balanced by contributions of three origins: the elasticity of the elastomer, the permittivity of the elastomer, and the elasticity of the spring. Equation (5) can also be obtained by invoking the Maxwell stress [9, 10].

Inserting (1) into (4), we obtain that

$$\frac{\Phi}{L_3}\sqrt{\frac{\varepsilon}{\mu}} = \frac{1}{\left(\lambda_1 \lambda_2^p\right)^2}\left(\frac{Q}{\sqrt{\mu\varepsilon} L_1 L_2}\right) \qquad (6)$$

The actuator has three dimensionless parameters: the prestretches in the two directions in the plane of the elastomer, λ_1^p and λ_2^p, as well as the normalized stiffness of the spring α. These parameters of design are prescribed once the actuator is constructed.

Equations (5) and (6) are the equations of state, relating the dimensionless loading parameters, $P/(\mu L_2 L_3)$ and $\Phi/\left(L_3\sqrt{\mu/\varepsilon}\right)$, to the dimensionless generalized coordinates, λ_1 and $Q/\left(L_1 L_2 \sqrt{\varepsilon\mu}\right)$.

These nonlinear equations of state can be displayed graphically on a plane spanned by the two dimensionless generalized coordinates as shown in Fig. 2. Plotted on this plane are the lines of constant force and the lines of constant voltage. Fig. 2 can be used to locate the state of the actuator under prescribed axial forces ($P/\mu L_2 L_3 = 0$, $P/\mu L_2 L_3 = -2$, and $P/\mu L_2 L_3 = -4$) and voltages ($\Phi/L_3\sqrt{\mu/\varepsilon} = 0.07$, $\Phi/L_3\sqrt{\mu/\varepsilon} = 0.1$, and $\Phi/L_3\sqrt{\mu/\varepsilon} = 0.2$). Plotting the equations of state in Fig. 2, we have set the parameters of design to a particular set of values.

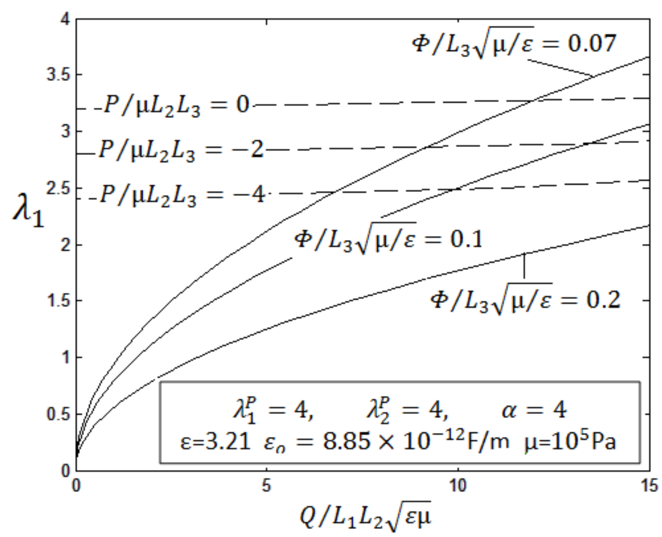


Fig. 2 A graphical representation of the equations of state

When the design variables $\left(\lambda_1^p, \lambda_2^p, and\ \alpha\right)$ are prescribed, the state of the actuator is characterized by two generalized coordinates: the stretch λ_1 in the axial direction, and the charge Q on one of the electrode.

3. Modes of Failure

The range of operation of an actuator is limited by various modes of failure. Each mode of failure restricts the state of the actuator to a region on the plane of the generalized coordinates. The common region that averts all modes of failure constitutes the set of allowable states. To illustrate the procedure to construct the region of allowable states, several representative modes of failure are considered [11, 12].

First *electromechanical instability* (EMI) of the elastomer is considered. When the applied voltage is increased, the elastomer reduces its thickness, so that the voltage induces a high electric field. The positive feedback between a thinner elastomer and a higher electric field may cause the elastomer to be reduced drastically, resulting in an electrical breakdown. This electromechanical instability can be analyzed by using a standard method in thermodynamics. [13].

Consider a three-dimensional space, with the generalized coordinates λ1 and Q being the horizontal axes, and the Helmholtz free energy A being the vertical axis. In this case, the free-energy function $A(\lambda_1, Q)$ is a surface. A point on the surface represents a state of the actuator, and a curve on the surface represents a path of actuation. Imagining a plane tangent to the surface at a state (λ_1, Q). The slopes of this tangent plane are $PL1$ and Φ, according to (3) and (4).

For a state (λ_1, Q) to be stable against arbitrary small perturbations in the generalized coordinates, the surface $A(\lambda_1, Q)$ must be convex at the point (λ_1, Q). This condition of stability is equivalent to the following set of inequalities:

$$\frac{\partial^2 A(\lambda_1, Q)}{\partial \lambda_1^2} > 0 \tag{7a}$$

$$\frac{\partial^2 A(\lambda_1, Q)}{\partial Q^2} > 0 \tag{7b}$$

$$\frac{\partial^2 A(\lambda_1, Q)}{\partial \lambda_1^2} \cdot \frac{\partial^2 A(\lambda_1, Q)}{\partial Q^2} > \left(\frac{\partial^2 A(\lambda_1, Q)}{\partial \lambda_1 \partial Q}\right)^2 \tag{7c}$$

Based on the three inequalities, (7a) ensures mechanical stability, (7b) electrical stability, and (7c) electromechanical stability. Using (1), it is noticeable that (7a) and (7b) are satisfied for all values of (λ_1, Q), but (7c) is violated for some values of (λ_1, Q). A combination of (1) and (7c) shows that the electromechanical instability sets when:

$$\frac{Q}{\sqrt{\mu\varepsilon}L_1L_2} = \sqrt{(1+\alpha)\lambda_1^4\left(\lambda_2^p\right)^2 + 3} \tag{8}$$

This equation corresponds to the curve marked by EMI in Fig. 3. The curve divides the (λ_1, Q)plane into two regions. Above the curve, the actuator is stable against small perturbation of the generalized coordinates. Below the curve, the actuator undergoes electromechanical instability.

The second mode of failure is the *electrical breakdown* (EB) of the elastomer. Even before the electromechanical instability sets, the electric field in the elastomer may become too high, leading to localized conduction path through the thickness of the elastomer. For the complexity of the microscopic process of electrical breakdown, it will not be addressed in this paper. To illustrate the procedure of design, we assume that electrical breakdown occurs when the true electric field exceeds a critical value Ec. For the ideal dielectric elastomer, D = εE, where the true electric displacement is $D = Q/\lambda_1\lambda_2^p L_1 L_2$, the condition for electric breakdown is

$$\frac{Q}{\sqrt{\mu\varepsilon}L_1L_2} = \lambda_1\lambda_2^p E_C\sqrt{\frac{\varepsilon}{\mu}} \tag{9}$$

Equation (9) corresponds to the straight line marked by EB on the (λ_1, Q)plane as shown in Fig. 3. The actuator in a state in the region above this straight line will not suffer from the electrical breakdown.

Loss of tension of the elastomer when large voltage Φ is considered, or axial force P is compressive and of a large magnitude, the stress in the plane of the elastomer may cease to be tensile. This loss of tension will cause the elastomer to buckle out of the plane, so that elastomer will no longer generate force of actuation. To avert this mode of failure, the stress is required to be tensile in every direction in the plane of the elastomer. That is, both the stress along the axial direction and the stress in the circumferential direction are required to be tensile, $S_1 > 0$ and $S_2 > 0$. Following [14], the nominal stress in the axial direction is obtained in terms of the two generalized coordinates:

$$\frac{S_1}{\mu} = \left(\lambda_1 - \lambda_1^{-3}\left(\lambda_2^p\right)^{-2}\right) - \left(\frac{Q}{\sqrt{\mu\varepsilon}L_1L_2}\right)^2 \lambda_1^{-3}\left(\lambda_2^p\right)^{-2} \tag{10}$$

Setting the critical condition in (10), we obtain that

$$\frac{Q}{\sqrt{\mu\varepsilon}L_1L_2} = \sqrt{\lambda_1^4\left(\lambda_2^p\right)^2 - 1} \tag{10a}$$

Similarly, nominal stress s_2 in terms of the two generalized coordinates can be obtained:

$$\frac{S_2}{\mu} = \left(\lambda_2^p - \left(\lambda_2^p\right)^{-3}\lambda_1^{-2}\right) - \left(\frac{Q}{\sqrt{\mu\varepsilon}L_1L_2}\right)^2 \left(\lambda_2^p\right)^{-3}\lambda_1^{-2} \tag{11}$$

Setting the critical condition s_2=0 in (11), the following equation can be obtained:

$$\frac{Q}{\sqrt{\mu\varepsilon}L_1L_2} = \sqrt{\lambda_1^2\left(\lambda_2^p\right)^4 - 1} \tag{11a}$$

The critical conditions for loss of tension, s_1=0 and s_2=0, are plotted in Fig. 3. A comparison of (8) and (10a) shows that, for spring-roll actuators, loss of tension in the axial direction will always precede electromechanical instability. In contrary, other types of dielectric elastomer actuators may fail by electromechanical instability [15, 16].

IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 2, September 2011
ISSN (Online): 1694-0814
www.IJCSI.org

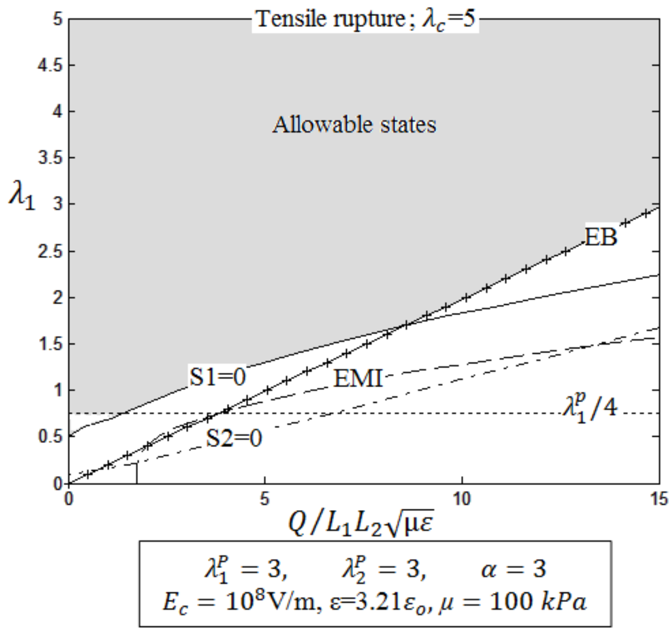


Fig. 3 A graphical representation of modes of failure

Next *tensile rupture* of the elastomer is considered. When an elastomer is stretched too severely, the elastomer may rupture. The critical condition for tensile rupture is not well quantified. Here the simple criterion that the elastomer will rupture when either stretch, λ_1 or λ_2 exceeds a critical value λ_c is used. A representative value $\lambda_c = 5$ is included in Fig.3.

The *compressive limit* of the spring is finally considered. The spring in the spring-roll actuator is designed to be under compression. When the spring is compressed excessively, however, it may deform plastically. The length of the spring at its relaxed state is $\lambda_1^p L_1$, and the length of the actuated spring is $\lambda_1 L_1$. We assume that the spring deforms plastically when λ_1^p/λ_1 exceeds a critical value *c*, which we set to be *c* = 4. In the (λ_1, Q) plane, Fig. 3, the region above the line $\lambda_1 = \lambda_1^p/c$ will guarantee that the spring remains elastic.

The modes of failure discussed in this section are all averted in the shaded region in Fig. 3. As evident from the above discussion, this region of allowable states will depend on the critical conditions for various modes of failure.

4. Actuation Range

The actuation range is an important issue because the applied voltage may take different values within the actuation range.
In this paper the concept of actuation range depends upon whether the actuator has a fixed load actuator or variable multi-load actuator. The fixed load actuator is the actuator subject to a fixed axial force (i.e., a dead weight). In this case the actuation range is very small value starting at the state of zero charge, and ending at the state where the line of a failure mode intersects the axial force line. Fig. 4 shows the actuation range for a fixed load actuator. The actuation range extends from $\lambda_1 = 4.4$ to $\lambda_1 = 4.515$ (the two highlighted black points) when $\lambda_1^p = 5$, $\lambda_2^p = 5$, $\alpha = 10$, $P/\mu L_2 L_3 = -1$, and $E_C = 10^8 v/m$.

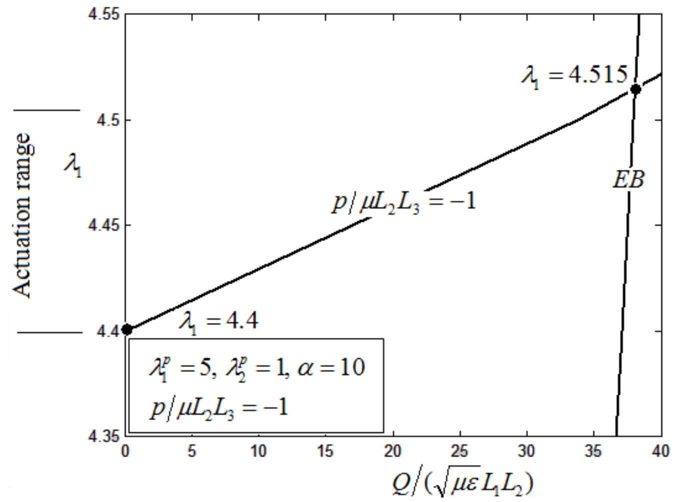


Fig. 4 The actuation range for a fixed load actuator when $\lambda_1^p = 5$, $\lambda_2^p = 5$, $\alpha = 10$, $P/\mu L_2 L_3 = -1$, and $E_C = 10^8 v/m$.

The load of the exchangeable multi-load actuator can be replaced by another unequal load and this load in turn can be replaced by another unequal one and so on. The actuation range of the exchangeable multi-load actuator starts at the state of zero charge of heaviest load and ends at the state where the line of a failure mode intersects the line of the lightest load. Fig. 5 shows the actuation range of the exchangeable multi-load actuator. The actuator given in fig. 5 subjects to one of the following dimensionless loads ; $P/\mu L_2 L_3 = -1$, $P/\mu L_2 L_3 = -10$, $P/\mu L_2 L_3 = -20$, or $P/\mu L_2 L_3 = -30$. The actuation range in this case starts at $\lambda_1 = 1.95$ and ends at $\lambda_1 = 4.52$ (the two highlighted black points).

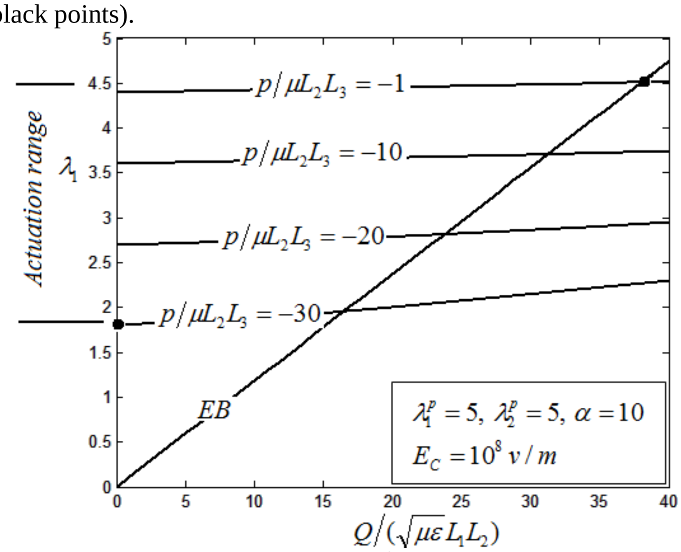


Fig. 5 The actuation range of the exchangeable multi-load actuator

5. Spring-Roll Dielectric Elastomer Actuator Design

Parameters of design are prescribed once the actuator is constructed therefore they must well studied and selected to be optimal.

5.1 Spring-Roll Dielectric Elastomer Actuator Optimal Design Parameter

To optimize an actuator, the actuator should have certain features such as higher actuation and the ability to burden heavier loads. To address this issue, we have to figure out the space of design parameters values that confirm these features.

Spring-roll dielectric elastomer optimal design parameters according to [17] are $\lambda_1^p = 5,\ \lambda_2^p = 2,\ \alpha = 10$. These parameters of design are prescribed once the actuator is constructed.

In this paper we will address the effect of λ_2^p on actuation λ_1, axial force $P/\mu L_2 L_3$, and $\Phi/\left(L_3\sqrt{\mu/\varepsilon}\right)$ to prove that the value $\lambda_2^p = 2$ that was considered as an optimal value should be changed to $\lambda_2^p = 5$.

The actuation λ_1 as a function of λ_2^p

The obtained results in simulation show that when λ_2^p increases, the actuation λ_1 slightly decreases. There is no big difference between actuation at $\lambda_2^p = 1$ and actuation at $\lambda_2^p = 5$ even at higher values of the dimensionless charge. From fig. 6, it is clear that λ_2^p slightly affects λ_1.

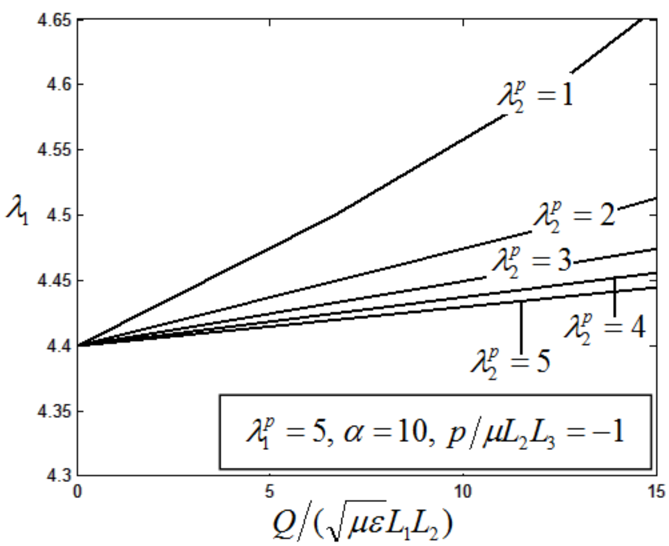


Fig. 6 The effect of λ_2^p on the actuation λ_1 at $\lambda_1^p = 5,\ \alpha = 10,\ and\ P/\mu L_2 L_3 = -1$

The dimensionless force (load) $P/\mu L_2 L_3$ as a function of λ_2^p

The dimensionless axial force $P/\mu L_2 L_3$ is a nonlinear function of λ_2^p. λ_2^p slightly affects the value (modulus) of the axial force. Fig. 7 shows the effect of λ_2^p on the axial force.

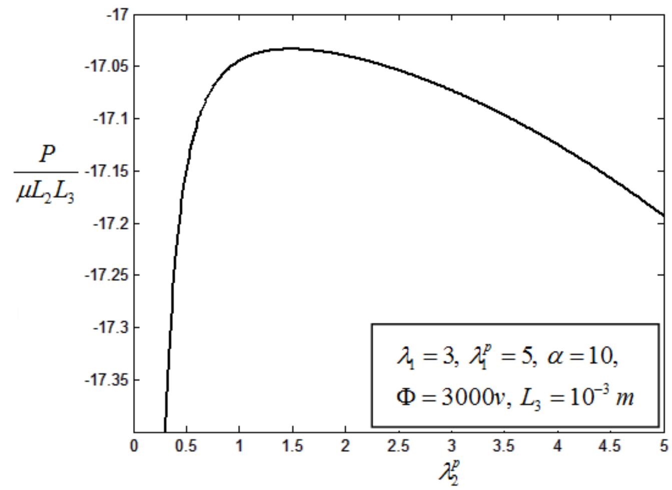


Fig. 7 The effect of λ_2^p on the axial force $P/\mu L_2 L_3$ at $\lambda_1 = 3,\ \lambda_1^p = 5,\ \alpha = 10,\ L_3 = 10^{-3}$, and $\Phi = 3000v$

The effect of λ_2^p on the applied voltage Φ

λ_2^p Slightly affects both actuation λ_1 and dimensionless axial load $p/\mu L_2 L_3$ but it has a great effect on the applied voltage Φ. When the applied voltage decreases, λ_2^p will increase. Let us take the applied voltage at a critical state Φ_C where one failure mode sets in as an example for the relation between the applied voltage and λ_2^p. When $\lambda_1^p = 5,\ \lambda_2^p = 1,\ \alpha = 10,\ p/\mu L_2 L_3 = 0,\ and\ L_3 = 1mm$, the critical dimensionless applied voltage is 0.215 and the critical applied voltage is $12756\ v$, as shown in fig. 8.

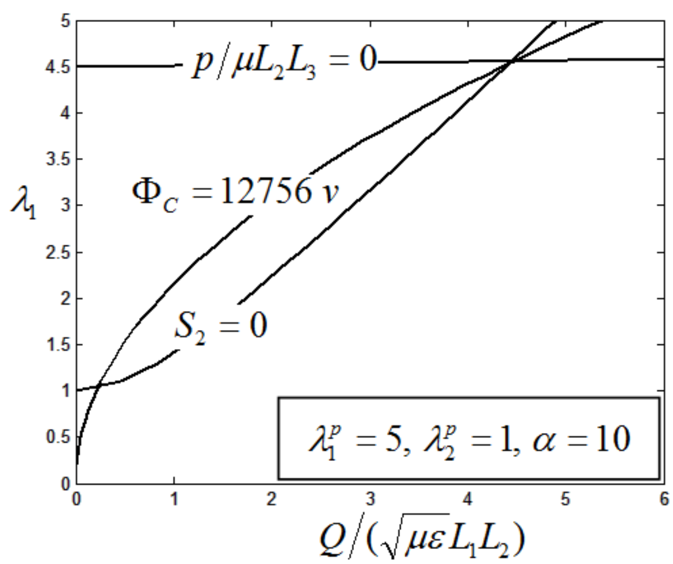


Fig. 8 Critical applied voltage at $\lambda_1^p = 5,\ \lambda_2^p = 1,\ \alpha = 10,\ p/\mu L_2 L_3 = 0,\ and\ L_3 = 1mm$

When $\lambda_1^p = 5,\ \lambda_2^p = 2,\ \alpha = 10,\ p/\mu L_2 L_3 = 0,\ and\ L_3 = 1mm$, the critical dimensionless applied voltage is 0.182 and the critical applied voltage is $10798v$, look at fig. 9.

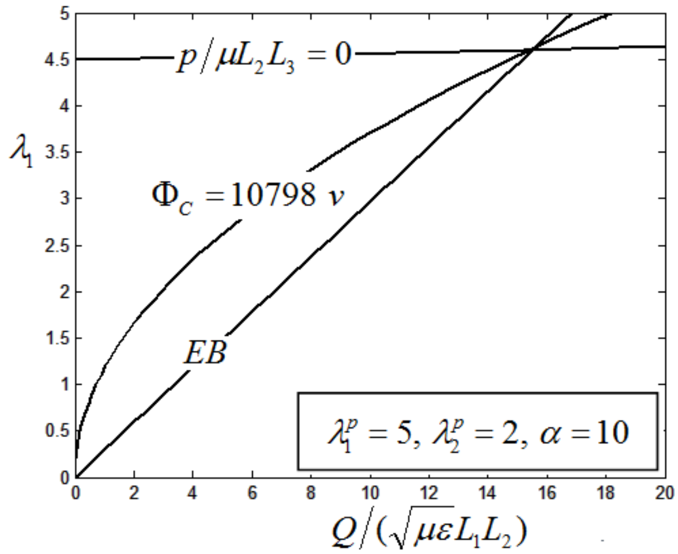


Fig. 9 Critical applied voltage at $\lambda_1^p = 5,\ \lambda_2^p = 2,\ \alpha = 10,$ $p/\mu L_2 L_3 = 0,\ and\ L_3 = 1mm$

When $\lambda_1^p = 5, \lambda_2^p = 3,\ \alpha = 10,\ p/\mu L_2 L_3 = 0,\ and\ L_3 = 1mm$, the critical dimensionless applied voltage is 0.122 and the critical applied voltage is $7238v$, as shown in fig. 10.

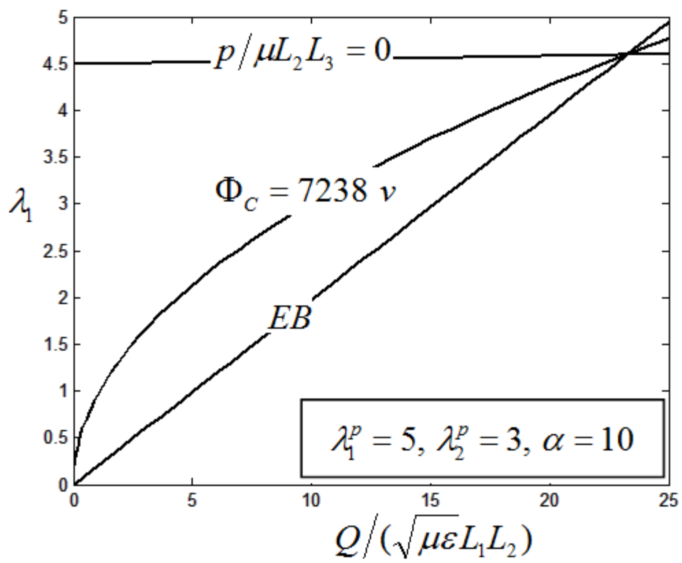


Fig. 10 Critical applied voltage at $\lambda_1^p = 5,\ \lambda_2^p = 3,\ \alpha = 10,$ $p/\mu L_2 L_3 = 0,\ and\ L_3 = 1mm$

When $\lambda_1^p = 5, \lambda_2^p = 4,\ \alpha = 10,\ p/\mu L_2 L_3 = 0,\ and\ L_3 = 1mm$, the critical dimensionless applied voltage is 0.092 and the critical applied voltage is $5458v$, look at fig. 11.

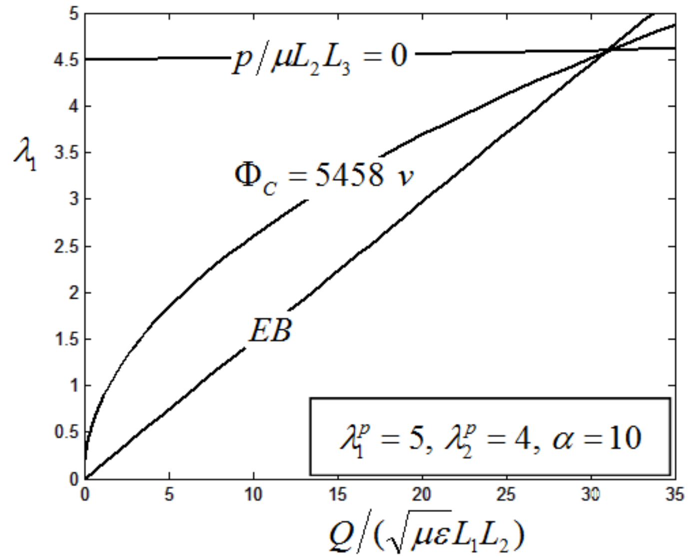


Fig. 11 Critical applied voltage at $\lambda_1^p = 5,\ \lambda_2^p = 4,\ \alpha = 10,$ $p/\mu L_2 L_3 = 0,\ and\ L_3 = 1mm$

When $\lambda_1^p = 5, \lambda_2^p = 5,\ \alpha = 10,\ p/\mu L_2 L_3 = 0,\ and\ L_3 = 1mm$, the critical dimensionless applied voltage is 0.073 and the critical applied voltage is $4331v$, look at fig. 12.

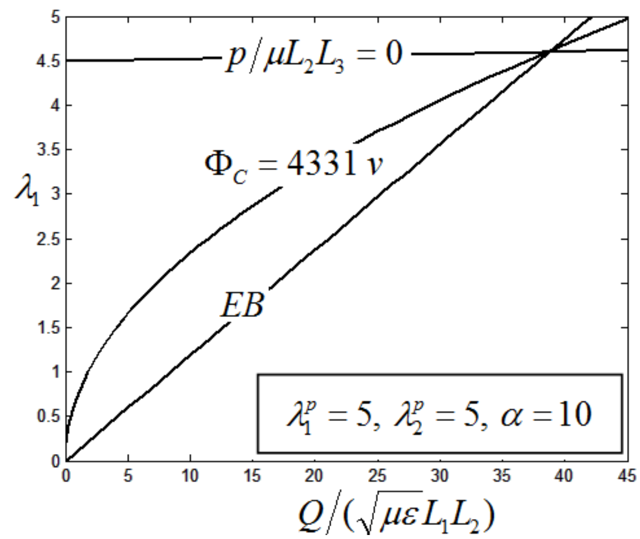


Fig. 12 Critical applied voltage at $\lambda_1^p = 5,\ \lambda_2^p = 5,\ \alpha = 10,$ $p/\mu L_2 L_3 = 0,\ and\ L_3 = 1mm$

The voltage at a critical state Φ_C decreases from 12756v at $\lambda_2^p = 1$ to 4331v at $\lambda_2^p = 5$. In general, the applied voltage decreases as λ_2^p increases and high voltage problems can be avoided therefore the optimal design parameters should be changed from;

$$\boxed{\lambda_1^p = 5,\ \lambda_2^p = 2,\ \alpha = 10}$$

to:

$$\boxed{\lambda_1^p = 5,\ \lambda_2^p = 5,\ \alpha = 10}$$

5.2 Spring-Roll Dielectric Elastomer Actuator Optimal Design Parameters

Spring-roll dielectric elastomer actuator is designed according to equations of state (eq. 5, and eq. 6). The first step of designing the elastomer actuator is to be aware of the constant values of equations of state;

μ: The shear modulus of the dielectric elastomer (VHB 4910 material) equals $10^5\,Pascal$ [18].

ε_r: The relative dielectric constant equals 3.21 for VHB 4910 material, at a frequency of 1 KHz.

ε_o: The permittivity equals 8.85×10^{-12} Farad/meter.

E_C: The electric breakdown of VHB 4910 material equals $10^8\, v/m$.

$\lambda_1^p,\ \lambda_2^p,\ \alpha$: The design parameters are dealt as constant values; $\lambda_1^p=5,\ \lambda_2^p=5,\ \alpha=10$.

L_3: The thickness of the dielectric elastomer (VHB 4910 material), 3M company produces it in a thickness of 1 mm therefore it is dealt as a constant value.

Spring-roll dielectric elastomer specifications prescribed by the customer

The customer has to specify three specifications before the actuator is constructed; axial force/forces in Newtons, length of the actuator at relax $L_{1\,relax}$ in meters, and actuator maximum required length $L_{1\max}$ in meters. Dimensionless axial force/forces should not be positive (tensile force/forces) and should not situate in the region of modes of failure. If this happens, the customer is reviewed to modify his request.

Determining the maximum actuation (stretch) $\lambda_{1\max}$

$$L_{1\,relax}=\lambda_1^p L_1$$
$$L_{1\max}=\lambda_{1\max}L_{1\,relax}$$
$$\lambda_{1\max}=L_{1\max}/L_{1\,relax}$$

Determining the range of the applied voltage

The applied voltage may take infinite different values along the actuation range starting at $\Phi=0$ and ending at $\Phi=\Phi_C$. $E=\Phi/\lambda_3 L_3$ where E is the electric field across the elastomer membrane. When $\Phi=\Phi_C$, $E=E_C$, where E_C is the electric breakdown field. Then $E_C=\Phi_C/\lambda_3 L_3$. Since the dielectric elastomer is incompressible material, $\lambda_3=1/(\lambda_2^p\lambda_1)$.

$$\Phi_C=\frac{10^8 L_3}{\lambda_2^p\lambda_1} \tag{12}$$

where;

$E_C=10^8\, v/m$ for VHB 4910 dielectric elastomer material.

λ_2^p according to new derived optimal design parameters equals "5".

3 M company produces VHB 4910 membrane in a thickness of 1 mm. Φ_C can be determined.

Determining L_2, the width of the elastomer membrane

Substituting $\lambda_{1\max}$ *and* Φ_C in eq. 6, dimensionless charge $Q/(\sqrt{\mu\varepsilon}L_1L_2)$ can be determined. Substituting the determined $Q/(\sqrt{\mu\varepsilon}L_1L_2)$ and $\lambda_{1\max}$ in eq. 5, L_2 can be determined.

Determining *K*, the stiffness of the spring

$L_1=L_{1\,relax}/\lambda_1^p$, L_2 has been determined in section 6.4. Substituting L_1 *and* L_2 in the following eq; $\alpha=KL_1/\mu L_2L_3$, K can be determined.

Determining *n*, number of turns of actuator membrane around the spring

Approximately the number of turns of the actuator membrane can be determined by dividing the width of the actuator L_2 by the circumference of the spring.

$$n=\frac{\lambda_2^p L_2}{2\pi r} \tag{13}$$

where; r is radius of the used spring.

5.3 Samples of Spring-Roll Dielectric Elastomer Design Results

In appendix A we develop equations of state based on Mat-Lab program help design spring-roll dielectric elastomer actuators. The customer has to specify the relaxed length of the actuator $L_{1\,relax}$, the maximum required length the actuator has to achieve $L_{1\max}$, and axial load the actuator will treat with.

Table 1: Samples of spring-roll dielectric elastomer design specifications

p	L_{1relax}	$L_{1\max}$	r	L_1	L_2	L_3	Φ_c	$\lambda_{1\max}$	k	n
-4	0.01	0.02	0.001	0.02	0.0014	0.001	10000	2	679.7	1.08
-8	0.02	0.05	0.001	0.004	0.0034	0.001	8000	2.5	846	2.7

-12	0.03	0.09	0.001	0.006	0.0067	0.001	6667	3	1114	5.3
-16	0.04	0.14	0.0025	0.008	0.013	0.001	5714	3.5	1624	4.14
-20	0.05	0.05	0.0025	The actuator is overloaded						
-24	0.06	0.07	0.0025	The actuator is overloaded						
+28	0.07	0.14	0.0025	The load is a tensile forces						
+32	0.08	0.184	0.0025	The load is a tensile force						
-36	0.09	0.225	0.005	0.0180	0.0152	0.001	8000	2.5	846	2.4
-40	0.1	0.29	0.005	0.02	0.021	0.001	6897	2.9	1048	3.3
-44	0.11	0.352	0.005	0.022	0.028	0.001	6250	3.2	1274	4.5
-48	0.12	0.42	0.005	0.024	0.039	0.001	5714	3.5	1624	6.2
-52	0.13	0.234	0.005	0.026	0.0164	0.001	11111	1.8	629	2.6
-56	0.14	0.294	0.005	0.028	0.0198	0.001	9524	2.1	708	3.2
-60	0.15	0.345	0.005	0.03	0.231	0.001	8696	2.3	771	3.7
-64	0.16	0.432	0.005	0.032	0.03	0.001	7407	2.7	937	4.8
-68	0.17	0.527	0.005	0.034	0.0404	0.001	6452	3.1	1189	6.4
-72	0.18	0.702	0.005	0.036	0.092	0.001	5128	3.9	2555	14.6
-76	0.19	0.76	0.01	0.038	0.1132	0.001	5000	4	2980	9
-80	0.2	0.82	0.01	0.04	0.143	0.001	4878	4.1	3576	11.4
-84	0.21	0.399	0.01	0.042	0.0275	0.001	10526	1.9	654	2.2
-88	0.22	0.484	0.01	0.044	0.0325	0.001	9090	2.2	738	2.6
-92	0.23	0.598	0.01	0.046	0.0409	0.001	7692	2.6	889	3.3
-96	0.24	0.576	0.01	0.048	0.0387	0.001	8333	2.4	807	3.1
+100	0.25	0.975	0.01	The load is a tensile force						

Using the above mentioned software, we can design actuator specifications; the dimensions of the actuator membrane, the stiffness of the used spring, the maximum applied voltage, and maximum achieved actuation. Table 1 includes samples of design specifications where the dimensions $L_1, L_2, L_3, L_{1\max}, L_{1relax}, and\ r$ are measured in *meters* , $\Phi_{\max}$ *or* $\Phi_{critical}$ is measured in volts, stiffness of the spring k is measured in *newtons per meter* , and the axial load p is measured in *newtons* .

6. Spring-Roll Dielectric Elastomer Actuator Voltage Supply

In spring-roll dielectric elastomer actuator system to achieve a specific actuation λ_1, a specific voltage should be applied to the electrodes of the actuator and specific charges should flow to them, therefore a voltage supply whose output ranges from several volts up to $15000\,v$ is required. This voltage supply should be adjusted automatically to any voltage between 0 and 15000v whatever the value of the required voltage is.

A charge pump driven by a low voltage switching power supply and a voltage drive is presented in this paper [19].

6.1 Charge Pump

Increased voltage levels are obtained in a charge pump as a result of transferring charges to a capacitive load and do not involve amplifiers or transformers. The charge pump is constructed by n cascaded voltage doublers. Charge pump operates by switching on and off a large number of switches which charge and discharge a large number of capacitances, transferring energy to the output load. Switched-capacitor charge pumps have exponentially growing voltage gain as a function of the number of stages (voltage doublers) up to 2^n [20].

A switched-capacitor organization of a two phase DC-DC voltage doubler is shown in Fig. 13. It contains 2 clock controlled switches and 2 capacitors.

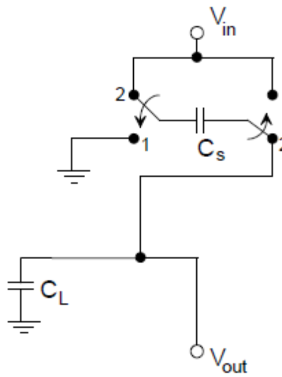


Fig. 13 The DC-DC TPVD voltage doubler

For a simple explanation of the voltage doubler operation, let us assume that the switches and capacitors are all ideal. That is, we assume that there is no leakage current in capacitors, switches dissipate no energy and the electric charge transferring is instantaneous. Fig. 14.a shows the equivalent circuit of the voltage doubler when the circuit is in the *Kth* iteration cycle and the clock is in phase 1. At this time instance, the load capacitor C_L holds the previous voltage value.

$$V_{out}^{[K]} = V_{out}^{[K-1]} \tag{14}$$

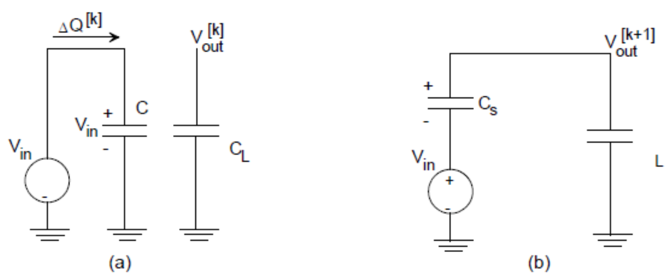


Fig. 14 The equivalent circuits in two clock phases.

The voltage across the capacitor C_S changes from $-(V_{in}-V_{out}^{[K-1]})$ to V_{in}. The charge $\Delta Q^{[K]}$ transferred from the voltage source V_{in} to C_S is obtained from

$$\Delta Q^{[K]} = C_S\left(V_{in} - \left(-\left(V_{in} - V_{out}^{[K-1]}\right)\right)\right) = C_S\left(2V_{in} - V_{out}^{K}\right) \tag{15}$$

Equation (15) implies that the voltage source *Vin* would stop transferring charge to C_S if $V_{out}^{[K]} = 2V_{in}$.

Fig. 14.b shows the equivalent circuit of the voltage doubler when the circuit is in the $(k+1)th$ iteration cycle and the clock is in phase 2. According to the charge conservation law at the node connecting C_S and C_S, and evaluating charges stored in capacitors as *Q* = *C V*, the relationship between voltages at $k\,th$ and $(k+1)th$ iteration can be expressed by

$$V_{in}\times C_S + V_{out}^{[K]}\times C_L = \left(V_{out}^{[K+1]} - V_{in}\right)\times C_S + V_{out}^{[K+1]}\times C_L \tag{16}$$

If we set $r = \dfrac{C_S}{C_L + C_S}$, to represent the capacitor ratio, then

$$V_{out}^{[K+1]} = (1-r)V_{out}^{K} + 2rV_{in} \tag{17}$$

where 0 < *r* < 1. Thus, V_{out} can be represented as a sequence of the iteration index *k*.

Fig. 15 shows the voltage gains *Av* as a function of the iteration index *k*, with different *r*. The smaller the *r*, the larger the ratio of the grounded capacitor C_L to the switched capacitor *Cs*. It is clear that the final (steady state) value of the voltage gain *Av* is 2 independently of the capacitor ratio *r*. That is, the circuit in Fig. 14 works as a voltage doubler provided that the voltage source V_{in} supplies enough charge to the charge pump. The larger C_L (smaller *r*) requires more clock cycles (bigger *k*) to reach the desired output voltage. The value of *r* does not influence the final voltage gain.

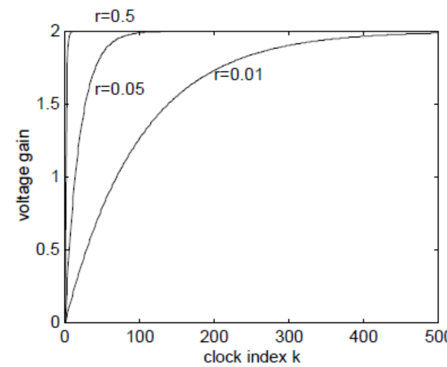


Fig. 15 The voltage gain as a function of the clock index *k*

6.2 Proposed Actuator Voltage Supply

A new organization of spring-roll dielectric elastomer actuator voltage supply is presented in this paper. Switched multistage charge pump driven by a controlled low voltage switching power supply and a voltage driver is used as a variable output high voltage supply. Coarse adjustment of the output voltage is automatically accomplished by connecting a specific number of the stages of the charge pump to the actuator. Fine adjustment of the output voltage is automatically accomplished by controlling the width of the pulse of the low voltage switching power supply. Fig. 16 shows the block diagram of the proposed voltage supply.

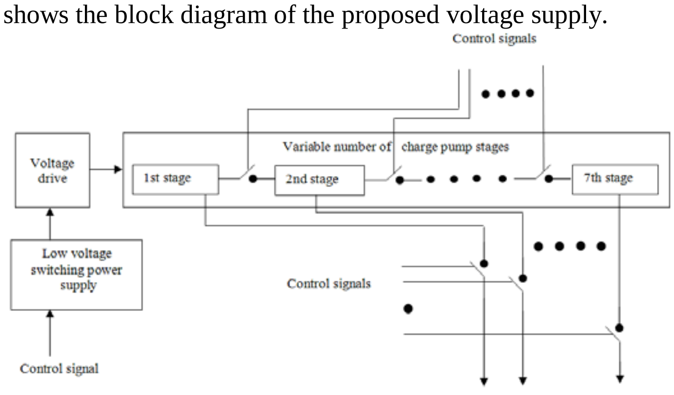


Fig. 16 Block diagram of the proposed voltage supply

The voltage driver is designed to supply 120 volt at a 50% duty cycle in the low voltage switching power supply. The input and output voltages of each stage of the charge pump are indicated in table 2.

Table 2: The input and output and output of each stage of the charge pump

Charge pump stage number	***Input voltage***	Output voltage
1	120	240
2	240	480
3	480	960
4	960	1920

5	1920	3840
6	3840	7680
7	7680	15360

At reset state all switches are switched off, duty cycle of low voltage switching power supply is 50% and the output of the voltage driver is 120 volt. Suppose that the actuator needs instantaneous 4000 volt to achieve a specific actuation, the control circuit connects stages umber 1, 2, 3, 4, and 5 in series and the output is taken from the fifth stage (the automatic coarse adjustment), then the control circuit increases the duty cycle (pulse width) of the voltage switching power supply until the output of the fifth stage reaches 4000 volt (the automatic fine adjustment).

7. Conclusion

In this paper we search two problems, the first is to develop an optimal design parameters based spring-roll dielectric elastomer actuator, and the second is to develop a controlled wide range output high voltage supply. As for the first problem, we prove that λ_2^p has slight effect on both actuation λ_1 and axial load, but has a high effect on the applied high voltage, it reduces the required applied high voltage down to 35% when raised to 5. Therefore the optimal design parameters are changed to; $\lambda_1^p = 5, \lambda_2^p = 5, \alpha = 10$.

In this paper we develop procedures of designing a spring-roll dielectric elastomer actuator, involve a software help designing the actuator [appendix 1], and mention samples of the results of designed spring-roll dielectric elastomer actuators.

In this paper the concept of actuation range depends upon whether the actuator is a fixed load actuator or an exchangeable multi-load actuator.

As for the second problem a switched multistage charge pump driven by a controlled low voltage switching power supply and a voltage driver is proposed as a wide range output high voltage supply.

Solving the two problems a complete spring-roll dielectric elastomer actuator system is developed.

For future work, the instantaneous change of actuation of the spring-roll dielectric elastomer actuator is required. This actuation can be controlled by controlling both charges pumped to the electrodes of the actuator and the voltage applied to them. In a next paper we shall develop a sub control system to control charges pumped to electrodes of the actuator and another sub control system to control the applied voltage. Modes of failure of the actuator will be avoided using the proposed control system. Sensors of charges and sensors of high voltage will be used. Intelligent system will be proposed.

Appendix A

```
function [l1, l2, d1max, phic, k, n]=actdesign (p, l1relax, l1max, r)
% We develop this equations of state based Mat-Lab program to design spring-roll
% dielectric elastomer actuators.
% l1max is the maximum required actuator length.
% l1relax is the spring-roll dielectric elastomer actuator at relax.
% p (lowr case letter)is the compressive load in Newtons.
% P (upper case letter) is the dimensionless axial force.
% phic is the applied voltage at a critical state.
% d1max is maximum actuation.
% r is the radius of spring.
% n number of turns of the actuator membrane around the spring.
% k is the stiffness of the spring.
% l1, l2 and l3 are the length, width and thickness of the actuator membrane.
% mu is the shear modulus of the actuator material (dielectric elastomer).
% epsilon is the dielctric of the actuator material.
% d1p & d2p are the prestretches in in the length and width of the
% membrane.
% Q (upper letter) is the dimensionless charge.
% Given (l1max, l1relax, and p), [l1, l2, d1max, phic, k]can be determined.
mu=10.^5;
epsilon=3.21*8.85*10.^-12;
d1p=5;
d2p=5;
a=10;
l3=10.^-3;
if p<0
   l1=l1relax/d1p;
   d1max=l1max/l1relax;
   phic=(10.^8*l3)/(d2p*d1max);
   Q=(d1max.^2)*(d2p.^2)*(phic/l3)*(sqrt(epsilon/mu));
   P=d1max - ((d1max.^-3)*(d2p.^-2)) - ((Q.^2)*(d1max.^-3)*(d2p.^-2)) + a*(d1max-d1p);
   if P>-35;
      l2=p/(P*mu*l3);
      k=(a*mu*l2*l3)/l1;
      n=(d2p*l2)/(2*pi*r);
   else
      disp ('The actuator is overloaded');
% The word "overloaded" means that the actuator may subjects to a failure mode.
      l1=[];
      l2=[];
      d1max=[];
      phic=[];
      k=[];
      n=[];
   end
else
```

```
  disp ('The Load is a tensile force');
  % The load should be a compressive load (-ve)and should not be tensile (+ve).
  l1=[];
  l2=[];
  d1max=[];
  phic=[];
  k=[];
  n=[];
end
```

References

[1] Rui Zhang, Patrick Lochmatter, Andreas Kun and Gabor Kovacs, "Spring Roll Dielectric Elastomer Actuators for a Portable Force Feedback Glove," Proceedings of SPIE Vol. 6168, 61681T, (2006).

[2] Q. Pei, R. Pelrine, S. Stanford, R. Kornbluh, M. Rosenthal, Synthetic Metals, 135-136, 129-131(2003).

[3] Q. Pei, R. Pelrine, S. Stanford, R. Kornbluh, M. Rosenthal, K. Meijer, R. "Full, Smart Structures and Materials," Proc. of SPIE Vol. 4698, 246 (2002).

[4] Guggi Kofod, "The static actuation of dielectric elastomer actuators: how does pre-stretch improve actuation?," J. Physics. D: Appled Physics. 41 215405 (11pp) (2008)

[5] Woosang JUNG, Yutaka TOI, "Computational Modeling of Electromechanical Behaviors of Dielectric Elastomer Actuators," Proceedings of the International MultiConference of Engineers and Computer Scientists 2010 Vol III, IMECS 2010, March 17-19, 2010, Hong Kong.

[6] X. Zhao, W. Hong, Z. Suo, Physical review B 76 (2007).

[7] G. Kofod, M. Paajanen, S. Bauer, Applied Physics" a-Materials Science & Processing 85, 141(2006).

[8] Z. G. Suo, X. H. Zhao, and W. H. Greene, Journal of the Mechanics and Physics of Solids 56, 467-486 (2008).

[9] R. Zhang, P. Lochmatter, A. Kunz, G. Kovacs, Smart Structures and Materials, Proc. of SPIE Vol. 6168 (2006).

[10] G. Kovacs, P. Lochmatter, M. Wissler, Smart Materials and Structures, Vol. 16, S306-S317 (2007).

[11] R. E. Pelrine, R. D. Kornbluh, J. P. Joseph, Sensors and Actuators A 64, 77 (1998).

[12] J. S. Plante, S. Dubowsky, International Journal of Solids and Structures 43, 7727 (2006).

[13] A. N. Norris, Applied Physics Letters 92, 026101 (2008).

[14] X. Zhao, Z. Suo, Applied Physics Letters 91 (2007).

[15] Tianhu He, Xuanhe Zhao and Zhigang Suo, "Equilibrium and stability of dielectric elastomer membranes undergoing inhomogeneous deformation," School of Engineering and Applied Sciences, Harvard University, (2-10-2008).

[16] Christoph Keplinger, Martin Kaltenbrunner, Nikita Arnold, and Siegfried Bauer, Röntgen's electrode-free elastomer actuators without electromechanical pull-in instability,Applied physical science, (December 15, 2009).

[17] Medhat H. Ahmed and Besada A. Anees, "Parameter Optimization in Spring-Roll Dielectric Elastomer Actuator Design",The Journal of Engineering Sciences, Faculty of Engineering , University of Assiut, Assiut, Egypt, 2011.

[18] Mickael Moscardo, Xuanhe Zhao, Zhigang Suo,_ and Yuri Lapusta, "On designing dielectric elastomer actuators," Journal of Applied Physics 104, 093503, (2008).

[19] Sandeep Pemmaraju, "High voltage charge pump circuit for an ion mobility spectrometer", submitted for the degree of Master of science in Engineering, Electrical Engineering, Boise State University, 2004.

[20] Janusz A. Starzyk, Ying-Wei Jan, and Fengjing Qiu, "A DC-DC charge pump design based on voltage doublers" IEEE transactions on circuits and systems – I: fundamental theory and applications, vol. 48, no. 3, March 2001.

Medhat H A Awadalla obtained his B.Sc. degree from Helwan University in 1991 in the Electronics and Communications Department and his M.Sc in the field of reconfigurable computer architecture in 1996 from Helwan University. He received his PhD from Cardiff University, UK in the field of mobile robots in 2005. He was a postdoctoral fellow at Cardiff University in 2006 and currently he is working as an Assistant Professor in Helwan University. His research interests include real time systems, parallel and distributed systems, grid computing, sensor networks, and robotics.

Besada Adeeb Anees obtained his B.Sc. degree from Menoufia University in 1977 in the Electronics Department and his M.Sc in the field of control in 2006 from Helwan University. He is currently working in his PhD scheme in the field of Artificial Muscles. His research interests include control systems, Fuzzy logic, neural networks, swarm based control.

Performance Modeling and Analysis of Distributed-Based Polling Networks

Salman Ali AlQahtani
Computer Science Department, King Fahad Security College,
Riyadh, Saudi Arabia

Abstract
In this paper, the performance evaluation of distributed polling networks using the analytical equations and a developed simulation model is considered. The exhaustive service versus K-limited service are also investigated and compared. Other main measurements such as average cycle time, walking time are also compared and calculated. In general, the result of both our designed simulation and analytical equations are almost in total agreements. Our simulation model and performance evaluations can be used as a basis to investigated or develop many polling-based wireless protocols. The differences between the analytical and simulation values came from the approximation done on the analytical equations.
Keywords: Distribution polling, performance analysis, queuing system, system modulation.

1. Introduction

Polling is a general way of multiplexing service requests for a single server from multiple stations. The basic concept of polling network system is to have multiple queues accessed by a single server in cyclic order [1-5]. polling systems whose study dates from the late 1950's found wide applications, for example in telecommunications, transportation, healthcare, etc., and have been the subject of numerous studies [6-14]. This is because it provides general performance analysis criteria for a large number of demand-based and multiple access schemes in computer and communication schemes. Nowadays, polling has been included as a resource sharing mechanism in the Medium Access Control (MAC) protocol for several wireless networks, such as wireless networks [6,7], IEEE 802.16 broadband wireless networks [8], IEEE 802.11 wireless LANs [9-11], and broadband wireless networks [12]. In addition, it have been used in network such as Ethernet passive optical network [13].The concept behind this is because, polling-based one performs better in heavy loads, Compared with another contention-based MAC protocol [14].

In this paper, we develop a simulation model in order to investigate the performance of distributed poling networks with exhaustive and K-limited services at certain conditions and compared that with the derived analytical equations. We aimed from this to make a basis for extending this protocol in more MAC wireless protocols. Their simulation results are simulated and compared with its analytical equations.
The rest of this paper is organized as follows. In section 2, the distributed poling network model is presented. In section 3, the simulation model is explained. In Section 4, the comparisons between the simulation and analytical results are compared. Finally, conclusions are provided in sections 5.

2. Polling Network Modeling

A polling network is a computer communications network that uses *polling* to control access to the network. Each node or station on the network is given exclusive access to the network in a predetermined order. Permission to transmit on the network is passed from station to station using a special message called a *poll*. Polling may be centralized (often called *hub polling*) or decentralized (*distributed*). In hub polling, the polling order is maintained by a single central station or *hub*. When a station finishes its turn transmitting, it sends a message to the hub, which then forwards the poll to the next station in the polling sequence. In a decentralized polling scheme, each station knows its successor in the polling sequence and sends the poll directly to that station. To simplify matters, we will assume a distributed polling scheme.
We divide time into alternating types of intervals: *polling intervals*, during which the poll is transferred between stations, and *transmission intervals*, during which the station with the poll transmits packets [1-5].
Polling networks come in three flavors: gated, exhaustive, and partially gated as follows [3]:

- *Exhaustive Policy*: If an exhaustive policy is in use, the server serves all packets at a queue that it finds upon arrival there, and the new packets that arrive after the server (while serving).
- *Gated Policy*: If a gated policy is in use, the server serves all packets at a queue that it finds upon arrival there, but no new packets that arrive after the server will be served.

- *Limited Policy*: If a limited policy is in use, the server serves a limited number of packets.

The polling network model for this simulation uses distributed polling. There are N queues indexed by *i*, 0 <= *i* <= N - 1. There is a single server that moves successively from queue i to queue (*i* + 1) mod N as shown in Fig. 1. The rate at witch packets arrive at a station for transmission on the network is the same for all stations. To simplify the model, we assume all packets except the poll are of the same length (constant packet lengths). A station that has permission to transmit (received the poll) transmits exhaustively, that is, until all massages in its input queue have been transmitted. This includes any packets that might arrive while the station has the poll and is transmitting other packets on the network. Packets arrive at each station according to a Poisson or Bernoulli process, independent of all other arrivals. The arrival rates at the station are identical.

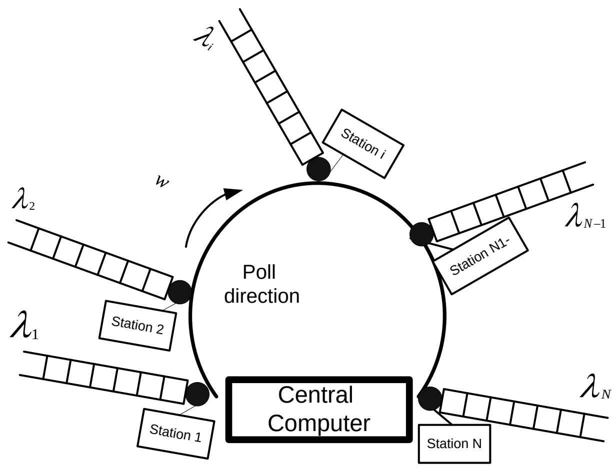


Fig. 1 Polling Network Model

The server has a switchover time to go from queue *i* to queue (*i* + 1) mod N with fixed delay. This is called walk time (*w*) and it is required to transfer the poll from one station to another and synchronize the station for transmission to the central computer. We assume that the distance between stations are equals, so that the walks time are equals. The total time required to poll each station and return to the starting station in the polling sequence is called cycle time and the average cycle time is $\bar{T_c}$ as shown in Fig. 2.

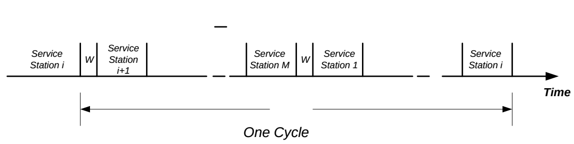


Fig. 2 Cycle Time of one Poll (M<=N)

The server model for polling system with exhaustive service is shown in Fig. 3.

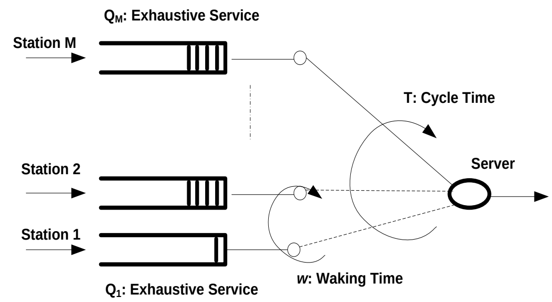


Fig. 3 Server Model for Polling Network

Each packet at queue *i* requires a service time that is related to the packet length, but since the packets size are equally likely then the service time is assumed to be constant time. The packets service times, walk time (*w*) and packets arrival processes are all mutually independent.

3. Simulation Process Model

As mentioned in section 2, the polling scheme with exhaustive service may be described with the following rules.

- The server polls the stations by sending a POLL packet.
- If the buffer queue of the polled station is not empty, then the station responds to each POLL packet by sending the data packets from its buffer queue exhaustively, that is, until all massages in its input queue have been transmitted. This includes any packets that might arrive while the station has the POLL and is transmitting other packets on the network
- If the buffer queue of the polled station is empty, then the station responds to each POLL packet by sending an empty (NULL) packet.
- The exchange ends after the queue buffer is empty.
- After the end of exchange, the server moves on to Poll next station.
- When all the stations have been visited, the sequence is cyclically repeated.

In case of K-limited service, the server will serve K packets at maximum at each visit. In this case the above rules will have the following modification:

- The exchange ends after K packets services finish or the queue buffer is empty whichever comes first.
- After exactly K packets services finish, the exchange ends regardless of the state of station's queue even there are new arrival packets during the period of the this services.

In this research we use C++ platform to simulate the studied system. As shown in Fig. 4 , we have three main types of events:

- *Packet Arrival event:* Packets arrive at queue *i* according to a Poisson process with the rate λ_i. where the inter-arrival time is exponentially distributed, and the Next Arrival Time (Packets) = Clock Time + (- 1/ λ_i ln(u)) Where 0 <= u <=1
- *Poll Arrival Event:* Next Arrival Time (Poll) = Clock Time + WalkTime(*w*) where *w* is small constant value of time. The sequence of Next station= (current station sequence +1) mod N
- *Completion Event:* Time of service Completion= Clock Time + Service Time (*s*) where s is constant value of time.

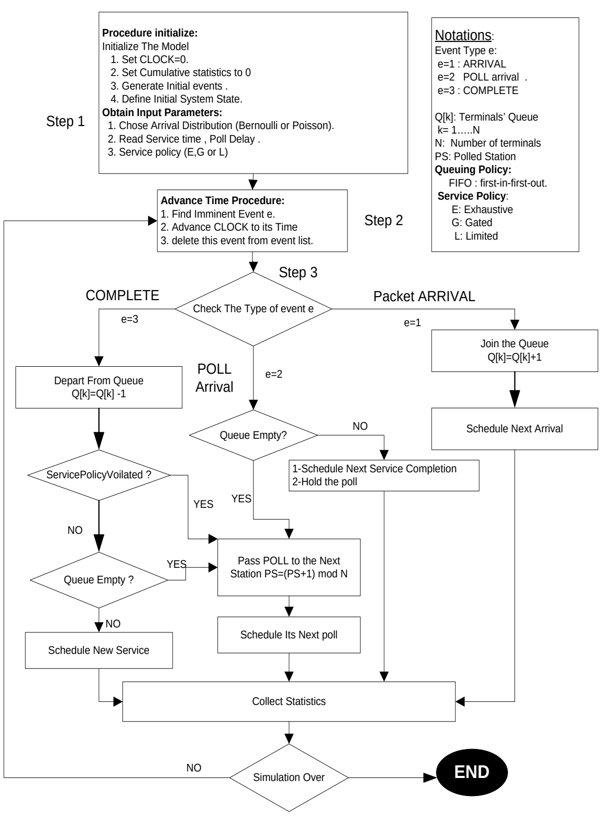


Fig. 4 : Simulation Model Flowchart

4. Performance Analysis

In this section, we present our performance study for the distributed exhaustive and K-limited poling network. The delay, throughput and performance of the K-limited services polling networks are compared with that of analytical equations under different network conditions.

4.1 Performance measurements and assumptions

In determining the performance, the following assumptions are made:

- Each station has the same Poisson arrival statistics with average arrival rate λ_i packets/second.
- The walk time (w) between stations is constant and the same for every consecutive station.
- The channel propagation times between stations are equal and are included in the walk time (w).
- Packet length distributions are the same for packets arriving at each station.

The analytical performance measures are derived in details in [2-5]. In this paper, the following selected parameters will be used to measure the performance:

- *Throughput* (S): the ratio of the total average arrival rate to the network to the total capacity of the network (both in packets/second).

$$S = \frac{M\lambda \overline{X}}{R} \tag{1}$$

Where X is the first moment of packet length and R is the channel capacity

- *Average cycle time* (T_c): the total time required to poll each station and return to the starting station in the polling sequence

$$T_c = \frac{Mw}{1-S} \tag{2}$$

- *Average waiting delay* (W): it is divided into to components
  - ➢ the waiting delay in the station buffer while other station are being served.
  - ➢ the waiting delay in the station buffer while the particular station is being served.

$$W = \frac{Mw(1-S/M)}{2(1-S)} + \frac{S\,\overline{X}}{2R(1-S)} \tag{3}$$

Assuming constant packet lengths

- *Average number of packets stored in a station buffer* (N) : it is divided into two parts
  - ➢ Those packets that arrive while a station inactive
  - ➢ Those packets that arrive while the station is being served.

$$N = \lambda W = \frac{Mw\lambda(1-S/M)}{2(1-S)} + \frac{S\lambda \overline{X}}{2R(1-S)} \qquad (4)$$

- *Average transfer delay (T)*: the total average time between packet arrival at station and its delivery to the central computer.

$$T = \frac{\overline{X}}{R} + \tau + W \qquad (5)$$

where τ is the end-to-end prorogation delay for the bus.

In the simulation, we assume that: (1) one server and M stations; (2) the length of time slot is 10 ms; (3) the Switchover time when the server poll the stations from the (i)th to the (i+1)th is one slot; (4) the arriving process of the packets is according to Poisson arrival process. (Arrival rate =0.4 packets/sec , Tc=0.2 sec, w=0.05 msec)

4.2 Comparing analytical and simulation results

In this section we compare the results of our simulation model with the analytical equations described in section 4.1. The comparison of simulation results and numerical results obtained from solution of the analytical model in terms of average delay and average number of stored in station buffer are depicted in Fig. 5 and Fig. 6. From these figures, the two results are almost in total agreements. The differences between the analytical and simulation values came from the approximation done on the analytical equations.

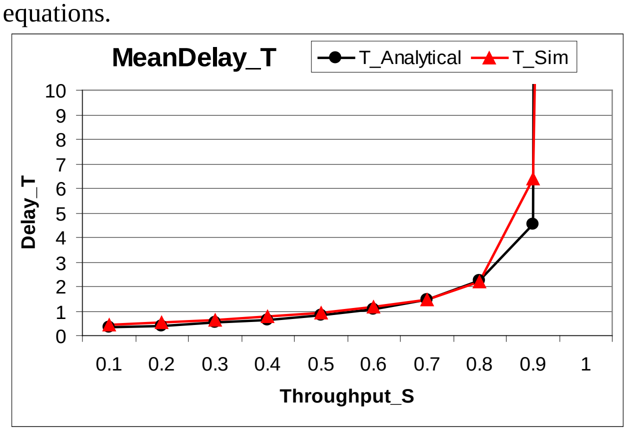


Fig. 5 Mean total delay versus Throughput

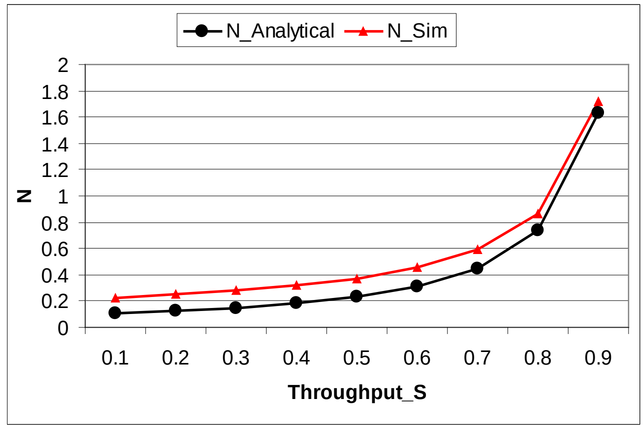


Fig. 6 Average number stored in station buffer versus Throughput

4.3 System performance at different conditions

Effects of increasing the total number of stations: the effect of increasing the number of stations on the performance of polling networks is shown in Fig. 7 and Fig. 8. As we expected, increasing the number of stations will increase the cycle time, and thereby increasing the waiting time of a packet. Consequently, increasing the waiting time will increase the average number of stored packets at station.

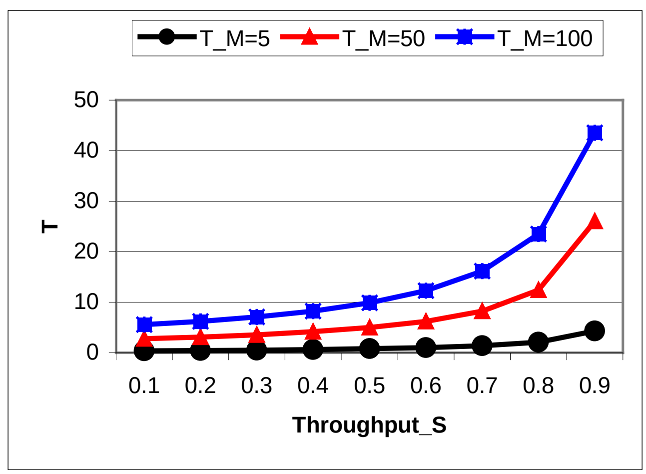


Fig. 7 Mean delay versus throughput at different number of stations

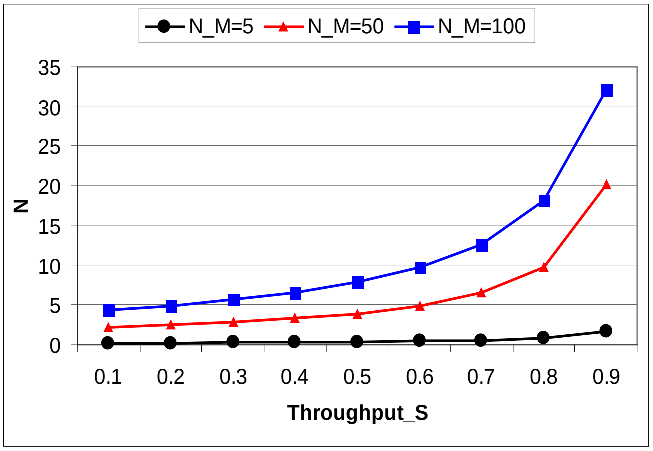


Fig. 8 average number of packets stored at a station versus throughput at different number of stations

Effects of increasing the walking time: the effect of increasing the walking time on the performance of polling networks is shown in Fig. 9 and Fig. 10. As we expected, increasing the walking time will increase the cycle time, and thereby increasing the waiting time of a packet. Consequently, increasing the waiting time will increase the average number of stored packets at station.

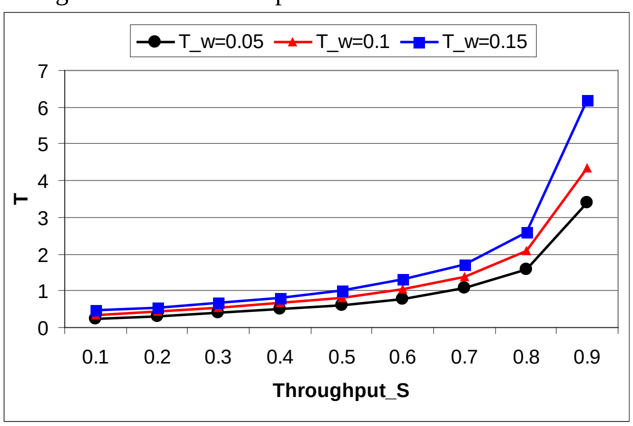


Fig. 9 Mean total delay versus throughput at different values of walking time

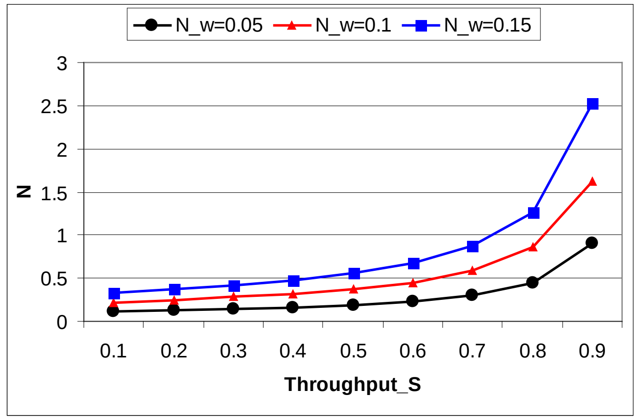


Fig. 10 average number of packets stored at a station versus throughput at different values of walking time

Average Transfer delay and average cycle time: The total time required to poll each station and return to the starting station in the polling sequence is called cycle time. Fig. 11 shows T and Tc/2 versus S with different walking times. This figure verifies that Tc/2 is an excellent approximation for T if Mw is sufficiently large.

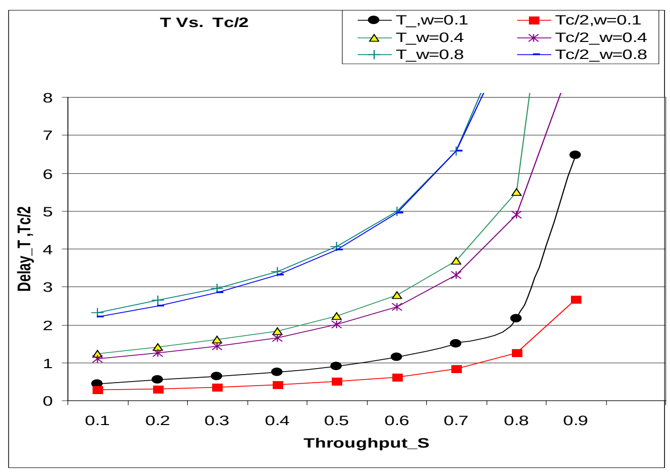


Fig. 11 average transfer delay and Tc/2 versus throughput at different values of walking time

4.4 Exhaustive service versus K-limited service

It is well-known that in a polling system, exhaustive service has the highest efficiency, 1-limited service has the lowest efficiency. But this issue is correct in certain conditions. In this section the performance of polling networks using exhaustive service policy versus K-limited service policy in terms of average transfer delay are compared. Using the parameters: Arrival rate=0.4, M=5,w=0.1 ms, we show how the mean transfer delay vary with parameter K . As shown in Fig. 12, the performances of two service policies are the same for the case of low and medium traffic load. For high traffic load, the transfer delay of K-limited service policy is less than exhaustive. However, when the parameter K is large enough, it functions nearly like the exhaustive service.

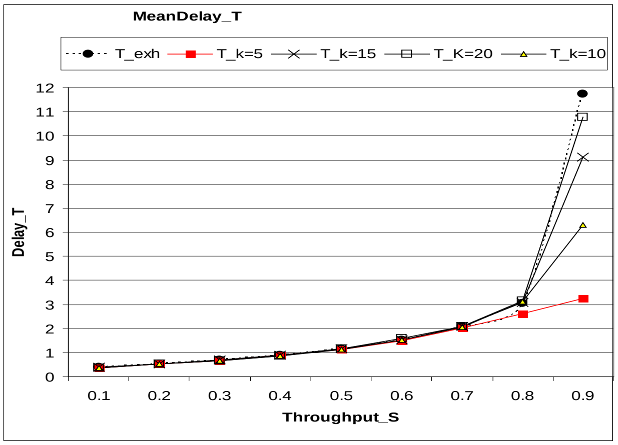


Fig. 12 Transfer delay versus throughputs for Exhaustive and K-limited

5. Conclusions

In this paper we have modeled the performance of distributed poling networks under exhaustive and K-limited services at certain conditions. From examination of the equations and plots given in previous sections, we can make some comments on the design of polling networks. In fact, it is important to keep average of transfer delay and average of stored packets per station as small as possible. So, to do that, the number of stations should be restricted when the walking time is large and, correspondingly, if large number of stations is required, we should try our best to keep the walking time as small as possible.

Also, both average of transfer delay and average of stored packets per station increase with throughput. Thus, as throughput increases, the performance in terms of delay will decrease, and more storage is required at the network stations.

In case of low and medium traffic load environments, the exhaustive and k-limited service disciplines are almost the same. However as throughput increases, k-limited service disciplines will provide more fairness than exhaustive. But as K becomes large enough, the performance K-limited discipline approaches the exhaustive.

References

[1] H. Takagi, "On the analysis of a symmetric polling system with single-message buffers," Performance Evaluation, Vol. 5, 1985, pp. 149-157.

[2] H. Takagi, "Analysis of Polling Systems," The MIT Press, 1986.

[3] J. L. Hammond and P. J. O'Reilly, "Performance Analysis of Local Computer Networks", Addison-wesley,1986.

[4] V.M. Vishnevskii and O.V. Semenova, "Mathematical methods to study the polling systems", Automation and Remote Control, vol. 67 , no. 2, 2006, pp. 173-220.

[5] Hong Wei Ding, Dong Feng Zhao and Yi Fan Zhao, "Queue-Length Analysis of Continuous-Time Polling System with Vacations Using M-Gated Services", Applied Mechanics and Materials , vol 20 - 23, January, 2010, pp. 427-431.

[6] Yan Li Guangxi Zhu ," Performance Analysis of Three Gated Service Polling Scheme on Wireless Network ", 2nd International Symposium on Wireless Pervasive Computing, 2007.

[7] V. Vishnevsky and O. Semenova, "Adaptive Dynamical Polling in Wireless Networks", Cybernetics and Information Technologies, vol 8, no 1, 2008, pp. 3-11.

[8] Zsolt Saffer and Miklós Telek, " Analysis of Globally Gated Markovian Limited Cyclic Polling Model and its Application to IEEE 802.16 Network", Proceedings of the 5th International Conference on Queuing Theory and Network Applications, Beijing, China, on July 24-26, 2010.

[9] R. Y. Law, V. C. Leung and H. C. Chan, "Polling-based protocols for packet voice transport over IEEE 802.11wireless local area networks", IEEE Wireless Communications, Feb. 2006, pp.22-29.

[10] Tao Li, D. Logothetis, M. Veeraraghavan, "Analysis of a polling system for telephony traffic with application to wireless LANs". IEEE Transactions on Wireless Communications, vol.5, no.6, June 2006, pp.1284 – 1293.

[11] V. M. Vishnevsky, A. I. Lyakhov and N. N. Guzakov, "An Adaptive Polling Strategy for IEEE 802.11 PCF," In Proc. of 7th Int. Symp. on Wireless Personal Multimedia Communications (WPMC'04). Vol. 1. Abano Terme, Italy, September 12-15, 2004, pp. 87-91.

[12] V. M. Vishnevsky, D. V. Lakontsev, O. V. Semenova and S. A. Shpilev, "Polling Model for Investigation of the Broadband Wireless Networks," Automation and Remote Control, Vol. 67, no 12, 2006, pp. 123-135.

[13] C. Geun Park, B. Kim and D. Hwan Han, "Queueing Analysis of Gated Polling System for Dynamic Bandwidth Allocation Scheme in an EPON", J. Appl. Math. & Computing Vol. 16, No. 1 - 2, 2004, pp. 469 – 481.

[14] R. D. Vander Mei, "Towards a Unifying Theory on Branching-Type Polling Systems in Heavy Traffic," Queueimg Systems, Vol. 57, No 1, 2007, pp. 29-46.

Comparison of the elastic coefficients and Calculation Models of the Mechanical Behavior one- Dimensional Composites

Saleh Alsubari[1], Hassan Chaffoui[2]

[1]**Department Mathematics, UFR-MASI,** ***Faculty of Science and Technology, Hassan II University, Mohammedia/20650, Morocco .***

[2]**Department physics, UFR-MASI,** ***Faculty of Science and Technology, Hassan II University, Mohammedia/20650, Morocco .***

Abstract

In this paper, we present the mechanical models that are devoted to the elastic properties of one-dimensional composite. We have compared the equivalent coefficients of one-dimensional composite, resulting from different models. The validation of the results was made through effective experiments on a one-dimensional composite consisting of fibers of alumina and a matrix of aluminum. This study allows us to better assess the rigidity of composite structures, and the results of calculation of the mechanical behavior, resulting from each model. It appears that the finite element model is the best suited to the approach of a refined conception. For more insurance, we have chosen to make our calculations by finite element in the three-dimensional case, using the technique of homogenization by asymptotic development.

Keywords: *homogenization method, finite elements method, material composites*

1. Introduction

The composite materials are used in the majority of the high mechanical performances structures. Nowadays their presence in all the technical fields, justifies this importance. In this work we present the mechanical models are devoted to the elastic properties of composites. There is a basic and practical reason, for that to most of the applications structural of composite materials, instantaneous resistance of these materials is not the decisive criterion for the dimensioning, which must face-as well as behavior with the impact, of deterioration or ageing under static load. The other reason is that, while the elastic properties can be the subject of rather correct forecasts, the characteristics of the instantaneous ruptures and a fortiori the characteristics of the differed ruptures, behavior to the shock, deterioration or all characteristics strongly dependent on qualitative aspects, like the conditions of manufacture and environment, cannot be the subject of sufficiently general quantitative precise details. The objective of this work is initially to locate the state of progress the work of the mechanical modeling of the composite structures and to determine the models available with their possibilities, their limitations [4, 11, 18]. It provides a catalogue of concrete results, either resulting from the bibliography or calculated by our method starting from the theoretical concepts which make it possible to define models of increasing complexity which will be useful, bases for the modeling of other types of composites.

2. Principle of the equivalent medium

It is completely unimaginable to consider a direct calculation which requires a discretization of the whole composite structure. Generally, the shape of the composite, the geometric complexity of its basic constituents and the tedious nature of the mesh by finite element do not allow taking consideration of the composite geometry. To circumvent these difficulties, many researchers have proposed models that appeal to rational micromechanical notions. Thus, we can predict the mechanical performance of composites and their ultimate resistance. This is done through the nature and the provision of basic components. In this case, the composite area is considered as homogeneous but anisotropic: this is the principle of equivalent homogeneous area. This principle is based on replacing a real, highly heterogeneous composite material by equivalent homogeneous material. The approach is based on the notion of the details macroscopic behavior, neglecting the influence of the microscopic or those inherent in the composition of the composite. To conduct this analysis of behavior, it is necessary to define an elementary sample, represent a volume of the structure and its heterogeneity, on a microscopic scale. This representative volume plays the classic role of continuous mechanics of volume element. This idea involves a concept of statistics average, in which the actual constitution of the material is idealized in considering a continuous medium. Once the model of continuity is admitted, the concept of homogeneity is deducted. However, as close as the disparity between the different components is too large or that geometric continuity of heterogeneity is not acquired, it will be required have homogenization models of percolation or the methods of self consistent [4, 8, 16]. The mechanical characteristics of the material and analysis properties are defined from a basic of elementary volume and representative, small enough to account for microscopic composite to study, but large enough as to the fiber diameters and their separating distances. This representative volume sets the state of local stress, induced by external displacement of loads and boundary conditions [19]. This state represents requires of the volume, placed inside the real composite. The solution of the problem is made through the analysis of the stresses and deformations in the elementary volume, to establish relations of elastic behavior. Generally, an exact solution requires a complex procedure which is not always applicable without introducing simplifying assumptions. This is what limits the scope of this type of analysis prediction.

3. Empirical methods

We observe, in the literature, two types of models interested in the prediction of the composite materials elastic behavior. The first models are global and allow estimating the mechanical performance on a global scale. These models are limited to the determination of equivalent elasticity coefficients. In contrast, the second family of models is local and allows, not only, to determine the equivalent elasticity coefficients, but it also allows to go back on a microscopic scale to locate the possible interactions that can exist between the fibers and the matrix and identify zone of damage. We take into account the work of Hashin-Shtrikman [9,10], [13, 14] Hill, Hashin-Rosen [11], Christensen [5] and Hashin [12] who have token an approach to the fibrous composite case, in which fibers can have sections of circular, elliptical or square shapes and can be distributed randomly or following a hexagonal and identical diameters. The developed models are based on a variational method using the elastic strain energy theorems, in a representative elementary volume of the composite. Through other works in this framework, there are many semi-empirical methods in which weights coefficients, are introduced to adapt to a restricted application environment. In this context, it was noted the models developed by Tsai [21] [17, 18] which are related to the composite Puck reinforced one-directional fiber of square or circular and distributed sections following repetitive rectangular or hexagonal shapes. To account for potential irregularities distribution or alignment of the fibers, Tsai introduced a factor of contiguity of the fibers. However the empirical determination of these factors limits the scope of certain models in calculation of forecast. The distinction between the models is located in the strict framework of their assumptions or on their adjustment more and less empirical by correcting coefficients, so that results are consistent with the experience.

4. Homogenization of periodic structures composites.

The 1970s knew a new orientation of the homogenization technique of composite materials. The first work generated mathematical progress on the side. In particular, the use of functional analysis for the formulation of the mechanics heterogeneous environments problems. The work of Sanchez - Palencia [20], Duvaut [6], chaffoui [7] showed that the theory of homogenization can provide an excellent mathematical response for the determination of the mechanical properties of composites with periodic structure [1, 2, 3]. By supposing that the period is small compared of dimensions studied area, going to the extremities allows to calculate macroscopic and microscopic sizes exactly which are often inaccessible to the experiment. We noted that this technique is well adapted to calculations by finite elements.

5. Applications

It is about comparing the equivalent coefficients of a one-dimensional composite, resulting from the various models. The validation of the results was made through experimental tests carried out on a one-dimensional constituted an aluminum of fibers alumina of matrix. The rate of impregnation out of matrix is 50 %. The fibers are parallel to the longitudinal direction of the E_3 module and their sections are circular of ray $R = 5{,}75\ \mu m$. Mechanical characteristics of the components are:

- For the matrix: $E_m = 70\ Gap$, $\nu_m = 0{,}33$
- For the fibre: $E_f = 380\ GPa$, $\nu_f = 0{,}25$.

For more insurance, we have chosen to make our calculations of finite elements in the three-dimensional case, by using the technique of homogenization by asymptotic development, where the basic period form parallelepiped (Fig. 5. c), and the two-dimensional case which form the base period a square (Fig. 5. b). However geometrical symmetries compared to y_1 and $y_{2,}$ lead to orthotropic isotropic and make it possible to reduce the field of the useful mesh for calculation, with the quarter of the period.(Fig. 6, a. 2D problem) et (Fig. 6. b. 3D problem).

Table 1 presents a comparison of the elastic modules resulting from the tests by ultrasonic evaluation non destructive (test 1) and from the mechanical tests (test 2), with the results, resulting from the empirical formulations, on the other hand, and with calculations of homogenization by finite elements We constant in the analysis of these results that the model calculations of homogenization by finite elements is the best performing, with compared the empirical methods.

6. Influence basic materials.

To determine the influence of the mechanical characteristics of the materials that constitute the composite, we varied the module of elasticity of fiber (fig.1-2) and of the matrix (fig.3-4). The variations of the elastic modules show the same observations mentioned above with knowing, the influence of coefficient of adjacency in the formulated of (Tsai) and the behavior of the results of the calculations by finite elements compared to the empirical formulae. Nevertheless, we note that the slopes of the curves are significantly different and depend on the module of the component that has been varied (E_f for fiber or E_m for the matrix).

Table 1: Comparison of the elastic coefficients.

	$E_3 (GPa)$	$E_1 = E_2 (GPa)$	$G_{12} (GPa)$	$G_{13} = G_{23} (GPa)$	ν_{12}	$\nu_{13} = \nu_{23}$
Test 1	233,3	141,9	51,8	54,6	0,368	0,291
Test 2	223,0	138,0	- - - - - - -	52,0	- - - - - - -	0,310
Puck	225,0	181,8	- - - - -	73,9	- - - - -	0,290
Hashin	225,0	125,4	- - - - - - -	54,9	- - - - - - -	0,290
Tsai	213,8	154,2	- - - - - - -	72,9	- - - - - - -	O,278
Withney	225,1	154,2	- - - - - - -	54,9	- - - - - - -	0,291
F. E. M.	223,5	140,5	50,1	55,1	0,371	0,286

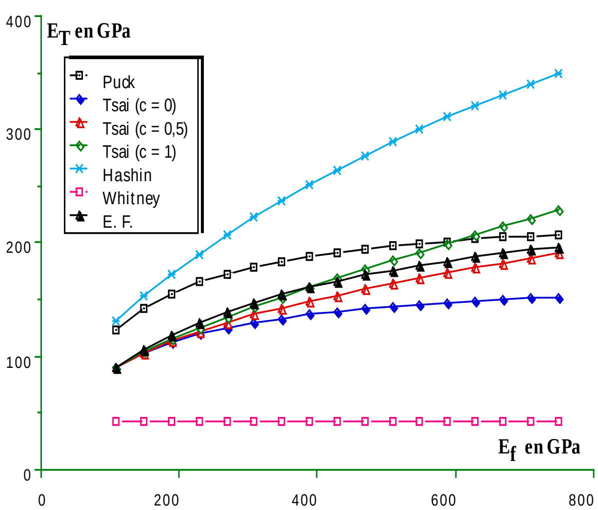


Fig. 1 *Variation of the module* E_L *in function* E_f

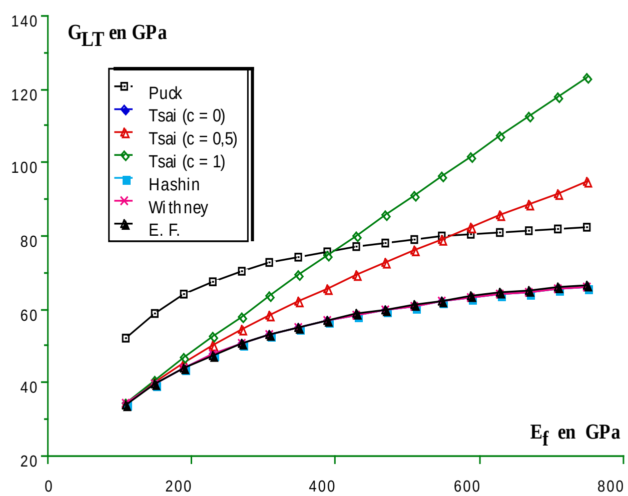


Fig. 2 *Variation of the module* G_{LT} *in function* E_f

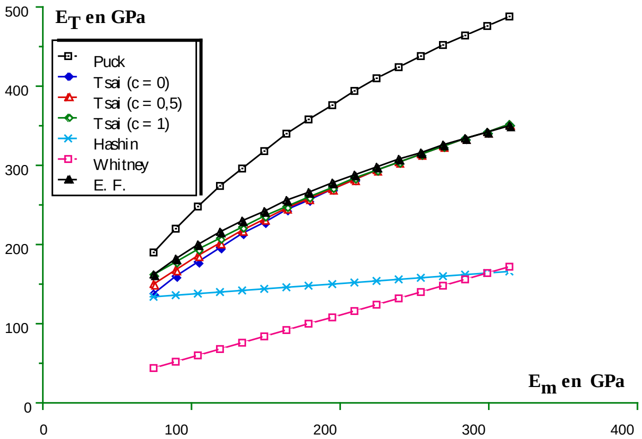


Fig. 3. *Variation of the module* E_T *in function* E_m

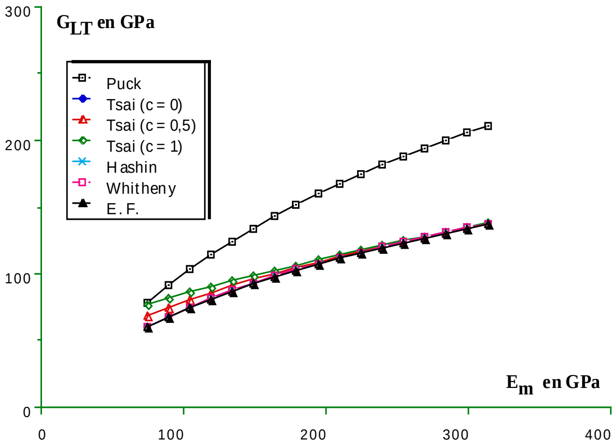


Fig. 4. *Variation of the module* G_{LT} *in function* E_m

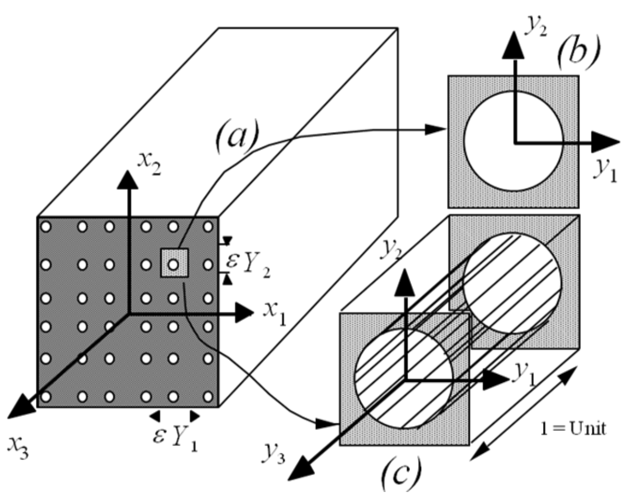


Fig.5 Geometry composite of one-dimensional and definition of the period

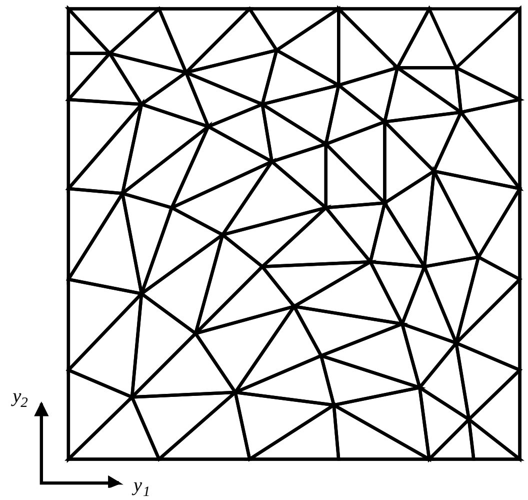


Fig. 6. a Mesh quarter of the period (2D problem)

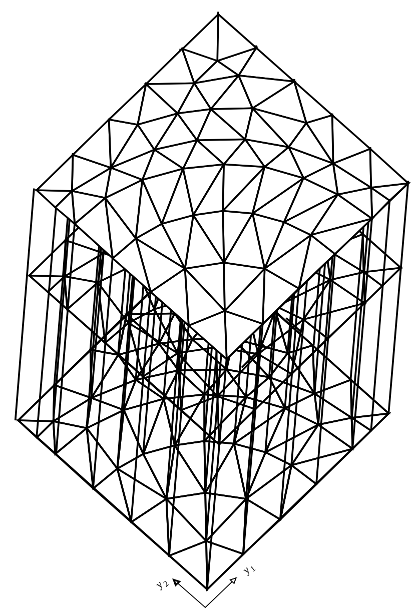


Fig. 6. b Mesh quarter of the period (3D problem)

7. Conclusions

This study allows to better assessing the composite rigidity of the structures, and the results calculation of the mechanical behavior of each model forecasting. It appears that the finite element model the best adapted to the approach of a refined conception. We constant in the analysis of these results that the model of calculations of homogenization by finite elements is the best, compared with the empirical methods.

References

[1] S Alsubari, H. Chaffoui.. Homogénéisation du comportment mécanique d'une plaque composite, 9th international conference on numerical analysis and optimization, FST Mohammedia, 17-19 december, 2008, pp.116-121.

[2] H. Chaffoui, M. EL Hammouti; A. Yeznasni, D. Play, Homogénéisation des plaques sandwiches, non symétriques par rapport au plan moyen, Rev. des composites et des Mat. Avancés, V. 9, N° 2, 2000, pp. 219-238.

[3] S. Alsubari, H. Chaffoui, Etude du comportement mécanique d'une plaque composite à renforts textiles « Homogénéisation par sous domaine », Journée de Mécanique des Structures, JMS2008, 26 Novembre 2008, FST Settat, pp 1-6

[4] M. Brun, O. Lopez-Pamies, P. Ponte Castaneda. Homogenization estimates for fiber reinforced elastomers with periodic microstructures. International Journal of Solids and Structures 44, 2007,pp. 5953-5979.

[5] R. Mr. Christensen, mechanics of composite materials, Ed. John Wiley & Sounds, 1979.

[6] G. Duvaut, Analyzes functional and mechanics of the continuous mediums:"request being studied of composite materials elastic with structures periodical - Homogenization. Theo. and app mech. W.T Koiter, Ed. North-Holland Publishing Company, 1976, pp. 119-132.

[7] H. Chaffoui., D. Play, Etude du comportement mécanique pour la conception des structures textiles : "Application aux bandes transporteuse", Rev. Euro. des Eléments Finis, V. 7, N° 6, 1998, pp. 737-754.

[8] George Chatzigeorgiou , Nicolas Charalambakis , Francois Murat. Homogenization problems of a hollow cylinder made of elastic materials with discontinuous properties. International Journal of solids and structures 45, 2008, pp. 5165-5180.

[9] Harald Berger, Sreedhar , Kari, Ulrich Gabbert An analytical and numerical approach for calculating effective material coefficients of piezoelectric fibre composites . international journal of solids and structure. 42, 2005 , pp. 5692-5714.

[10] Z Hashin and S. Shtrikman, A Vartiational Approach with the theory of the elastic behavior of polycrystals, J mech. and Phy. of solids, 10, 1962, pp. 343-352.

[11] Z Hashin and B W Rosen, The Elastic module reinforced material fibers, ASME J of APP mech., 31, 1964, pp 223-232.

[12] Z Hashin, on the elastic reinforced fibre material behavior of arbitrary geometry of transverse plan, J mech. and Phy. of solids, 13, 1965, pp. 119-126.

[13] R. Hill, elastic properties of Reiforced solids: "Certain theoretical principles", J mech.and Phy. of solids, Vol11, 1963, pp 357-372.

[14] R. Hill, theory of the mechanic material fiber estengthened properties:"Elastic behavior", J mech. and Phy. of solids, 12, pp. 1964, 199-212.

[15] Marcin Kaminski .Homogenized properties of periodic n-component composites International Journal of Engineering Science 38 , 2000 , pp. 405-427

[16] T.M.H. Nguyen , Blond , A. Gasser , T. Pritl, Mechanical homogenisation of masonry wall without mortar . European Journal of Mechanics A/solids 28 ,2009, pp. 535-544.

[17] A. Puck, Grundlagen DER spannungs und veromungs to analyze, Dopl. ING. Kunstoffe, C 57, 1967 .

[18] A. Puck, Zur Beanspruchung und performing von GFK mehrschichtenverbund bauelementen, Kunstoffe, 1967.

[19] S. Sakata , F. Ashida , T. Kojima , M. Zako. Three-dimensional stochastic analysis using a perturbation based homogenization method for elastic properties of composite material considering microscopic uncertainty International Journal of Solids and Structures 45, 2008 , pp.894–907

[20] E .Sanchez-Palencia, Behavior local and macroscopic of a type of physical environments heterogeneous, int. J Engng. Sci., 12, 1974, pp 331-351.

[21] S. W Tsai and h. T Hahn, Introduction to composite materials, Technomic Pub Co, Westport, Conn., 1980.

S. Alsubari is a PhD student since (2008) in the university of Hassan II – Mohammedia - Casablanca, Faculty of Science and Technology, Morocco. His research interests include method homogeneisation by asymptotic development for calculation of material coefficients of plates composites. Mechanical modeling and scientific calculation by analytical numerical, he has published two papers for international journals and more than five conferences paper.

H. Chaffoui is a Professor at the university of Hassan II - Mohammedia - Casablanca, Faculty of Science and Technology, Morocco. His research interests includes contribution to the modeling and optimization of composites structures. He has published numerous papers for international journals and/or conferences. Head of Computing Center (1995 to 2002), director of the mechanic laboratory and head Department of Physics (2004-2009).

Subword Unit Concatenation for Malay Speech Synthesis

Sabrina Tiun[1], Rosni Abdullah[2] and Enyakong Tang[3]

**[1] Pusat pengajian Sains Komputer, Universiti Sains Malaysia
Minden, 11800, Malaysia**

**[2] Pusat Pengajian Sains Komputer, Universiti Sains Malaysia
Minden, 11800, Malaysia**

**[3] Multimedia University,
Cyberjaya, 63100 , Malaysia**

Abstract

In this paper, we propose an idea to use subword synthesis units to generate novel word sound of Malay language. The subword synthesis units were carefully created based on Malay phoneme adjacency analysis. The phoneme adjacency analysis was used to find out which adjacent phonemes that would not introduce audible distortion if the phonemes were concatenated. Besides phoneme adjacency factor, we also include prosodic features in selecting the subword synthesis unit. Thus, by considering both phoneme adjacency and prosodic feature in selecting the subword units, a smooth sound of novel word can be obtained.

Keywords: *Subword Concatenation, Phoneme Adjacency Analysis, Concatenative synthesis, Speech Synthesis, Malay language.*

1. Introduction

The work presented in this paper is part of work done for Malay speech synthesizer pilot system called UTMK-MSS system. UTMK-MSS system was developed based on the approach that prioritizes naturalness before flexibility aspect. The reason why naturalness prioritized before flexibility is because most Malay Speech synthesizer systems are flexible but their naturalness quality needs to be improved.

Naturalness quality can be achieved in speech synthesizer system if both segmental quality and prosodic quality of the voice systems are high. Synthesis voice with high prosodic quality is when there are no mismatched of duration, pitch and intensity between the joined speech units. In the other hand, high segmental quality is when the joined units do not have audible discontinuity that is caused by spectral mismatched [1].

Thus, in order for the UTMK-MSS system to speak naturally, the system need to choose speech units that once concatenated will introduce less or none audible distortion caused by the prosodic and segmental mismatched. In UTMK-MSS system, prosodic mismatches are avoided by selecting the most appropriate speech units using NLP approach (see Fig.1). Whereas to avoid segmental mismatched, only larger chunks are considered: Word and subword.

Subword unit is the synthesis unit that will be used to generate novel word unit. However, the concatenation of this unit potentially can degrade the naturalness of generated speech that caused by the segmental (spectral) mismatched. Thus, we avoid the possible of subword segmental mismatched by only synthesizing novel word from the list of subword unit that do not cause audible distortion.

These subword units were created based on the result of phonemes adjacency analysis. It was an analysis that listed out which Malay phonemes can be adjacent with other Malay phonemes without audible discontinuities. We gathered the findings of other researchers about non-audible phonemes concatenation and used it as a guide for non-audible Malay phonemes adjacency, which then used to build the subword unit lookup as in Fig. 1.

The phoneme adjacency analysis procedure and result, and also how we build the subword lookup are being described in section 2. In section 3, we explain how novel word sounds are created using subword concatenation. Afterwards, the evaluation on the subword concatenation will be in section 4 and 5. Finally in section 6, we end our paper with conclusion and discussion.

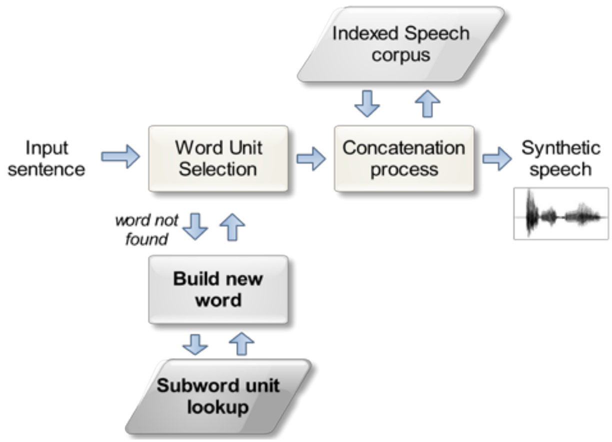


Fig. 1: The overall of UTMK-MSS process.

2. Malay Phoneme Adjacency Analysis

The subword lookup (see Fig. 1) is built based on the result of phonemes adjacency analysis. Building the subword unit lookup required two steps; (1) to find out what phonemes adjacent to other phonemes that will not introduce audible distortion, and (2) to build a list of subwords from the existing word synthesis units based on the list of phonemes in step (1).

2.1 Phoneme Adjacency Analysis

Table 1: Malay consonants and vowels.

Consonants	***Stops (7):*** /g/, /d/, /b/, /p/, /t/, /k/, /ʔ/ ***Affricates (2)*** : /ʤ/, /ʧ/ ***Fricatives (9):*** /s/,/f/,/v/,/z/,/ʃ/,/θ/, /ð/, /x/, /ɣ/, /h/ ***Liquid (2):*** /r/, /l/ ***Nasal (4):*** /m/, /n/, /ɲ/,/ŋ/ ***Semi-vowel/glides (3):*** /w/,/j/
Vowels	/a/, /e/, /i/, /o/, /u/, /ə/
Dipthongs	/ai/, /ao/, /au/

The phonemes in Malay consists of 27 consonants, 6 vowels and 3 dipthongs [2][3] (see Table 1).

In [4]'s study, the stop, fricatives and affricates would not introduce discontinuities if spliced is happened between these consonants. Therefore, we perceptually tested the joined of the inter-categories and intra-categories of those consonants for Malay. Two samples of phonemes were took out from each category and tested for their combination (see Table 2 and Table 3). The perceptual test was done by concatenating subwords from bisyllabic or polysyllabic word and had been recorded in isolated manners.

In perceptual test 1 (Table 2), we purposely chose /h/ to represent fricatives consonant. This is because /h/ is identified as glide in [3], but [5] and [2] stated that /h/ belongs to fricative consonants. Since we are interested in fricative consonants, therefore /h/ has to be considered in the test.

Table 2: Perceptual test I.

	Stop	***Fricative***	***Affricate***
Stop	/p/ and k/	/k/ and /z/	/k/ and /ʧ /
Fricative	/ɣ/ and /t/	/f/ and /z/ /ɣ/ and /s/ /h/ and /z/	/h/ and /ʧ/ /f/ and /ʧ/
Affricate	/ʤ/ and /t/	/ʤ/ and /s/ / ʤ / and /h/	/ʧ/ and /ʧ/ /ʤ/ and /ʧ/

Table 3: Perceptual test II.

	Alveolar	***Dental***	***Palatal***
Alveolar	/n/ and /z/ /s/ and /z/ /t/ and /z/ /l/ and /z/	/k/ and / ð /	/z/ and /j /
Dental	/ð/ and /t/	/ð / and /s/	/ð / and /j/
Palatal	/j/ and /z/	/ ð/ and /j/	/ɲ/ and /j/

Based on the perceptual test 1, the joint of phonemes either within the inner or between the outer categories did not produce audible discontinuities. As for affricate /ʧ/, the result obtained was similar to [4]'s finding, and therefore, the suggestion made by [5] was disregarded. In the other hand, perceptual test on /h/ turned out to be as smooth as the other tested fricatives.

Tests were also carried out to find whether it is also true for consonants from alveolar, dental and palatal, if were replaced with each other, will not produce audible distortion. This information is useful if one wants to obtain subword ended with vowel, but, unable to find the matched adjacent consonant of the next subword.

For an example, if the word to be synthesised is *diazan (*'uttering the moslem call prayer'*)*, but, the speech corpus only contains the words *azan* ('the moslem call for prayer') and *dialas* ('putting a cover'). The subword unit /*dia*/ can be extracted from the word *dialas* ('putting a cover') and concatenates it with the subword unit /zan/ from the word '*azan*' ('the moslem call for prayer').

In Malay, the consonants under these three categories are: (1) alveolar - /t/,/d/,/s/, /z/, /n/, /l/, (2) dental - /θ/, /ð/ and (3) palatal - /ɲ/, /j/. Again, two samples of phonemes from each category are chosen for the perceptual test. The joints of the inter-categories and intra-categories of those consonants were tested out perceptually (see Table 3). The combination of all of the phonemes sounded natural except a 'click' happened between the joint points. Since by using fading-out and fading-in effects, the 'click' sound can be removed or reduced, the phonemes of alveolar, dental and palatal are included in the subword lookup list.

2.2 Building the Subword Lookup

Using the phonemes adjacency result in section 2.1, we built a list of subword that will not introduce audible discontinuities when the subwords are joined together. Since the subwords units will be used to synthesize new word, therefore, the list of subwords unit must be created from the existed words in the speech corpus.

Creating the subword unit lookup was done according to these steps: Firstly, all words (strings) from the speech corpus were extracted out. Secondly, automatically using [7]'s program, letters were converted into phonemes and syllabified into phonemized words. Finally, based on the result of phoneme adjacency analysis, syllables were united with their adjacent syllables, if their adjacent phonemes were not any of the tested consonants (see Table 2 and Table 3). This is how the subword unit lookup was built. Fig. 2 presents the overall process of building the subword unit lookup based on the phoneme adjacency analysis.

The words units in the speech corpus are annotated with Part-of-Speech (POS), types of phrase break and prominent marks (or none if the word is not prominent). Therefore, the subwords will also inherit the same prosodic features from word which the subwords were obtained.

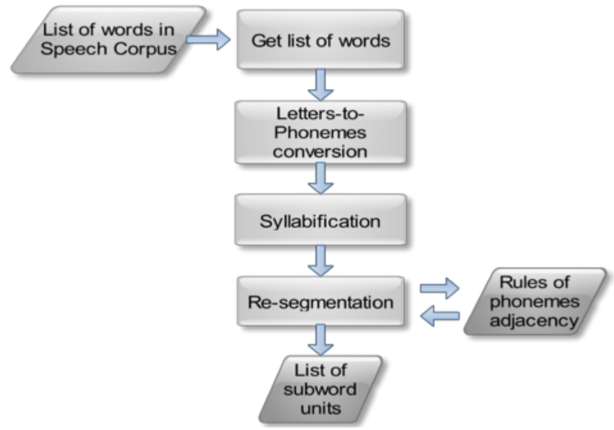


Fig. 2 The overview process of subword unit lookup.

Table 4 shows the example of indexed subword unit lookup. In Table 4, subword string that contains numerical of '0' or '1', indicating the position of the syllables of the subword unit within the word. Numerical '0' signifies that the subword located at the first syllable of a word, and '1' indicates the subword unit is a final syllable(s). The subword without numerical '0' and '1' means that the subword is a middle position syllable(s). For example, in Table 4, the subword /*sil1*/ is from the final syllable of the word *komunikasi* ('communication'), inherits noun as POS, with the starting time of 7.54 ms and ended at 7.98 of the wave file (with node ID 2879).

Table 4: The example of subword unit lookup entries.

Subword	***POS & Prosodic***	***Start Time***	***End Time***	***String***	***ID***
0komu	N,1	7.07	7.26	komunikasi	2879
nika	N,1	7.26	7.54	komunikasi	2879
si1	N,1	7.54	7.98	komunikasi	2879

3. Subword Units Concatenation

The subword concatenation is a concatenation of synthesis unit which is smaller than word units, or a process known as 'build new word' in Fig.1.

As mentioned before, in order not to degrade a synthesized utterance, both segmental and prosodic must not have any mismatches. Segmental mismatch will be avoided by joining subword units that will not introduce audible distortion, whereas for prosodic mismatch, we avoid it by selecting subword based on the subword prosodic features and its position in a sentence.

As pointed out by [6] in their syllable re-combining rules that the position of syllable within a word is not important except when the word located at the start and the end of a phrase. The rules can be also applied to subword. However, we want to be very particular on the location of subword not only at the phrase level but also at the word level, especially on the first and last position of subword segment in a word. Subword located at the first and the last position in a word must be replaced with the same subword with similar positions in order to avoid audible distortion.

The subword matching steps are similar to the steps of building the subword unit lookup. In the subword matching, the first step is to convert the letters in the unmatched word into phonemes. Then, the phonemized words are segmented into syllables using the same syllabification rules of [7]'s program. Symbol of '0' and '1' are appended to the first syllable and the last syllable respectively. Using the phoneme adjacency rules in Table 2 and Table 3, the syllables segment in the word are re-segmented.

A program's output below shows the output of every step to create subwords strings of the unmatched word for subword matching.

```
UNMATCHED WORD    : menjelaskan
PHONEMISED WORD   : m e n ʤ ə l a s k a n
SYLLABIFICATION   : men.ʤə.las.kan
SYLLABLE POSITION : 0men.ʤə.las.kan1
RE-SEGMENTATION   : 0men.ʤəlas.kan1
```

The string delimited by "." of the RE-SEGMENTATION output in output source code above will be used to be matched against the subword unit lookup. In the matching, both subword string and prosodic features are used to retrieve the matched subword unit. The reason to use the prosodic features instead of POS is because we assume matched prosodic features will ensure no or less prosodic mismatch occurs between the joined subword. We do not prioritize POS since we want a higher number of subword unit candidates to be retrieved.

Besides the prosodic features, subword units that come from the same wave are given priority compared to subword unit from other wave file with similar POS and prosodic features.

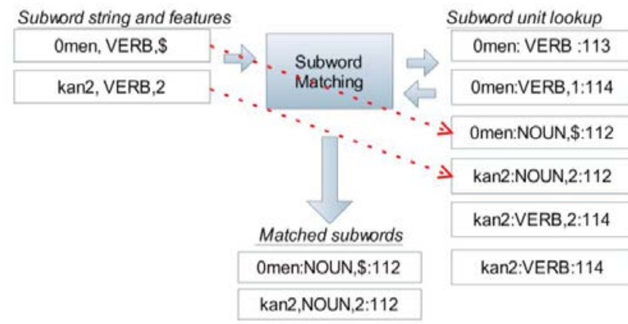


Figure 3 An example of which prosodic features and subword originated from the same wave file are given higher priority.

For example, in Fig. 3, the subword unit of /*0men*/, *NOUN,$*' was chosen based on the prosodic value of '*$*' although its type of POS is different from the target subword. The subword unit of /*kan*2/ with SSTC-ID of '*112*' was chosen instead of the other subword /*ka*n2/, since it comes from the same wave file as the other chosen subword unit.

4. The Smoothness Test

In order to evaluate whether the joint between the subword units is smoothly concatenated, we ran a smoothness test. The smoothness test is a subjective test where subjects will listen and choose which subword joints are smooth and which are not. In this paper, the smoothness test was conducated by replicating the same smoothness test conducted by [8] and [9].

We use the smoothness test setup of [8] [9] because our objective and their experiment objective are the same; that is to assess the joint of small speech units (the size of speech unit smaller than words). In the next subsections, we will describe in detail the data test use, description of participants and how the test conducted. The result obtained will be discussed in next section, which is section 5.

4.1 Participants, Test Data and Procedure

4.1.1 The Participants

The participant of our smoothness test was about 37 people that were invited through phone calls, meeting-in person and e-mails. All of the participants were Malay native speakers with no hearing problem. The reason why only native Malay was chosen was because native speaker's judgment is more competent than the non-native speaker, due to the native speaker better language knowledge than non-native speakers [12]. The gender distribution of male and female was also balanced with 51% were female and 49% male. We also only invited participants who were not working as language technologist. A language technologist defined here is a person who is working in speech technology, natural language processing or understanding, computational linguistics and other related fields in speech and language. This is to ensure the test is done based on the people who has no idea about the state-of-arts of speech synthesis and thus, real user's judgement on smoothness quality in speech synthesis can be captured. Since aged people might have hearing problems, and young people might not sensitive enough to express their judgement, the smoothness test were only taken by people who are aged above 20 but not over 50 years old.

4.1.2 Test Data

Since, we unable to have all the words in the target sentences to be synthesised with subwords unit, the sound test for the smoothness test were made up by combining the concatenation of word and subword synthesis units in total of only five sentences (.wav files).

Table 5: The test data for smoothness test.

Sound	*Unit concatenation*
Estc1	{penggunaan}{0me}{dia1}{0ter}{diri1}{0da}{lam1}{0ben}{tuk1}{0komu}{nika}{si1}{#}{yang}{0men}{gunakan1}{radio}
Estc2	{agak}{0je}{las1{0un}{tuk1}{0me}{nerang}{kan1}{dakwah}{tertentu}{#}{0kera}{na1}{dapat}{memahami}{0ting}{kah1{laku}{0se}{seorang1}
Estc7	{bab}{0ter}{sebut1}{0men}{je}{laskan}{dan}{0meng}{huraikan1}{motif}{dan}{sikap}{mereka}
Estc8	{kehidupan}{yang}{0ber}{lumba1}{0un}{tuk1}{0men}{cari1}{kekayaan}{membuat}{0ki}{ta1}{jahil}
Estc10	{matlamat}{penyampaian}{dakwah}{0a}{da}{lah1} {supaya}{0ma}{nu}{sia1}{0men}{ja}{di1}{baik}

In Table 5, the curly bracket symbol shows the length of synthesis units, whether word or subword were used in the synthesised sentences. The symbol '#' represent a silence unit and the appended number '0' and '1' meaning part of the subwords sound are initial syllable or final syllable.

4.1.3 The Procedure

[8] and [9] conducted their smoothness tests by asking the participant to listen to a synthesised sentence and expressed their judgement on the smoothness of joint synthesis units. In the smoothness test of [8] and [9], the participants were asked to mark words that did not sound smoothly. In our smoothness evaluation test, we asked the participants to mark (by ticking) the empty box next to the word that they perceived as not smooth.

A simple GUI program was developed to conduct the evaluation test. The participants used headphones or speakers to listen to the test sound when they clicked to the corresponding buttons. The participants can replay the test sentences as many times as they want to.

5. Performance Measure for Smoothness Test

The performance measure we used for the smoothness test was by adapting the performance measures of phrase break in [10] and [11]. By replacing (see Table 6 formula acronym definition):

- Total junctures (N) as the total of words and joint subwords (TSWW).
- Total breaks (B) as total of joint subwords (TSW).
- The deletion error (D) as total of joint subwords perceived as word (SWW).
- Insertion error (I) as the total of word perceived as joint subwords (WSW).

The formula for BC and JC in [10] and [11] were used according to our assessment objective.

Table 6: Data collected from smoothness test

Abbreviation	*Description*	*Total words*
TSWW	*Total of all words and joint subwords*	1656
TSW	*Total of all joint subwords*	864
TW	*Total of all words*	792
SWSW	*Total of not smooth joint subwords*	364
WSW	*Total of not smooth words*	140
SWW	*Total of smooth joint subwords*	500
WW	*Total of smooth words*	652

The BC formula in [10] was adapted to measure the percentage of subword sounds that were perceived as not smooth by the participants. We named this formula as SWSW-CORRECT.

$$SWSW\text{-}CORRECT = \frac{TSW - SWW}{TSW} *100\% = \frac{864 - 500}{864} *100\% = \mathbf{42.13\%} \quad (1)$$

We adapted the JC formula in [10] for SWW-INCORRECT calculation that measures how many percent of subword and word that were not smoothly perceived by the participants. In JC formula, S value is included since it is referred as the type of phrase break predicted. However, we ignored S since we do not differentiate the joining of subword. We also ignore I, and instead we replaced it the value of WW. This is because we have to get rid of any word that was perceived as smooth.

$$SWW\text{-}CORRECT = \frac{TSWW - SWW - WW}{TSWW} *100\% = \frac{1656 - 500 - 652}{1656} *100\% = \mathbf{30.43\%} \quad (2)$$

We again adapted the BC formula to calculate the percentage of word sounds that were smoothly perceived by the participants. We named the modified BC formula as WW-CORRECT.

$$WW\text{-}CORRECT = \frac{TW - WSW}{TW} *100\% = \frac{792 - 140}{792} *100\% = \mathbf{82.32\%} \quad (3)$$

Table 7: The assessment scores result.

Description	***Measurement***	***Words (%)***
Subword perceived as not smooth	SWSW-CORRECT	42.13
Subwords and words perceived as not smooth	SWW-INCORRECT	30.43
Word perceived as smooth	WW-CORRECT	82.32

6. Discussion and Conclusion

The objective of the perceptual smoothness test was to find out the percentage of non-smooth joint subwords. Since the joint subwords came from the list of subwords that were assumed would not create audible distortion, therefore, the value of SWSW-CORRECT should be lower and closer to zero percent. Based on the result presented in Table 7, we can conclude that the result of 42.13% is considered as bad result for our objective, which is to have none or less audible distortion for subwords concatenation. However, the result of word perceived as smooth (WW-CORRECT) with only at 82.32%, gives an impression that bad result of SWSW-CORRECT is not solely contributed by the subword concatenation. Since the word synthesis unit is not relevant for segmental mismatch (spectral mismatch), therefore, the audible distortion (non-smoothness) in the smoothness test was more likely caused by the prosodic mismatch. However, with only 30.43% of non-smooth word and joint subwords sounds perceived by participants, suggested that the approach of selecting word and subword unit for synthesis was not bad after all.

We would like to suggest that possibly, the culprit of the non-smoothness in the joint subwords was not only at the segmental mismatch but also at the prosodic mismatch. This was due to the fact that word unit was also perceived as non-smooth by the participants when word unit has nothing to do with segmental mismatch. In other words, the whole process of selecting speech unit for synthesis requires further investigation for an improvement, not only at the subwords units but also at the word units.

References

[1] Klabbers, E. and Veldhuis, R., “Reducing audible spectral discontinuities”, IEEE Transcations On Speech and Audio Processing, Vol.9, No.1, 2001, pp.39–51.
[2] Maris, Y., The Malay Sound System, Kuala Lumpur: Fajar Bakti, 1980.
[3] Teoh, B. S., The Sound System of Malay Revisited, Kuala Lumpur: Dewan Bahasa dan Pustaka, Kuala Lumpur, 1994.
[4] Yi, J. R. and Glass, J. R. (1998). “Natural-sounding speech synthesis using variable-length units”, Proceeding of ICSLP, 1998, pp.1167–1170.
[5] Samsudin, N., “Word selection characteristics based on the manner of articulation”, Technical Report, Universiti Sains Malaysia, 2004.
[6] Lewis, E. and Tatham, M., “Word and syllable concatenation in text-to-speech synthesis”, Proceedings of the European Conference on Speech Communication and Technology, 1999, pp. 615–618.
[7] Tan, T. P., Grapheme to Phoneme System, Unit Terjemahan Melalui Komputer (UTMK), Universiti Sains Malaysia, Penang, Malaysia, 2008.

[8] Tomokiyo, L. M., Black, A. and Lenzo, K. A., "Foreign accents in synthesis: Development and Evaluation", 2005, INTERSPEECH2005.
[9] Tomokiyo, L. M., Peterson, K., Black, A. and Lenzo, K. A. "Intelligibility of machine translation output in speech synthesis", 2006, INTERSPEECH2006.
[10] Taylor, P. and Black, A., "Assigning phrase breaks from part-of-speech sequences", Computer Speech and Language Vol.(12), 1998, pp: 99–117.
[11] Atterer, M. and Klein, E., "Integrating linguistic and performance-based constraints for assigning phrase breaks", Proceedings of the 19th international conference on Computational linguistics, 2002, pp. 1–7.
[12] Jones, C., Berry, L. and Stevens, C., Synthesized speech intelligibility and persuasion: Speech rate and non-native listeners, Computer Speech and Language, Vol. (21), 2007, pp. 641–651.

First Author Sabrina Tiun received her Ph.D from Universiti Sains Malaysia recently. Her research work and interests range from Natural Language Processing, Computational Linguistics to Speech Processing. She is currently working as a lecturer in Universiti Sains Malaysia, Penang, Malaysia.

Second Author Rosni Abdullah is a professor and the Dean of the school of Computer Sciences in Universiti Sains Malaysia, Penang, Malaysia. She is known for her actively involvement in the research of Parallel Processing, Bioinformatics and Cloud Computing.

Third Author Tang Enya Kong is an Associate Professor at Multimedia University, Selangor, Malaysia. He is one of the prominent researchers in Natural Language Processing community in Malaysia. His works covers from Machine Translation, Natural Language Processing, Computational Linguistics, Knowledge Representation and Speech Processing.

IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 2, September 2011
ISSN (Online): 1694-0814
www.IJCSI.org

Nonlinear Robust Regression Using Kernel Principal Component Analysis and R-Estimators

Antoni Wibowo and Mohammad Ishak Desa

Faculty of Computer Science and Information Systems
Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor Malaysia

Abstract

In recent years, many algorithms based on kernel principal component analysis (KPCA) have been proposed including kernel principal component regression (KPCR). KPCR can be viewed as a non-linearization of principal component regression (PCR) which uses the ordinary least squares (OLS) for estimating its regression coefficients. We use PCR to dispose the negative effects of multicollinearity in regression models. However, it is well known that the main disadvantage of OLS is its sensitiveness to the presence of outliers. Therefore, KPCR can be inappropriate to be used for data set containing outliers. In this paper, we propose a novel nonlinear robust technique using hybridization of KPCA and R-estimators. The proposed technique is compared to KPCR and gives better results than KPCR.
Keywords: *Kernel principal component analysis, kernel principal component regression, robustness, nonlinear robust regression, R-estimators.*

1. Introduction

Kernel principal component analysis (*KPCA*) has been proposed to be used for nonlinear systems by mapping an original input space into a higher-dimensional feature space, see [3, 9, 17, 18] for the detailed discussion, and becomes an attractive algorithm because it does not involve nonlinear optimization, it is as simple as the *principal component analysis* (*PCA*) and it does not need to specify the number of principal components prior to modeling compared to other nonlinear methods. In recent years, many nonlinear algorithms based on KPCA have been proposed including *kernel principal component regression* (*KPCR*) which can be viewed as a non-linearization of *principal component regression* (*PCR*) and dispose the effects of multicollinearity in regression models [6, 8, 14, 15, 16, 22]. However, KPCR was constructed based on the *ordinary least squares* (*OLS*) for estimating its regression coefficients which was sensitive to the presence of outliers. An observation is called outlier if it does not follow the OLS based linear regression model. When we use OLS to estimate regression coefficients then outliers have a large influence to the prediction values since squaring residuals magnifies the effect of the outliers. Therefore, KPCR can be inappropriate to be used when outliers are present.

In previous years, several techniques have been developed to overcome the negative effects of outliers such as *R-estimators* which was a robust method based on the ranks of the residuals [7, 11]. However, the previous works applied it for tackling the effect of outliers in the linear regression model. We should notice that the estimate of regression coefficients using R-estimators is obtained through solving a nonlinear optimization problem. To obtain the estimate of regression coefficients, we can use several techniques for solving this nonlinear optimization problems such as genetic algorithms (GAs), simulated annealing and particle swarm optimization [2, 4, 5, 12, 13, 21, 23]. However, applying R-estimators in the ordinary regression still yields linear models which have limitations in applications.

In this paper, we propose a novel robust technique using hybridization of KPCA and R-estimators to overcome the limitation of KPCR and R-estimators in the linear regression. We use KPCA to dispose the effects of multicollinearity in regression and to construct nonlinearity of prediction model by transforming original data into a higher-dimensional feature space and perform a kernel trick to have a multiple linear regression in this space. Then, we perform R-estimators in this linear regression and solve the optimization problems of the R-estimators for obtaining a nonlinear robust regression. We refer the proposed method as the *robust kernel principal component R regression* (*R-KPCRR*).

We organize the rest of manuscript as follows: Section 2, we review theories and methods of R-estimators and

KPCA, followed by R-KPCRR and its algorithm. In Section 3, we compare the performance of the proposed method using several data sets. Finally, conclusions are given in Section 4.

2. Theories and Methods

2.1 R-Estimators

The ordinary multiple linear regression model is given by

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \tag{2.1}$$

where $\mathbf{y} = (y_1 \quad y_2 \quad \cdots \quad y_N)^T \in \mathbf{R}^N$, $\tilde{\mathbf{X}} = (\mathbf{x}_1 \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_N) \in \mathbf{R}^{N\times p}$, $\mathbf{x}_i = (x_{i1} \quad x_{i2} \quad \cdots \quad x_{iN})^T \in \mathbf{R}^p$, $\mathbf{X} = (\mathbf{1}_N \quad \tilde{\mathbf{X}}) \in \mathbf{R}^{N\times(p+1)}$ with $\mathbf{1}_N$ is $N\times 1$ vector with all elements equal to one, $\boldsymbol{\beta} = (\beta_0 \quad \beta_1 \quad \cdots \quad \beta_p)^T \in \mathbf{R}^{p+1}$ is a vector of regression coefficients, $\mathbf{e} = (e_1 \quad e_2 \quad \cdots \quad e_N)^T \in \mathbf{R}^N$ is vector of residuals and $\mathbf{R}$ is the set of real numbers and i = 1, 2, …, N.

When we use OLS to find the estimate of $\boldsymbol{\beta}$, say $\hat{\boldsymbol{\beta}}$, then the estimate is found by minimizing

$$\sum_{i=1}^{N} e_i^2, \tag{2.2}$$

where $e_i = y_i - \dot{\mathbf{x}}_i^T\boldsymbol{\beta}$ and $\dot{\mathbf{x}}_i^T = (1 \quad \mathbf{x}_i^T)$. The solution can be found by solving the following linear equation

$$\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}^T\mathbf{y} \tag{2.3}$$

However, it is well known that the prediction of the OLS based regression will be distorted when outliers are present. To overcome the presence of outliers, we can use R-estimators which minimize

$$\sum_{i=1}^{N} a_N(R_i)e_i \tag{2.4}$$

where R_i is the rank of e_i and $a_N(i)$ is a score function which is monotone and satisfies $\sum_{i=1}^{N} a_N(i) = 0.$ The common choice of $a_N(i)$ are

$$a_N(i) = i - (N+1)/2 \tag{2.5a}$$

and

$$a_N(i) = \Theta^{-1}(i/(N+1)) \tag{2.5b}$$

which are called the Wilcoxon and Van der Waerden scores, respectively, with Θ^{-1} is the inverse of cumulative normal distribution function.

2.2 KPCA

Assume that we have a function $\psi : \mathbf{R}^p \to \mathbf{F}$, where $\mathbf{F}$ is the feature space which it is an Euclidean space with dimension p_F ($p_F \geq p$). Then, we define the matrices $\mathbf{C} = (1/N)\boldsymbol{\Psi}^T\boldsymbol{\Psi} \in \mathbf{R}^{p_F\times p_F}$ and $\mathbf{K} = \boldsymbol{\Psi}\boldsymbol{\Psi}^T \in \mathbf{R}^{N\times N}$ where $\boldsymbol{\Psi} = (\psi(\mathbf{x}_1) \quad \psi(\mathbf{x}_2) \quad \ldots \quad \psi(\mathbf{x}_N))^T \in \mathbf{R}^{N\times p_F}$ and assume that $\sum_{i=1}^{N}\psi(\mathbf{x}_i) = \mathbf{0}$. The relation of eigenvalues and eigenvectors of the matrices $\mathbf{C}$ and $\mathbf{K}$ were studied by Scholkopf *et al.* [18] .

Let $\hat{p}_F$ be the rank of $\boldsymbol{\Psi}$ where $\hat{p}_F \leq \min(N, p_F)$ which implies that both rank($\mathbf{K}$) and rank($\boldsymbol{\Psi}^T\boldsymbol{\Psi}$) are equal to $\hat{p}_F$. It is evident that the eigenvalues of $\mathbf{K}$ are nonnegative real numbers since the matrix $\mathbf{K}$ is symmetric and positive semidefinite [1]. Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{\tilde{r}} \geq \lambda_{\tilde{r}+1} \geq \cdots \geq \lambda_{\hat{p}_F} > \lambda_{\hat{p}_F+1} = \cdots = \lambda_N = 0$ be the eigenvalues of $\mathbf{K}$ and $\mathbf{B} = (\mathbf{b}_1 \quad \mathbf{b}_2 \quad \ldots \quad \mathbf{b}_N)$ be the matrix of the corresponding normalized eigenvectors $\mathbf{b}_s$ (s =1, 2, …, N) of $\mathbf{K}$. Then, let $\boldsymbol{\alpha}_l = \mathbf{b}_l \big/ \sqrt{\lambda_l}$ and $\mathbf{a}_l = \boldsymbol{\Psi}^T\boldsymbol{\alpha}_l$ for l = 1, 2, …, $\hat{p}_F$. The eigenvectors $\mathbf{a}_l$, however, cannot be found explicitly since we do not know $\boldsymbol{\Psi}^T\boldsymbol{\Psi}$ explicitly. However, we can obtain the principal component of $\psi(\mathbf{x})$ corresponding to nonzero eigenvalues of $\boldsymbol{\Psi}^T\boldsymbol{\Psi}$ by using a *kernel trick*. The l-th principal component of $\psi(\mathbf{x})$ (l = 1, 2, …, $\hat{p}_F$) as given as follows:

$$\psi(\mathbf{x})^T\mathbf{a}_l = \sum_{i=}^{N}\alpha_{li}\psi(\mathbf{x})^T\psi(\mathbf{x}_i) \tag{2.6a}$$

where α_{li} is the i-th element of $\boldsymbol{\alpha}_l$. According to Mercer Theorem, if we choose a continuous, symmetric and positive semidefinite kernel $\kappa : \mathbf{R}^P \times \mathbf{R}^P \to \mathbf{R}$ then there exists $\varphi : \mathbf{R}^P \to \mathbf{F}$ such that $\kappa(\mathbf{x}_i, \mathbf{x}_j) = \varphi(\mathbf{x}_i)^T\varphi(\mathbf{x}_j)$ [10, 17]. Instead of choosing ψ explicitly, we choose a kernel κ and employ the corresponding function φ as ψ. Let $K_{ij} = \kappa(\mathbf{x}_i, \mathbf{x}_j)$ then $\mathbf{K}$ and $\boldsymbol{\alpha}_l$ (l = 1, 2, …, $\hat{p}_F$) are explicitly known now. Therefore, Eq. (2.6a) is also explicitly known and can be written as

$$\psi(\mathbf{x})^T \mathbf{a}_l = \sum_{i=}^{N} \alpha_{li} \kappa(\mathbf{x}, \mathbf{x}_i) \qquad (2.6b)$$

2.3 Nonlinear Robust Regression Using KPCA and R-Estimators

The centered multiple linear regression in the feature space is given by

$$\mathbf{y}_0 = \boldsymbol{\Psi}\boldsymbol{\gamma} + \tilde{\mathbf{e}} \qquad (2.7)$$

where $\boldsymbol{\gamma} = \left(\gamma_1 \quad \gamma_2 \quad \ldots \quad \gamma_{\mathrm{P_F}}\right)^T$ is a vector of regression coefficients in the feature space, $\tilde{\mathbf{e}}$ is a vector of random errors and $\mathbf{y}_0 = \left(\mathbf{I}_N - (1/N)\mathbf{1}_N \mathbf{1}_N^T\right)\mathbf{y}$ where $\mathbf{I}_N$ is the $N \times N$ identity matrix.

Since the rank of $\boldsymbol{\Psi}^T\boldsymbol{\Psi}$ is equal to $\hat{p}_F$, then the remaining $(p_F - \hat{p}_F)$ eigenvalues of $\boldsymbol{\Psi}^T\boldsymbol{\Psi}$ are zero. Let λ_k $(k = \hat{p}_F + 1, \hat{p}_F + 2, \ldots, p_F)$ be the zero eigenvalues of $\boldsymbol{\Psi}^T\boldsymbol{\Psi}$ and $\mathbf{a}_k$ be the normalized eigenvectors of $\boldsymbol{\Psi}^T\boldsymbol{\Psi}$ corresponding to λ_k. Furthermore, we define $\mathbf{A} = \left(\mathbf{a}_1 \quad \mathbf{a}_2 \quad \ldots \quad \mathbf{a}_{p_F}\right)$. It is evident that $\mathbf{A}$ is an orthogonal matrix, that is, $\mathbf{A}^T = \mathbf{A}^{-1}$. It is not difficult to verify that

$$\mathbf{A}^T\boldsymbol{\Psi}^T\boldsymbol{\Psi}\mathbf{A} = \mathbf{D}$$

where

$$\mathbf{D} = \begin{pmatrix} \mathbf{D}_{(\hat{p}_F)} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix},$$

$$\mathbf{D}_{(\hat{p}F)} = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \lambda_{\hat{p}_F} \end{pmatrix},$$

and $\mathbf{O}$ is a zero matrix. By using $\mathbf{A}\mathbf{A}^T = \mathbf{I}_{p_F}$, we can rewrite the model (2.7) as

$$\mathbf{y}_0 = \mathbf{U}\boldsymbol{\vartheta} + \tilde{\mathbf{e}} \qquad (2.8)$$

where $\mathbf{U} = \boldsymbol{\Psi}\mathbf{A}$ and $\boldsymbol{\vartheta} = \mathbf{A}^T\boldsymbol{\gamma}$. Let

$$\mathbf{U} = \left(\mathbf{U}_{(\hat{p}_F)} \quad \mathbf{U}_{(p_F - \hat{p}_F)}\right) \text{ and } \boldsymbol{\vartheta} = \left(\boldsymbol{\vartheta}_{(\hat{p}_F)}^T \quad \boldsymbol{\vartheta}_{(p_F - \hat{p}_F)}^T\right)^T,$$

where sizes of $\mathbf{U}_{(\hat{p}_F)}$, $\mathbf{U}_{(p_F - \hat{p}_F)}$, $\boldsymbol{\vartheta}_{(\hat{p}_F)}$ and $\boldsymbol{\vartheta}_{(p_F - \hat{p}_F)}$ are $N \times \hat{p}_F$, $N \times (p_F - \hat{p}_F)$, $\hat{p}_F \times 1$ and $(p_F - \hat{p}_F) \times 1$, respectively. The model (3.3) can be written as

$$\mathbf{y}_0 = \mathbf{U}_{(\hat{p}_F)}\boldsymbol{\vartheta}_{(\hat{p}_F)} + \mathbf{U}_{(p_F - \hat{p}_F)}\boldsymbol{\vartheta}_{(p_F - \hat{p}_F)} + \tilde{\mathbf{e}}. \qquad (2.9)$$

It is easy to verify that $(\mathbf{U}_{(p_F - \hat{p}_F)}\boldsymbol{\vartheta}_{(p_F - \hat{p}_F)})^T \mathbf{U}_{(p_F - \hat{p}_F)}\boldsymbol{\vartheta}_{(p_F - \hat{p}_F)} = 0$ which implies $\mathbf{U}_{(p_F - \hat{p}_F)}\boldsymbol{\vartheta}_{(p_F - \hat{p}_F)}$ is equal to $\mathbf{0}$. Consequently, the model (2.9) reduces to

$$\mathbf{y}_0 = \mathbf{U}_{(\hat{p}_F)}\boldsymbol{\vartheta}_{(\hat{p}_F)} + \tilde{\mathbf{e}}. \qquad (2.10)$$

where $\mathbf{U}_{(\hat{p}_F)} = \boldsymbol{\Psi}\mathbf{A}_{(\hat{p}_F)} = \mathbf{K}\boldsymbol{\Gamma}_{(\hat{p}_F)}$ and $\boldsymbol{\Gamma}_{(\hat{p}_F)} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_{\hat{p}_F})$. It is evident that the elements of $\mathbf{U}_{(\hat{p}_F)}$ are the principal components of $\psi(\mathbf{x}_i)$ for i = 1, 2, …, N. Then, if we only use the first $\hat{r}$ $(\leq \hat{p}_F)$ vectors of $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2$, $\ldots, \boldsymbol{\alpha}_{\hat{p}_F}$, model (2.10) becomes

$$\mathbf{y}_0 = \mathbf{U}_{(\hat{r})}\boldsymbol{\vartheta}_{(\hat{r})} + \boldsymbol{\varepsilon}, \qquad (2.11)$$

where $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_N)^T$ is a vector of residuals influenced by dropping the term $\mathbf{U}_{(p_F - \hat{p}_F)}\boldsymbol{\vartheta}_{(p_F - \hat{p}_F)}$ in model (2.11), respectively. We usually dispose of the term $\mathbf{U}_{(p_F - \hat{p}_F)}\boldsymbol{\vartheta}_{(p_F - \hat{p}_F)}$ for tackling the effects of multicollinearity on the PCA based regressions where the number $\hat{r}$ is called the retained number of nonlinear principal components (PCs) for the KPCR. We can use the ratio λ_l / λ_1 (l = 1, 2, …, $\hat{p}_F$) for detecting the presence of multicollinearity on $\mathbf{U}_{(\hat{r})}$. If λ_l / λ_1 is smaller than, say < 1/1000, then we consider that multicollinearity exists on $\mathbf{U}_{(\hat{r})}$ [11].

Let us consider model (2.11) again. We can see that model (2.11) has the same structure with model (2.1) which implies that we can directly apply R-estimators in model (2.11). For this purpose, we define $\mathbf{U}_{(\hat{r})} = (\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_N)^T \in \mathbf{R}^{N \times \hat{r}}$ and obtain $\varepsilon_i = y_{oi} - \mathbf{u}_i^T \boldsymbol{\vartheta}_{(\hat{r})}$. Then, we minimize

$$\sum_{i=1}^{N} a_N\left(\tilde{R}_i\right)\varepsilon_i, \qquad (2.15)$$

To find the estimators of $\boldsymbol{\vartheta}_{(\hat{r})}$, where $\tilde{R}_i$ is the rank of ε_i, by using a nonlinear optimization solver .

Let $\hat{\boldsymbol{\vartheta}}^*_{(\hat{r})}$ be the estimator of $\boldsymbol{\vartheta}_{(\hat{r})}$ using the above R-estimators. Then, the prediction value of $\mathbf{y}$ *with the first* $\hat{r}$ *vectors of* $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_{\hat{p}_F}$ *using R-estimators*, say $\tilde{\mathbf{y}}$, is given by

$$\tilde{\mathbf{y}} = \bar{y}\mathbf{1}_N + \mathbf{K}\boldsymbol{\Gamma}_{(\hat{r})}\hat{\boldsymbol{\vartheta}}^*_{(\hat{r})} \qquad (2.16)$$

and the residual between $\mathbf{y}$ and $\tilde{\mathbf{y}}$ is given by

$$\tilde{\boldsymbol{\varepsilon}} = \mathbf{y} - \tilde{\mathbf{y}}. \qquad (2.17)$$

The *prediction of the R-KPCRR with the first* $\hat{r}$ *vectors of* $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_{\hat{p}_F}$ is given by

$$g_{(\hat{r})}(\mathbf{x}) = \bar{y} + \sum_{i=1}^{N} d_i \boldsymbol{\kappa}(\mathbf{x}, \mathbf{x}_i) \qquad (2.18)$$

where $(d_1 \quad d_2 \quad \cdots \quad d_N)^T = \boldsymbol{\Gamma}_{(\hat{r})}\hat{\boldsymbol{\vartheta}}^*_{(\hat{r})}$ and $g_{(\hat{r})}$ is a function from $\mathbf{R}^N$ into $\mathbf{R}$.

We summarize the above procedures of the R-KPCRR as follows:

1 . Given $(y_i \quad x_{i1} \quad x_{i2} \quad \cdots \quad x_{iN})$ for i =1, 2, . . . , N.
2. Calculate $\bar{y} = (1/N)\mathbf{1}_N^T\mathbf{y}$ and $\mathbf{y}_0 = (\mathbf{I}_N - (1/N)\mathbf{1}_N\mathbf{1}_N^T)\mathbf{y}$.
3. Choose a kernel $\kappa : \mathbf{R}^p \times \mathbf{R}^p \to \mathbf{R}$ and a function $a_N : \mathbf{R} \to \mathbf{R}$.
4. Construct $K_{ij} = \kappa(\mathbf{x}_i, \mathbf{x}_j)$ and $\mathbf{K} = (K_{ij})$.
5. Diagonalize $\mathbf{K}$.
   Let rank($\mathbf{K}$) = $\hat{p}_F$ and $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{\hat{r}} \geq \lambda_{\hat{r}+1} \geq \cdots \geq \lambda_{\hat{p}_F} > \lambda_{\hat{p}_F+1} = \ldots = \lambda_N = 0$ be the eigenvalues of $\mathbf{K}$ and $\mathbf{b}_1 \quad \mathbf{b}_2 \quad \ldots \quad \mathbf{b}_N$ be the corresponding normalized eigenvectors $\mathbf{b}_s$ (s =1, 2, …,N) of $\mathbf{K}$.
6. Choose $\hat{r}$ $(\leq \hat{p}_F)$ and construct $\boldsymbol{\alpha}_l = \mathbf{b}_l / \sqrt{\lambda_l}$ for l = 1, 2, …, $\hat{r}$. Then, define $\boldsymbol{\Gamma}_{(\hat{r})} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_{\hat{r}})$.
7. Calculate $\mathbf{U}_{(\hat{r})} = \mathbf{K}\boldsymbol{\Gamma}_{(\hat{r})}$ and let $\varepsilon_i = y_{oi} - \mathbf{u}_i^T\boldsymbol{\vartheta}_{(\hat{r})}$.
8. Let $\tilde{R}_i$ be the rank of ε_i.
9. Solve problem (2.15) using a nonlinear optimization solver and let $\hat{\boldsymbol{\vartheta}}^*_{(\hat{r})}$ be solution of (2.15).
10. Calculate $(d_1 \quad d_2 \quad \cdots \quad d_N)^T = \boldsymbol{\Gamma}_{(\hat{r})}\hat{\boldsymbol{\vartheta}}^*_{(\hat{r})}$.
11. Given a vector $\mathbf{x} \in \mathbf{R}^p$, the prediction of the R-KPCRR with the first $\hat{r}$ vectors of $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_{\hat{p}_F}$ is given by

$$g_{(\hat{r})}(\mathbf{x}) = \bar{y} + \sum_{i=1}^{N} d_i \boldsymbol{\kappa}(\mathbf{x}, \mathbf{x}_i).$$

We should note that this algorithm works under the assumption $\sum_{i=1}^{N}\psi(\mathbf{x}_i) = \mathbf{0}$. When $\sum_{i=1}^{N}\psi(\mathbf{x}_i) \neq \mathbf{0}$, we replace $\mathbf{K}$ by $\mathbf{K}_N$ = $\mathbf{K}-\mathbf{EK}-\mathbf{KE}+\mathbf{EKE}$ in Step 4, where $\mathbf{E}$ is the $N \times N$ matrix with all elements equal to $1/N$.

Table 1: Growth of the Son of the Count de Montheillard.

Age (yr, mth [day])	Height (cm)	Age (yr, mth [day])	Height (cm)
0	51.4	9,0	137.0
0,6	65.0	9,7[12]	140.1
1,0	73.1	10,0	141.6
1,6	81.2	11,6	141.9
2,0	90.0	12,0	149.9
2,6	92.8	12,8	154.1
3,0	98.8	13,0	155.3
3,6	100.4	13,6	158.6
4,0	105.2	14,0	162.9
4,7	109.5	14,6[10]	169.2
5,0	111.7	15,0[2]	175.0
5,7	111.7	15,6[8]	177.5
6,0	117.8	16,3[8]	181.4
6,6[19]	122.9	16,6[6]	183.3
7,0	124.3	17,0[2]	184.6
7,3	127.0	17,1[9]	185.4
7,6	128.9	17,5[5]	186.5
8,0	130.8	17,7[4]	186.8
8,6	134.3		

3. Case Studies

3.1 Data Sets

We generated data sets from a trigonometric function and sinc function to test the performances of KPCR and R-KPCRR. The generated data from the trigonometric function and sinc function are given as follows:

$$f(x) = 4.5\sin(2x) + 2.5\cos(x), \qquad (3.1)$$

with $x_i \in [-2\pi : 0.15 : 2\pi]$ and $x_{tj} \in [-2\pi : 0.2 : 2\pi]$;

$$f(x)=\begin{cases}5\sin(|x|)/|x| & \text{if } x\neq 0\\ 5 & \text{otherwise.}\end{cases} \tag{3.2}$$

with $x_i\in[-8:0.25:8]$ and $x_{tj}\in[-6:0.3:6]$, respectively. The notation $[-z:l:z]$ stands for $[-z,z+l,z+2l,\cdots,z]$ where *l* is a real number.

Generally, the generated data from the those functions can be written as $y_i=f(x_i)+e_i$ *i* where *i* = 1, 2,· · ·, *N*. We also generate $y_{tj}=f(x_{tj})+e_{tj}$ where *j* = 1, 2, · · ·,N_t; where N_t is a positif integer. The random noises e_i and e_{tj} are real numbers generated by a normally distributed random with zero mean and standard deviation σ_1 and σ_2, respectively, with $\sigma_1,\sigma_2\in[0,1]$. For shake of comparisons, we set σ_1 and σ_2 equal to 0.2 and 0.3, and call the set of $\{(x_i,y_i)\}$ and $\{(x_{tj},y_{tj})\}$ *the training data set* and *the testing data set*, respectively. In addition, we also used a subset of the famous set of observation taken on the height of the son of the Count de Montbeillard between 1959 and 1977 [19]. Only the first ten years of data were used in this analysis. The growth of son data are given in the Table 1. In these data, we artificially generate the testing data by the relation $x_{jt}=0.1\times rand(1)+\mathrm{x}_j$ and $y_{jt}=0.25\times rand(1)+\mathrm{y}_j$ where *rand*(1) generates a random number which is uniformly distributed in the interval (0,1).

Then, we compare the performance of the above methods using the three data sets with and without outliers. For this purpose, we generated 200 sets of the training data and 200 sets of the testing data. Furthermore, we use the *mean absolute error* (*MAE*) to estimate the prediction error for the training data set which is given by

$$\mathrm{MAE}=(1/N)\sum_{i=1}^{N}|y_i-\tilde{y}_i|. \tag{3.3}$$

The MAE is also used to prediction error of the testing data sets and denoted by MAEt.

In this case studies, outliers are created artificially by moving some (x_i,y_i)s and (x_{tj},y_{tj})s away from designated locations. We generate eight potential outliers for each of the first, second and third data sets where the positions of outliers in x-direction and x_t-direction are chosen randomly in the domain of x_i and domain of x_{tj}, respectively. The positions of outliers in y-direction and y_t-direction are randomly selected in interval [−20, 20] from the correct positions of y_i and y_{tj}, respectively.

3.2 Results

In these case studies, we used the Wilcoxon and Van der Waerden scores for R-estimators and the standard genetic algorithm (GA) for solving the optimization problem of R-estimators. Then, we used the Gaussian kernel $\kappa(\mathbf{x},\mathbf{z})=\exp(-\|\mathbf{x}-\mathbf{z}\|/\rho)$ with parameter ρ is equal to five for both KPCR and R-KPCRR. We involved the estimate of $\vartheta_{(\hat{r})}$ by using KPCR, say $\hat{\vartheta}_{(\hat{r})}$, in the initial population of GA. In the intial population, the *i*-th gene of the other chromosomes (or candidate solutions of $\vartheta_{(\hat{r})}$) is randomly chosen by the formulae

$$\left(\hat{\vartheta}_{(\hat{r})}\right)_i+30\times rand(1)-15 \tag{3.4}$$

where $\left(\hat{\vartheta}_{(\hat{r})}\right)_i$ is the *i*-th element of $\hat{\vartheta}_{(\hat{r})}$ and *i* = 1, 2, …, $\hat{r}$. For the sake of comparisons, the numbers of population, maximum iterations, mutation rate and selection rate are 50, 1000, 0.2 and 0.5, respectively. For each chromosome in any population we sort ε_i in the descending order, say $\varepsilon_{[i]}$ and $\varepsilon_{[1]}\geq\varepsilon_{[2]}\geq\cdots\geq\varepsilon_{[N]}$, to determine its rank. Then, we define $c_k=\varepsilon_{[\lfloor 0.2\times k\times N\rfloor]}$ where *k*=1, 2, 3, 4 and rank ε_i is given by

$$\tilde{R}_i=\begin{cases}5 & \text{if } c_1\leq\varepsilon_i\\ 4 & \text{if } c_2\leq\varepsilon_i<c_1\\ 3 & \text{if } c_3\leq\varepsilon_i<c_2\\ 2 & \text{if } c_4\leq\varepsilon_i<c_3\\ 1 & \text{if } \varepsilon_i<c_4.\end{cases} \tag{3.5}$$

As the results, the three plots of the predictions of KPCR and R-KPCRR corresponding to the three data sets are presented in Figure 1, Figure 2 and Figure 3, respectively. We can see that the predictions of R-KPCRR are less distorted by the presence of outliers compared to KPCR. Table 2 illustrates the prediction errors of KPCR and R-KPCRR. In the case of data with outliers, R-KPCRR with Wilcoxon and Van der Waeden scores give lower MAEs and MAEts compared to KPCR. The MAEs' R-KPCRR with Wilcoxon scores for the trigometric, sinc and growth of son are 1.4527, 1.4250 and 4.8730 whereas the corresponding MAEts are 1.1363, 1.6742 and 6.5350, respectively. The MAEs' R-KPCRR with Van der Waeden score for the trigometric, sinc and growth of son are 1.4774, 1.4609 and 4.9876 whereas the corresponding MAEts are 1.1257, 1.7709 and 6.5776, respectively. Table 2 summarizes MAEs and MAEts of the three data sets without outliers. In this case, we can see that both KPCR and R-KPCRR perform well.

Table 2: MAE and MAEt for KPCR and R-KPCRR with outliers (Wilc=Wilcoxon, VDW=Van der Waerden).

Data	Method	MAE	MAEt
Trigono-Metric $(\hat{r}=10)$	KPCR	1.8804	1.6970
	R-KPCRR Wilc	1.4527	1.1363
	R-KPCRR VDW	1.4774	1.1257
Sinc $(\hat{r}=13)$	KPCR	2.3831	3.3284
	R-KPCRR Wilc	1.4250	1.6742
	R-KPCRR VDW	1.4609	1.7709
Growth of Son $(\hat{r}=14)$	KPCR	5.0125	7.0113
	R-KPCRR Wilc	4.8730	6.5350
	R-KPCRR VDW	4.9876	6.5776

Table 3: MAE and MAEt for KPCR and R-KPCRR without outliers (Wilc=Wilcoxon, VDW=Van der Waerden).

Data	Method	MAE	MAEt
Trigono-Metric $(\hat{r}=10)$	KPCR	0.0822	0.0823
	R-KPCRR Wilc	0.0821	0.0829
	R-KPCRR VDW	0.0814	0.0823
Sinc $(\hat{r}=13)$	KPCR	0.0725	0.0610
	R-KPCRR Wilc	0.0733	0.0613
	R-KPCRR VDW	0.0729	0.0614
Growth of Son $(\hat{r}=14)$	KPCR	0.9663	0.9840
	R-KPCRR Wilc	0.9172	0.9305
	R-KPCRR VDW	0.9229	0.9379

4. Conclusions

We have proposed a novel robust regression using the hybridization of KPCA and R-estimators. Our method yields a nonlinear robust prediction and can dispose the effects of multicollinearity in regression model. The proposed method was performed by transforming original data into a higher dimensional feature space and creating a multiple linear regression in the feature space. After that, we performed a kernel trick to have an explicit multiple linear regression and used R-estimators on this linear model to have a robust regression. Then, we solved the optimization problem of R-estimators using GA for obtaining the estimate of regression coefficients. In this paper, we used Wilcoxon and Van der Waerden scores on R-estimators.

We summarized several important points relating to our cases studies. First, the predictions of R-KPCRR are less distorted and give smaller MAEs and MAEts compared to KPCR when outliers are present in thedata. Second, without outliers, both R-KPCRR and KPCR perform equally well.

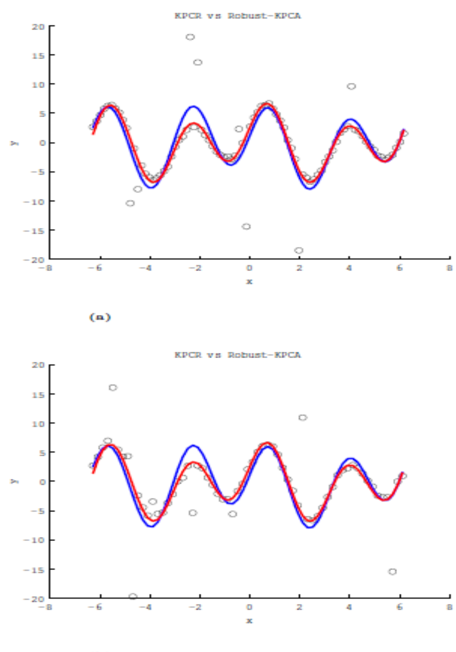


Figure 1: KPCR (Black) and R-KPCRR (red) using Wilcoxon scores with ρ and $\hat{r}$ equal to 5 and 10, respectively. The black circles are trigonometric data with random noises: (a) training data, (b) testing data.

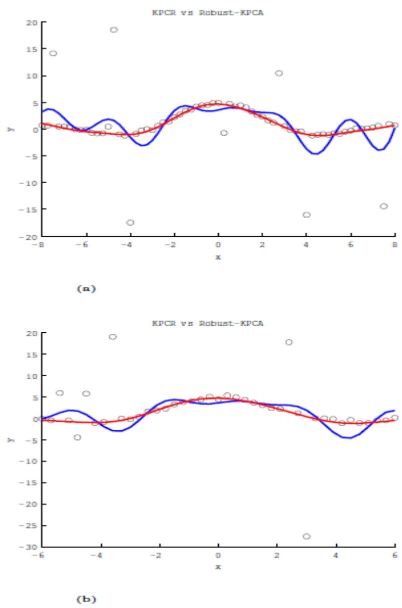


Figure 2: KPCR (Black) and R-KPCRR (red) using Van der Waerden scores with ρ and $\hat{r}$ equal to 5 and 13, respectively. The black circles are sinc data with random noises: (a) training data, (b) testing data.

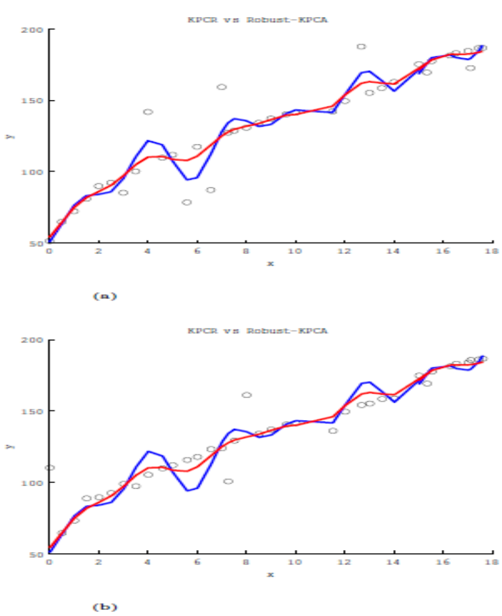


Figure 3: KPCR (Black) and R-KPCRR (red) using Wilcoxon scores with ρ and $\hat{r}$ equal to 5 and 13, respectively. The black circles are the growth of son data with random noises: (a) training data, (b) testing data.

Acknowledgments

The authors sincerely thank to Universiti Teknologi Malaysia and Ministry of Higher Education (MOHE) Malaysia for Research University Grant (RUG) with vot number Q.J130000.7128**.** We also thank to The Research Management Center (RMC) – UTM for supporting this research project.

References

[1] H. Anton, Elementary Linear Algebra, John Wiley and Sons, Inc., 2000.

[2] M.B. Aryanezhad and M. Hemati, "A new genetic algorithm for solving nonconvex nonlinear programming problems", Applied Mathematics and Computation, 86:186–194, 2008.

[3] J. Cho, J. Lee, S.W Choi, D. Lee, and I. Lee, "Fault identification for process monitoring using kernel principal component analysis", Chemical Engineering Science, pages 279–288, 2005.

[4] M. Gen, R. Cheng, and L. Lin, Network Models and Optimization Multiobjective Genetic ALgorithm Approach, Spinger, 2008.

[5] R.L. Haupt and S.E. Haupt, Practical Genetic Algorithms, John Wiley and Sons, 2004.

[6] L. Hoegaerts, J.A.K. Suykens, J. Vandewalle, and B. De Moor, "Subset based least squares subspace in reproducing kernel hilbert space", Neurocomputing, pages 293–323, 2005.

[7] P. Huber, Robust Statistics, John Wiley and Son Inc, 1981.

[8] A.M. Jade, B. Srikanth, B.D Kulkari, J.P Jog, and L. Priya, "Feature extraction and denoising using kernel pca", Chemical Engineering Sciences, 58:4441–4448, 2003.

[9] C. Lu, C. Zhang, T. Zhang, and W. Zhang, "Kernel based symmetrical principal component analysis for face classification", Neurocomputing, 70:904–911, 2007.

[10] H. Q. Minh, P. Niyogi, and Y. Yao, "Mercer's theorem, feature maps, and smoothing", Lecture Notes in Computer Science, Springer Berling, 4005/2006, 2009.

[11] D. C. Montgomery, E. A. Peck, and G. G. Vining, Introduction to Linear Regression, Wiley-Interscience, 2006.

[12] M.S. Osman, Mahmoud A. Abo-Sinn, and A.A. Mousa, "A combined genetic algorithm-fuzzy logic controller (ga-fls) in nonlinear programming", Applied Mathematics and Computation, 170:821–840, 2005.

[13] C. H. Park, W. I. Lee, W. Suck, and A. Vautrin, "Improved genetic algorithm for multidisciplinary optimization of composite laminates", Chemometrics and Intelligent Laboratory Systems, 68:1894–1903, 2008.

[14] R. Rosipal, M. Girolami, L. J. Trejo, and A. Cichoki, "Kernel pca for feature extraction and de-noising in nonlinear regression", Neural Computing and Applications, pages 231–243, 2001.

[15] R. Rosipal and L. J. Trejo, "Kernel partial least squares regression in reproducing kernel hilbert space", Journal of Machine Learning Research, 2:97–123, 2002.

[16] R. Rosipal, L. J. Trejo, and A. Cichoki, "Kernel

principal component regression with em approach to nonlinear principal component extraction", Technical Report, University of Paisley, UK, 2001.

[17] B. Scholkopf, A. Smola, and K.R. Muller, "Nonlinear component analysis as a kernel eigenvalue problem", Neural Computation, 10:1299–1319, 1998.

[18] B. Scholkopf and A.J. Smola, Learning with kernels, The MIT Press., 2002.

[19] G.A.F. Seber and C.J. Wild, Nonlinear Regression, John Wiley and Sons, Inc., 1998.

[20] S.N.Sivanandam and S.N.Deepa, Introduction to Genetic Algorithms, Springer, 2008.

[21] S. Sumathi, T. Hamsapriya, and P. Surekha, Evolutionary Intelligence, Springer, 2008.

[22] A. Wibowo and Y. Yamamoto, "A note of kernel principal component regression", To appear in Computational Mathematics and Modeling, Springer, 2011.

[23] X. Yu and M. Gen, Introduction to Evolutionary Algorithms, Springer, 2010.

Antoni Wibowo is currently working as a senior lecturer in the faculty of computer science and information systems, UTM. He received B.Sc in Math Engineering from University of Sebelas Maret (UNS) Indonesia and M.Sc in Computer Science from University of Indonesia. He also received M. Eng and Dr. Eng in System and Information Engineering from University of Tsukuba Japan. His interests are in the field of computational intelligence , operations research and data analysis.

Mohamad Ishak Desa is a professor in the faculty of computer science and information systems, UTM. He received his B.Sc. in Mathematics from UKM in Malaysia, a postgraduate diploma in system analysis from Aston University, UK. He also received a M.A. in Mathematics from University of Illinois at Springfield, USA and then, a PhD in operations research from Salford University, UK. He is currently the Head of the Operations and Business Intelligences Research Group in UTM. His interests are operations research, optimization, logistic and supply chain, and computational intelligence.

A Metric for Measuring Degree of Service Cohesion in Service Oriented Designs

Mohammad Daghaghzadeh[1], Ahmad Baraani Dastjerdi[2] and Hossein Daghaghzadeh[3]

[1] Department of Computer Engineering, Sheikhbahaee University of Isfahan, Isfahan, Iran

[2] Department of Computer Engineering, University of Isfahan, Isfahan, Iran

[3] Department of Computer, Falavarjan Branch, Islamic Azad University, Falavarjan, Isfahan, Iran

Abstract

Service Orientation has gained momentum as a fundamental design paradigm for developing distributed enterprise applications. However, comprehensive and quantitative metrics for estimating the appropriateness of the service designs are still lacking. One of the quality attributes as to a SOA is cohesion, which is a determining factor for many other desirable features of the software including reusability, agility and etc. The previous studies on the measuring the degree of cohesion failed to comprehensively consider the relationship among entities to evaluate service cohesion. Therefore, this paper puts forward a new metric for measuring the cohesiveness of service and whole design based on design level information, especially the information embedded in entity model. This metric takes into account both Entity-Entity and Entity-Activity relations. Furthermore, the metric conforms to mathematical properties that cohesion metrics must have. The metric is empirically evaluated in a case study.

Keywords: *Service orientation, Cohesion, Software Metrics, Service Identification*

1. Introduction

Service-oriented architecture (SOA) is an information technology (IT) architectural approach that supports the creation of business processes from functional units defined as services [1] [2]. It is actually an evolved architectural concept from object-oriented and component-based developments. SOA promises to provide more agility for organizations adhering to SOA design principles during the entire software development life cycle. In [3] authors consolidated nine design principles which are contract orientation, abstraction, autonomy, coupling, statelessness, cohesion, discoverability, reusability and composability.

However, SOA is still hype because there is no clear method to identify and evaluate the building blocks of a SOA, namely services, against SOA design principles. Designers still have to identify services based on their experience and intuition. Consequently, the designed services barely conform to the principles so that SOA still remains as an unreached promise land for the time being [4].

Cohesion is a prominent attribute of software services and is used as a major quality factor in service design. Cohesion is the degree of functional relevance of activities which are performed by a service to realize a business process [1]. It shows how much an individual service is instrumental in performing one single task [5]. High cohesion brings about ease of understanding of the design model and makes the system more agile. Furthermore, it is conducive to reaching a service granularity at an almost adequate level [2]. In [6] authors recognize seven types of cohesion which are coincidental, logical, communicational, external, sequential, implementation and conceptual cohesions. A service has communicational cohesion if its interface operations operate on the same data. As stated in [5] a service with communicational cohesion is analyzable, changeable, stable and testable. Here in this paper, we examine communicational cohesion because of its significant effect on the quality of software.

There are several works in the literature which have tried to propose some metrics for evaluating and measuring the compliance of the service design against some of the design principles. Most of these studies, which are examined in related work section, focus their attention on common input and output parameters of service operations in order to estimate the service cohesion. However, as some of them indicate, the cohesion between entities should be taken into account while measuring cohesion between service operations. By regarding interrelation among entities, reusability rises [7]. Therefore, in this paper, we aim to put forward a metric for cohesion in order to quantitatively measure the degree of cohesion within each service and whole service design. The introduced metric, considers the interrelation among entities in entity model.

The rest of the paper is arranged as follows. First, previous studies are discussed in section 2. Then, we present some definitions on the main concepts in section 3. Afterward, in section 4, the concept of entity cohesion is investigated. Then, we elaborate our metric in section 5. Next, a case study is conducted to show applicability of the introduced metric in section 6. Later, we evaluate our metric against indispensible mathematical properties that characterize the quality attributes in section 7. In section 8, we empirically validate our metric by investigating the correlation of our metric with subjects' ratings and two previously presented metrics. In the end, we give the conclusion and the future work.

2. Related Work

Cohesion is a software feature that has been attracting the attention of many researchers working on different kinds of software development systems including procedural, object-oriented, component-oriented and service-oriented systems. The metrics for previous systems do not work for service oriented systems without adaptation because of the unique characteristics of service orientation [5]. Therefore, many authors have proposed cohesion metrics for the SOA context by adapting previous metrics. In the context of service oriented systems, there are several works that either investigate cohesion along with other SOA quality attributes or merely focus on proposing a new cohesion metric. In [7], authors consider the number of activities in a service along with shared data flows across such activities. Although, they take into account the relative complexity of each entity, they do not pay attention to the relationships among entities. In [6] [9] [10], the proposed metrics merely take into account number of shared messages among operations. In fact, the number of shared parameters of the service operations is divided by the total number of parameters. One of these metrics is SIDC that we use in this paper to evaluate our metric. In [7], a new cohesion metric called CCM for communicational cohesion is proposed. Authors consider the relationships among entities by relating entities that are accessed by the same operation. However, as they confess at the end of the paper, strengths of relationships are not examined. In [11], two quantitative metrics for cohesion and coupling is put forward. Even though entity-entity relationships are considered, this is done in an incorrect way. The fundamental flaw in this work is the assumption that the higher relationship cardinality between two entities strengthens the cohesion between them. Moreover, different kinds of relationships among entities are not taken into account. Finally, the metric has not empirically evaluated.

3. Basic Concepts

3.1. Service Portfolio Model

Service portfolio model is a model that represents architectural elements and their relationships in establishing a service portfolio from business process decomposition [8].

In this paper, we leverage UML 2.0 profile for software services to model service portfolio which is presented in Figure 1.

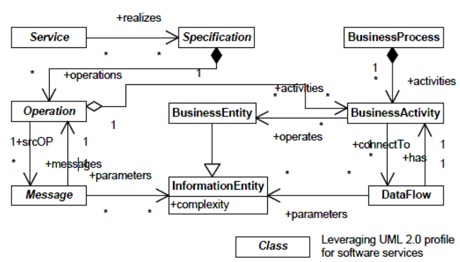


Fig. 1 Service portfolio model [8]

3.2. Entity Model

Entity model is a model that describes entities and their relationships. Entities indicate main stable domain abstractions of an enterprise. The entities are modeled in terms of organizations' proprietary rules and business policies. This brings about entities that are interrelated with each other [12]. Entity model can be described using entity relationship diagrams. Here, we assume that the reader is acquainted with notations used to draw an ER diagram.

- **Strong and weak entities**

Entities which are dependent on other entities for their existence are called weak entities and the relied entities are called strong.

- **Degree of relationship**

The number of entities which participate in a relationship is called degree of relationship. The general form for degree n is the n-ary relationship. A binary relationship is of degree two and a ternary relationship is of degree three.

- **Relationship cardinality ratio**

Determines the cardinal number of entities to which another entity can be connected through a relationship. The primary types of cardinality ratio for relationships are: one-to-one, one-to-many, and many-to-many.

- **Relationship participation**

Relationship participation states whether whole or part of an entity occurrence is associated to another entity. There are three types of relationship participation: mandatory-mandatory, mandatory-optional and optional-optional. In addition, there are some other types of relationship added to basic ER model which are aggregation, generalization and exclusive binary.

4. Entity Cohesion

There are several works in the literature which investigate the problem of entity clustering and Cohesion [13][14][15][16]. Generally, they carry out the grouping based on some prioritized grouping operations. According to [14] there are four kinds of relationship groupings which are: Dominance grouping, abstraction grouping, constraint grouping and relationship grouping**.** These operations should be done consecutively. In [14], authors define a precedence order of grouping operations based on the concept of cohesion borrowed from the software engineering field. They postulate that the level of cohesion between a strong and weak entity, and a dominance and dominated entity, are the highest level. The second level is the cohesion between a supertype and its subtype. Even though, they have not considered aggregation relationships between two entities, based on [13] it can be inferred that aggregation relationships bring about the same level of cohesion. They consider the cohesion between two entities that participate in a constraint relationship at the next level. At the fourth level, unary relationships have the highest cohesion, binary one-to-one next, then binary one-to-many, and finally binary many-to-many. Ternary and higher degrees of relationships are at the fifth level. At the end, there is no cohesion between entities that do not participate in any kind of relationship. Although, they posit the precedence order, they do not propose any figures as to the degree of cohesion. In [15], authors propose some weight numbers corresponding to each level. They use the concept of distance to do so. The distance between a weak entity and a regular entity is considered one. Furthermore, the distance between a supertype entity and each of its subtype entities is equal to 10. This distance is 100 for participating exclusive entities. For a binary relationship, the distance is deemed 1000. In the end, for ternary (or more) relationships, the distance is 10000.

Based on the mentioned studies, we propose the following distances for each kind of relationship (Table 1).

Table 1: Distances for relationship types

Relationship Type	**Distance**
Strong- Weak	1
Generalization	10
Aggregation	10
Exclusive Binary	100

For binary relationships which are not categorized under the above relationships we have the following distances (Table 2).

Table 2: Distances for binary relationships

Participation / **Cardinality ratios**	**M-M**	**M-O**	**O-O**
1-1	1	10	100
1-N	5	50	500
M-N	10	100	1000

In addition, for n-ary relationships (n>2), we multiply each distance in Table 2 by 10^{n-2} .
According to proposed distances, we define the distance between two distinctive entities as the number of edges in the shortest path between them.

$$Distance\ (E_1, E_2) = \begin{cases} \min_p \sum_{i=0}^{e_p} d_i & p > 0 \\ \infty & \mid p = 0 \end{cases} \quad (1)$$

where e_p is the number of edges in a path between E_1 and E_2 and p is the number of existing paths between E_1 and E_2. In addition, d_i is relationship distance which comes either from Table 1 or Table 2. Moreover, for the purpose of this paper, we deem that the distance between identical entities is one. We can measure the degree of cohesion between E_1 and E_2 by Formula 2,

$$\text{Cohesion}\ (E_1, E_2) = \frac{1}{Distance\ (E_1, E_2)} \quad (2)$$

For every entity model, the table in which the distance among all entities is specified can be constructed. In this context, we call such a table Distance Table.

5. The proposed service cohesion metric

In this section, we put forward a new metric for service and whole service design cohesion. Our metric takes into account relationships among entities as well as relationships among activities and entities. Moreover, in this metric, the number of activities in a service affects the degree of cohesion. The reason being is that adding operations to a service lessens its cohesion since the aggregated service does not focus on the semantic of one single task any longer [5].

To calculate cohesion between two activities in a service, we firstly build a complete bipartite graph with business entities on which each activity operates. Each edge with endpoints A and B has weight w coming from row A and column B in the distance table. Then, we take a greedy approach in order to match between the entities that each activity deals with. The summary of the approach can be seen in the pseudo code presented in Figure 2.

```
Void MatchingAlgorithm (BE1, BE2,DT, R)
{
Input: BE1, BE2 are the sets of business entities on which each activity operates and DT is the distance table for the corresponding entity model.
Output: R is a set that keeps the selected edges.
Build a complete bipartite graph G from BE1 and BE2;
Give each edge of graph G weight w that comes from corresponding row and column in DT;
R=Ø;
S=Ø;
While (|S| < |BE1|+|BE2|)
{
Select an edge e from G that has the minimum weight;
If no edge with e's endpoints is present in R and it does not make a circle with other edges in R then
 {
Add the new edge e to set R;
Add e's endpoints to set S;
}
}
}
```

Fig. 2 The algorithm for matching between the entities on which two activities work

To estimate cohesion between two activities, we must obtain average of cohesion among matched entities on which activities work. Therefore, cohesion between two activities i and j called Activity Cohesion (AC) is calculated as follows:

$$\mathrm{AC}\ (\mathrm{i},\mathrm{j}) = \frac{\sum_{\mathrm{i}=1}^{|\mathrm{R}|}\frac{1}{\mathrm{Weight}\ \ (\mathrm{R}(\mathrm{i}))}}{|\mathrm{R}|} \qquad (3)$$

Now, we can calculate Service Cohesion (SC) for each service k by Formula 4,

$$\mathrm{SC}_{\mathrm{k}} = \begin{cases} \sum_{i=1}^{a}\sum_{j=1,i>j}^{a}\frac{AC\ (\mathrm{i},\mathrm{j})}{\frac{a\ (a-1)}{2}} & \mathrm{a}>1 \\ 1 & \mathrm{a}=1, \end{cases} \qquad (4)$$

where a is the number of the activities in k'th service .

In the end, Service Design Cohesion (SDC) is computed by Formula 5,

$$\mathrm{SDC}= \frac{\sum_{k=1}^{s}\mathrm{SC}_{\mathrm{k}}}{s}, \qquad (5)$$

where s is the number of identified services in service portfolio.

6. Case study

In this section, we utilize a real-world business scenario to show the application of the proposed metric and evaluate its usefulness. In this scenario, the goods request process of a mine company is studied.

Every employee in each part of the company can compose a request and fill it out with his needed goods. Then, he sends the request to his boss. Afterward, the bus examines the request to see if the goods are really necessary and the amounts of requested goods do not exceed the determined share of the part. The boss may also add some other goods to the request or edit it. Next, the boss signs the request and sends it to the CEO or his deputy. He checks whether the request conforms to the company's high level policies and regulations. After some probable negotiations and editing, he either signs and then sends the request for store's boss or rejects the request. Store's boss examines the request to make sure that the request does not disturb the balance of the store's stocks.

We analyzed the enterprise and modeled the as-is business process. Then, we obtained the to-be business processes and entity model. Entity model is shown in Figure 3.

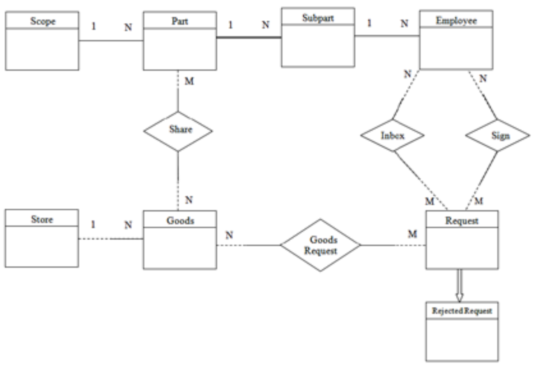


Fig 3.Entity model for good request case study

The distances among all entities are determined based on distances in Table 1 and Table 2 and can be seen in Table 3.

Table 3: The distance table for the good request case study

	Store	Goods	Scope	Part	Subpart	Employee	Request	Goods Request	Share	Inbox	Sign	RejectedRequest
Store	1	50	1055	1050	1055	1060	1050	100	100	1100	1100	1060
Goods		1	1005	1000	1005	1010	1000	50	50	1050	1050	1010
Scope			1	5	10	15	515	1055	55	65	65	525
Part				1	5	10	1010	1050	50	60	60	1020
Subpart					1	5	1005	1055	55	55	55	1015
Employee						1	1000	1050	60	50	50	1010
Request							1	50	1050	50	50	10
Goods Request								1	100	100	100	1060
Share									1	110	110	65
Inbox										1	100	65
Sign											1	65
RejectedRequest												1

We identified two services based on business process models and entity model, namely Request Processing, Goods Processing.

In the following table it is shown that each activity in Request Processing service operates on which business entities.

Table 4: Activities in Request Processing service and related entities

	Good	Scope	Part	Subpart	Employee	Request	Goods Request	Share	Inbox	Rejected Request
UpdateSign						×				
ReadRequest			×	×	×	×				
ReadSigns						×				
CreateRequest						×				
CheckInbox				×	×	×			×	
RejectRequest										×

To evaluate the cohesion between CheckInbox and ReadRequest activities we should build the graph shown in Figure 4.

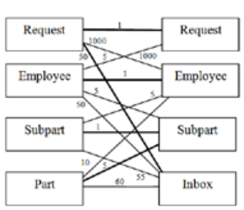


Fig. 4 The bipartite graph for CheckInbox and ReadRequest activities

According to algorithm in Figure 2 the bolded edges are chosen. Now we can calculate the cohesion between the activities as follows:

$$AC\,(4,1) = \frac{\frac{1}{50}+1+1+1+\frac{1}{5}}{5} = 0.644$$

In Table 5, the cohesion among all the activities in the service is shown.

Table 5: Cohesion among all the activities in Request Processing service

	UpdateSign	ReadRequest	ReadSigns	CreateRequest	CheckInbox	RejectRequest
UpdateSign	1	0.25	1	1	0.2555	0.1
ReadRequest		1	0.25	0.25	0.644	0.0257
ReadSigns			1	1	0.2555	0.1
CreateRequest				1	0.2555	0.1
CheckInbox					1	0.029
RejectRequest						1

So, SC for Request Processing is calculated as follows:

$$SC= \frac{\begin{array}{c}0.25+1+1+0.2555+0.1+0.25+0.25+0.664+\\ 0.0257+1+0.2555+0.1+0.2555+0.1+0.029\end{array}}{15} = 0.369$$

Cohesion of Goods Processing service is computed as we did it for Request Processing. So, cohesion of the service design is SDC= $\frac{0.369+0.644}{2} = 0.506$

7. Analytical Validation

In this section, we validate our metric through proving that it satisfies essential properties of a cohesion measure. We exploit properties based software engineering measurement framework [17] to validate our metric.

- **Non-negativity and Normalization**

AC is always between zero and one. The reason being is that all the weights of the edges, which are the distances between entities, are always between one and infinity. Consequently, $\frac{1}{\text{Weight}\ \ (R(i))}$ is always between one and zero so that the average is always between zero and one. As a corollary to this, SC which is the average cohesion of all pairs of activities in a service is between zero and one. Hence, SDC which is the average cohesion of all services embedded in a service design is always between zero and one.

- **Null and Maximum Values**

AC becomes zero when there is no relationship among the business entities that the activities operate on. Furthermore, AC becomes one at its maximum when all operated entities are either the same or participate in strong-weak relationships. Also, minimum of SC is zero when degrees of cohesion among all pairs of activities in a service are zero. Maximum of SC is one when either the cohesion among all pairs of activities in service is one or the service merely contains one activity. In addition, minimum of SDC is zero when the cohesion of all services in service design is zero. Maximum of SDC is one when the cohesion figures of all existing services in the design are one.

- **Monotonicity**

Our metric passes this criterion because adding a shared business entity as a cohesive interaction to the group of business entities on which activities work will not bring about decrease in the degree of cohesion between the activities. Addition of a cohesive interaction will add one edge with the weight of one to set R. Therefore, this will increase numerator and denominator of AC by 1, which does not decrease output of AC ($a \geq 0, b > 0, a \leq b \rightarrow \frac{a+1}{b+1} \geq \frac{a}{b}$).

- **Cohesive module**

Merging two services with non-related activities will result in a service with less degree of cohesion. The reason being is that this action will increase number of zeros in the numerator as it increases the value of $\frac{a\,(a-1)}{2}$ in the denominator.

8. Empirical Validation

In this section, we validate our metric by analyzing the data gathered from a group of 15 experts in SOA. The experts were students of master program in Sheikhbahaee University who had passed advanced software engineering course that SOA was one of its syllabuses. We asked the subjects to rate the cohesions of the services by a numerical scale between 0 and 1. The objects rated by the subjects are services embedded into two service portfolios constructed regarding the case study represented in section 6 (Case study 1) and the one presented in [18] (Case Study 2) as well as their corresponding entity models. The subjects had given guidelines as to how to do the experiment. Each subject did the experiment by himself, at home, and could use unlimited time to rate the cohesion of the given services. Our objective is to establish if any correlation exists between cohesion figures calculated by our metric and subjects' ratings. The Spearman Rank-Difference Correlation Coefficient, r_S, was employed to establish the correlation of the data collected in the experiment because the data acquired in the experiment is distribution free. The Spearman r_S is a non-parametric statistic employed to find out the relationship between two variables expressed as ranks [19].The correlation coefficient is used to determine how much a variable is able to predict the value of another variable. In our experiment the null hypothesis was as follows:

H_0: "there is no correlation between the SC metric and the subjects' rating of service cohesions".

We control the probability that the null hypothesis would be mistakenly rejected by two confidence levels: α_1=0.005 and α_2=0.05. Furthermore, the decision rules for rejecting the null hypothesis are:
For α_1: reject H_0 if r_S >= 0.689; for α_2: reject H_0 if r_S >= 0.447. The correlation coefficient for each subject is shown in Table 6.

Table 6: Spearman Rank Correlations regarding subjects' rating

Subjects	r_s	α_1	α_2
1	0.721	Reject H_0	Reject H_0
2	0.639	Reject H_0	Reject H_0
3	0.781	Reject H_0	Reject H_0
4	0.691	Reject H_0	Reject H_0
5	0.785	Reject H_0	Reject H_0
6	0.873	Reject H_0	Reject H_0
7	0.912	Reject H_0	Reject H_0
8	0.711	Accept H_0	Reject H_0
9	0.587	Accept H_0	Reject H_0
10	0.935	Reject H_0	Reject H_0
11	0.693	Accept H_0	Reject H_0
12	0.831	Reject H_0	Reject H_0
13	0.659	Accept H_0	Reject H_0
14	0.401	Accept H_0	Accept H_0

Based on the data in Table 2, we reject H_0 for 78 percent of the subjects with regard to α_1 and for 92 percent of the subjects considering α_2.
In addition, we calculated the correlation of our metric with SIDC and CCM metrics in two formerly introduced case studies. In Table 7, it is seen that our metric has high correlation with these two metrics.

Table 7: Spearman Rank Correlations regarding the case studies

	CCM	SIDC
Case Study 1	0.89	0.87
Case Study 2	0.92	0.94

9. Discussion

Despite the fact that our figures for service cohesion is close to the figures of SIDC, our metric shows its usefulness in the process of service identification. Even though the difference between figures is not too much, it has significant effect on the shape of service portfolio because an automatic service identification approach is carried out by contrasting these figures rather than by figures themselves. Accordingly, a little difference between two figures determines whether we add an activity to a specific service or not. To demonstrate our contention, we use an example. In the case study presented in section 6, if we did not consider the relationship between Request and RejectedRequest entities, an identification method would consider a separate service for activity RejectRequest despite the fact that these activities are cohesive. Hence, our metric is a rational measure for service cohesion and is effective in service identification.

10. Conclusion and future work

Comprehensive and quantitative metrics for estimating the appropriateness of service designs are still lacking. High cohesion is a determining factor for many other desirable features of software including reusability, agility and etc. Previous studies on measuring the degree of cohesion have failed to thoroughly consider relationship among entities to evaluate service cohesion. In this paper, we put forward a new metric for measuring the cohesiveness of the service and whole design. This metric takes into account both Entity-Entity and Entity-Activity relations. The metric is empirically evaluated in a case study and its correlation with experts' ratings and other metrics were investigated and examined. Furthermore, the metric conforms to mathematical properties that cohesion metrics must have. In the future, we intend to extend this metric in the way to embrace other kinds of cohesions as well. In addition, we are going to use this metric in a method for service identification. Furthermore, it is probable that the figures as to the distance between entities need to be adapted. Finally, we will conduct other case studies in different areas to show the applicability of the metric.

References

[1] A. Erradi, N. Kulkarni and P. Maheshwari, "Service Design Process for Reusable Services: Financial Services Case Study", In 5th International Conference on Service Oriented Computing (ICSOC'07), 2007, Vol. 4749, pp. 606-617.

[2] M.P. Papazoglou and W.-J. van den Heuvel, "Service-Oriented Design and Development Methodology", In International Journal of Web Engineering and Technology, Vol. 2, No. 4, 2006, pp.412-442.

[3] T. Kohlborn, "A Consolidated Approach for Service Analysis", MS. thesis, Chair of Interorganisational Systems , , Westfalische-Wilhelms Universitat Munster, Münster ,Germany, 2008.

[4] C. Llorente, "Model-Driven SOA with Oslo", The Architecture Journal, No. 21, pp. 10-15, 2009.

[5] G. Feuerlicht, "Service Aggregation Using Relational Operations on Interface Parameters", In 4th International Conference on Service-Oriented Computing (ICSOC'06), 2006, Vol. 4652, pp. 407-424.

[6] M. Perepletchikov, C. Ryan, and K. Frampton, "Cohesion Metrics for Predicting Maintainability of Service-Oriented Software", In Seventh International Conference on Quality Software (QSIC 2007), 2007, pp. 328-335.

[7] A. Rostampour, A. Kazemi, F. Shams, A. Zamiri and P. Jamshidi , "A Metric for Measuring the Degree of Entity-Centric Service Cohesion", IEEE International Conference on Service-Oriented Computing and Applications (SOCA'10), 2010, pp.1-5.

[8] Q. Ma, N. Zhou, Y. Zhu and H. Wang, "Evaluating Service Identification with Design Metrics on Business Process Decomposition", In Proceedings of IEEE SCC'2009,2009, pp.160-167.

[9] R. Sindhgatta, B. Sengupta and K. Ponnalagu, "Measuring the Quality of Service Oriented Design", In Proceedings of the 7th International Joint Conference on Service-Oriented Computing, 2009, pp. 485-499.

[10] B. Shim, S. Choue, K. Santae, and P. Sooyong, "A Design Quality Model for Service-Oriented Architecture", In 15th Asia-Pacific Software Engineering Conference, 2008, pp. 403-410.

[11] F. Meng, D. Zhan and X. Xu, "Business Component Identification of Enterprise Information System: A hierarchical clustering method", IEEE International Conference on e-Business Engineering (ICEBE'05), 2005, pp.473-480.

[12] T. Erl, "Service-Oriented Architecture (SOA): Concepts, Technology, and Design", US. : Prentice Hall, 2005.

[13] P.Jaeschke, A. Oberweis and W. Stucky, "Extending ER Model Clustering by Relationship Clustering", In Proceedings of ER'1993, 1993, pp. 451-462.

[14] T. J. Teorey, G. Wei, D. L. Bolton and J. A. Koenig, "ER Model Clustering as an Aid for User Communication and Documentation in Database Design", Communications of the ACM , Vol. 32, No. 8, 1989, pp. 438-450.

[15] J. Akoka and I. Comyn-Wattiau, "Entity-Relationship and Object-Oriented Model Automatic Clustering", In Data & Knowledge Engineering Journal, Vol. 20, No. 2 , 1996, pp. 87-117.

[16] M. Tavana, P. Joglekar and M. A. Redmond, "An Automated Entity-Relationship Clustering Algorithm for Conceptual Database Design", Information Systems, Vol. 32, No. 5, 2007, pp. 773-792.

[17] M. Perepletchikov, C. Ryan, and Z. Tari, "The Impact of Service Cohesion on the Analyzability of Service-Oriented Software", IEEE Transactions on Services Computing, Vol. 3, No. 2, 2010, pp. 89-103.

[18] A . Albani, S. Overhage and D. Birkmeier, "Towards a Systematic Method for Identifying Business Components", In the Proceedings of the 11th International Symposium on Component-Based Software Engineering, 2008, pp. 262-277

[19] S. Siegel and J.N. john Castellan, "Nonparametric Statistics for The Behavioral Sciences", US. : McGraw Hill, 1988.

IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 2, September 2011
ISSN (Online): 1694-0814
www.IJCSI.org

Grammatical Relations of Myanmar Sentences Augmented by Transformation-Based Learning of Function Tagging

Win Win Thant[1], Tin Myat Htwe[2] and Ni Lar Thein[3]

[1] **University of Computer Studies**
Yangon, Myanmar

[2] **Computer Software Department, University of Computer Studies**
Yangon, Myanmar

[3] **University of Computer Studies**
Yangon, Myanmar

Abstract

In this paper we describe function tagging using Transformation Based Learning (TBL) for Myanmar that is a method of extensions to the previous statistics-based function tagger. Contextual and lexical rules (developed using TBL) were critical in achieving good results. First, we describe a method for expressing lexical relations in function tagging that statistical function tagging are currently unable to express. Function tagging is the preprocessing step to show grammatical relations of the sentences. Then we use the context free grammar technique to clarify the grammatical relations in Myanmar sentences or to output the parse trees. The grammatical relations are the functional structure of a language. They rely very much on the function tag of the tokens. We augment the grammatical relations of Myanmar sentences with transformation-based learning of function tagging.

Keywords: *Function Tagging, Grammatical Relations, Transformation Based Learning, Context Free Grammar, Parse Tree.*

1. Introduction

Function tagging is the process of marking up each word in a text with a corresponding function tag like Subj, Obj, Tim, Pla etc. based both on its definition, as well as its context [1]. It has been developed using the statistical implementations, linguistic rules and sometimes both. Identifying the function tags in a given text is an important aspect of any Natural Language Application. We apply TBL for function tagging by extending the Naïve Bayesian based function tagging that is proposed in [2]. The number of function tags in a tagger may vary depending on the information one wants to capture. In the sentence below, the function tags are appended at the end of each word with '#'. For example:

သူ#PSubj သည်#SubjP ကျောင်း#PPla သို့#PlaP သွားသည်#Verb

Grammatical relations are the process of analyzing an input sequence in order to determine its grammatical structure with respect to a given grammar. They show the sentence structure of Myanmar language by using function tags of the words in a sentence. We describe a context free grammar (CFG) based grammatical relations for Myanmar sentences. In the simple sentence below, the grammatical relations are appended at the end of each phrase with '#'. For example:

သူသည်#Subj ကျောင်းသို့#Pla သွားသည်#Verb

In the complex sentence below, the grammatical relations are appended at the end of each phrase with '#'.
For example:

မိုးရွာ#Verb သောကြောင့်#CCS ကျွန်မ#Subj ဈေးသို့#Pla မသွားပါ#Verb

Function tagging and grammatical relations are the important steps in Myanmar to English machine translation. Statistical natural language processing (NLP) research in Myanmar language can only be given a push by the creation of annotated corpus for Myanmar language. In Myanmar language, the availability of the functional annotated tagged corpus is very less and so most of the techniques suffer due to data sparseness problem. We present a method that extends a pre-existing function tagger. Grammatical relations are augmented with transformation-based learning of function tagging.

2. Related Work

We [2] proposed 39 function tags for Myanmar Language and addressed the question of assigning function tags to Myanmar words and used a small functional annotated tagged corpus as the training data. In the task of function tagging, we used the output of morphological analyzer which tagged the function of Myanmar sentences with correct segmentation, POS (part-of-speech) tagging and chunking information. We used Naïve Bayesian statistics to disambiguate the possible function tags of each word in the sentence. We evaluated the performance of function tagging for simple and complex sentences. We concluded our remarks on tagging accuracy by giving examples of some of the most frequent errors. We showed some examples of common error types.

Yong-uk Park and Hyuk-chul Kwon [3] tried to disambiguate for syntactic analysis system by many dependency rules and segmentation. Segmentation is made during parsing. If two adjacent morphemes had no syntactic relations, their syntactic analyzer made new segment between these two morphemes, and found out all possible partial parse trees of that segmentation and combined them into complete parse trees. Also they used adjacent-rule and adverb subcategorization to disambiguate of syntactic analysis. Their syntactic analyzer system used morphemes for the basic unit of parsing. They made all possible partial parse trees on each segmentation process, and tried to combine them into complete parse trees.

Mark-Jan Nederhof and Giorgio Satta[4] considered the problem of parsing non-recursive context-free grammars, i.e., context-free grammars that generateed finite languages and presented two tabular algorithms for these grammars. They presented their parsing algorithm, based on the CYK (Cocke–Younger–Kasami) algorithm and Earley's alogrithm. As parsing CFG (context-free grammar), they have taken a small hand-written grammar of about 100 rules. They have ordered the input grammars by size, according to the number of nonterminals (or the number of nodes in the forest, following the terminology by Langkilde (2000)).

3. Myanmar Language

The Myanmar language, Burmese, belongs to the Tibeto-Myanmar language group of the Sino-Tibetan family. It is also morphologically rich and agglutinative language. Myanmar words are postpositionally inflected with various grammatical features.

3.1 Grammatical Hierarchy in Myanmar

The grammatical hierarchy is a useful notion of successively included levels of grammatical construction operating within and between grammatical levels of analysis [5]. This hierarchy is generally assumed in this study as a heuristic principle for the purposes of laying a foundational understanding of Burmese grammatical units and constructions. This hierarchy is a compositional hierarchy in which lower levels typically are filler units for the next higher level in the hierarchy (Longacre 1970, Pike and Pike 1982). Table 1 shows the hierarchy from the lowest level to the highest.

Table 1: Grammatical Hierarchy

Text
Paragraph
Sentence
Clause
Phrase
Word
Morpheme

3.2 Sentences of Myanmar Language

There are two kinds of sentences according to the syntactic structure of Myanmar language [6][7]. They are simple sentence (SS) and complex sentence (CS). Fig 1 shows the syntactic structure of Myanmar language.

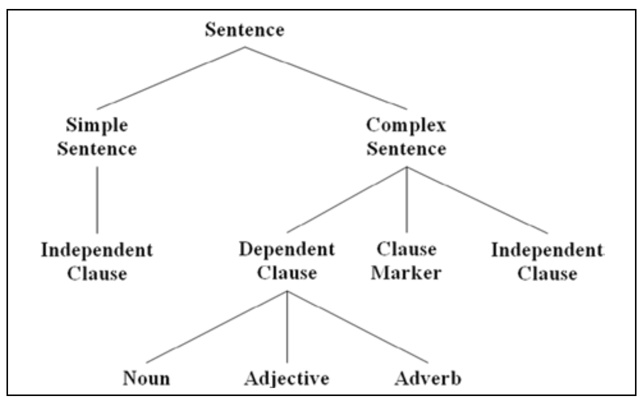


Fig 1: Syntactic Structure

3.2.1 Simple Sentence

It contains only one clause. There are two basic phrases such as subject phrase and verb phrase in a simple sentence. For example:
သူ (Subject phrase) အိပ်နေသည်(Verb phrase)

However, a simple sentence can be constructed by only one phrase. This phrase may be verb phrase or noun phrase.
For example:
စားပါ (Verb phrase)
(သူဘယ်သူလဲ) မမ (Noun phrase)

Besides, a simple sentence can be constructed by two or three phrases.
For example:
သွား (Object phrase) တိုက် (Verb phrase)
ရန်ကုန် တွင် (Place phrase) နေသည် (Verb phrase)

Myanmar phrases can be written in any order as long as the verb phrase is at the end of the sentence.
For example:
ဦးဘသည် မန္တလေးမှ ပြန်လာသည်။ (Subject, Place, Verb)
မန္တလေးမှ ဦးဘသည် ပြန်လာသည်။ (Place, Subject, Verb)

A simple sentence can be extended by placing many other phrases between subject phrase and verb phrase. All of the following are simple sentences, because each contains only one clause. It can be quite long.
For example:
ဦးဘသည် ပြန်လာသည်။
U Ba comes back.
ဦးဘသည် မန္တလေးမှ ပြန်လာသည်။
U Ba comes back from Mandalay.
ဦးဘသည် မန္တလေးမှ ရန်ကုန်သို့ ပြန်လာသည်။
U Ba comes back from Mandalay to Yangon.
ဦးဘသည် မန္တလေးမှ ရန်ကုန်သို့ မီးရထားဖြင့် ပြန်လာသည်။
U Ba comes back from Mandalay to Yangon by train.
ဦးဘသည် မန္တလေးမှ ရန်ကုန်သို့ မီးရထားဖြင့် မနက်က ပြန်လာသည်။
U Ba comes back from Mandalay to Yangon by train in the morning.
ဦးဘသည် မောင်မောင်နှင့်အတူ မန္တလေးမှ ရန်ကုန်သို့ မီးရထားဖြင့် မနက်က ပြန်လာသည်။
U Ba comes back from Mandalay to Yangon by train in the morning with Mg Mg.

It is also constructed by adding noun phrases such as subject phrase, object phrase, time phrase and verb phrase. These added noun phrases are called emphatic phrases.
For example:
ပါမောက္ခ ဦးဘသည် သား မောင်မောင်နှင့်အတူ အထက် မန္တလေးမှ မြို့တော် ရန်ကုန်သို့ အမြန် မီးရထားဖြင့် မနေ့ နံနက်က ချောချောမောမော ပြန်လာသည်။
Professor U Ba and **his son** Mg Mg came back **safely** from **upper** Mandalay to **capital** Yangon by **express** train in **yesterday** morning.

3.2.2 Complex Sentence

A complex sentence consists of two or more independent clauses (or simple sentences) joined by postpositions, particles or conjunctions. There are at least two verbs or more than two verbs in a complex sentence.

There are two kinds of clause in a complex sentence called independent clause(IC) and dependent clause (DC). DC is in front of IC. A complex sentence contains one independent clause and at least one dependent clause. DC is the same as IC but it must contain a clause marker (CM) in the end. A clause maker may be postpositions, particles or conjunctions [8][9]. There are three dependent clauses depending on the clause marker.

(1)**Noun DC** (joined by postpositions such as မှာ၊က၊ကို)
မမ ဈေးသို့ သွားသည် **ကို** ကျွန်မ မြင်သည်။
I see **that** Ma Ma goes to the market.
Noun DC : မမ ဈေးသို့ သွားသည် **ကို**
IC : ကျွန်မ မြင်သည်။

(2)**Adjective DC** (joined by particles such as သော ၊ သည့် ၊ မည့်)
မမ ပေး**သော** စာအုပ် ကို ကျွန်မ ဖတ်သည်။
I read the book **that** is given by Ma Ma.
Adjective DC :မမ ပေး**သော** (စာအုပ်)
IC :စာအုပ် ကို ကျွန်မ ဖတ်သည်။

(3)**Adverb DC** (joined by conjunctions such as သောကြောင့် ၊ လျက် ၊ သဖြင့်)
မိုးရွာနေ **သောကြောင့်** ကျွန်မဈေးသို့ မသွားပါ။
I do not go to the market **because** it is raining.
Adverb DC : မိုးရွာနေ **သောကြောင့်**
IC : ကျွန်မဈေးသို့ မသွားပါ။

4. Corpus Creation

Our corpus is to be built manually. We extended the functional annotated tagged corpus that is proposed in [2].We added sentences from newspapers and historical books of Myanmar to the existing corpus. The corpus consists of approximately 5000 sentences with average word length 15 and it is not a balanced corpus that is a bit biased on Myanmar textbooks of middle school. The corpus size is bigger and bigger because the tested sentences are automatically added to the corpus. Myanmar textbooks and historical books are text collections, as shown in Table 2. In our corpus, a sentence contains chunk, function tag, Myanmar word and its POS tag with category. Fig 2 shows the example corpus sentence.

Table 2: Corpus Statistics

Text types	**# of sentences**
Myanmar textbooks of middle school	1200
Myanmar grammar books	700
Myanmar websites	900
Myanmar newspapers	750
Myanmar historical books	1150
Others	300
Total	5000

VC@Active[မိုးရွာ/v.common] #CC@CCS[လျှင်/cc.sent] # NC@Subj [ကလေး/n.person,များ/part.number] # NC@PPla[လမ်း/n.location] # PPC@PlaP[ပေါ်တွင်/ppm.place] # NC@Obj[ဘောလုံး/n.objects] # VC@Active[ကန်ကြ/verb.common]# SFC@Null[သည်/sf.declarative]။

Fig 2: A sentence in the corpus

5. Function Tagging by Transformation Based Learning

Transformation-based learning starts with a supervised training corpus that specifies the correct values for some linguistic feature of interest, a baseline heuristics for predicting the values for that feature, and a set of rule templates that determine a space of possible features in the neighborhood surrounding a word, and their action is to change the system's current guess as to the feature for the word. The lexical and the contextual rules are generated from the training corpus [10].

We are not concerned with finding the correct attachment of prepositional phrases. We have stressed at several points that the Naive Bayesian assumptions are crude for many properties of natural language syntax. We describe a method for expressing lexical relations in function tagging that statistical function tagging [2] are currently unable to express. One of the strengths of this method is that it can exploit a wider range of lexical and syntactic regularities. In particular, tags can be conditioned on words and on more contexts. Transformation-based tagging encodes complex interdependencies between words and tags by selecting and sequencing transformations that transform an initial imperfect tagging into one with fewer errors [11]. The training of a transformation-based tagger requires an order of magnitude fewer decisions than estimating the large number of parameters of a Naïve Bayesian model.
A transformation consists of two parts, a triggering environment and a rewrite rule. Table 3 shows examples of the type of transformations that are learned given these triggering environments. The first transformation specifies that Cau should be retagged as PCau when the next tag is "CauP ". The first four transformations are triggered by tags and the last three transformations are triggered by words, as shown in Table 3.

Table 3: Examples of some transformations learned in transformation-based tagging

Source tag	**Target tag**	**Triggering environment**
Cau	PCau	the next tag is CauP
PObj	PPla	the second tag is CCC and the fourth tag is PlaP
Obj	Subj	the second tag is CCC and the fourth tag is Active
Obj	Subj	the second tag is CCC and the fourth tag is CCC and the fifth tag is Active
Subj	PcomplS	the lexical item of its next word is "ဖြစ်သည်"
Obj	PcomplS	the lexical item of its next word is "နက်သည်"
Pla	PcomplS	the lexical item of its next word is "ရှိသည်"

6. Error Analysis for Function Tagging

Transformation rules produced by TBL are then used to change the incorrect tags produced by the Naive Bayesian's method. Interestingly it gave an increase of 0.7% for Myanmar initially the accuracy decreased. This is due to the agglutinative nature of Myanmar and the lack of postpositional marker (PPM) in the sentences. There are about 1200 sentences in the test data for function tagging. Error analysis for function tagging is shown in Table 4.

Table 4: Error Analysis for function tagging

Actual Tags	**Assigned Tags**	**Counts**
PcomplS	Subj	133
PcomplS	Obj	108
PcomplS	Pla	52
PcomplS	Tim	24
PSubj	Subj	28
PObj	Obj	37
PTim	Tim	23
PPla	Pla	18
Subj	Obj	54

7. Grammatical Relations

Grammatical functions (or grammatical relations) refer to syntactic relationships between participants in a postposition. Examples are subject, object, time, and place and object complement. We use the context-free grammar (CFG) for grammatical relations of Myanmar sentences. The grammatical relations of the sentences are represented by parse tree. A parse tree is a tree that represents the syntactic structure of a string according to some formal grammar.

The LANGUAGE defined by a CFG is the set of strings derivable from the start symbol S (for Sentence). The core of a CFG grammar is a set of production rules that replaces single variables with strings of variables and symbols. The grammar generates all strings that, starting with a special start variable, can be obtained by applying the production rules until no variables remain. A CFG is usually thought in two ways: a device for generating sentences, or a device if assigning a structure to a given sentence [12]. We use CFG for grammatical relations of function tags.
A CFG is a 4-tuple <N,Σ,P,S> consisting of

- A set of non-terminal symbols N
- A set of terminal symbols Σ
- A set of productions P
  - A-> α
  - A is a non-terminal
  - α is a string of symbols from the infinite set of strings (ΣU N)*
- A designated start symbol S

S	→ SS| CS
SS	→ IC
CS	→ Subj? (Noun_DC| Adj_DC| Adv_DC) IC
Noun_DC	→ IC CCP
Adj_DC	→ IC CCA
Adv_DC	→ IC CCS
IC	→ Subj Obj Pla Active | Subj Active
Subj	→ Subj | PSubj SubjP
Obj	→ Obj | PObj ObjP
Pla	→ Pla | PPla PlaP
Sim	→ PSim SimP
Com	→ PCom ComP

Fig 3: A context free grammar for Myanmar language

7.1 Simple Sentence

Consider a simple declarative sentence "သူသည် စာအုပ်ကို ဆရာ့အား ပေးသည်" (He gives the book to the teacher). This sentence is represented as a sequence of function-tags as "PSubj[သူ]# SubjP[သည်]# PObj[စာအုပ်]#ObjP[ကို] #PIobj[ဆရာ့] # IobjP[အား]#Active[ပေးသည်]"

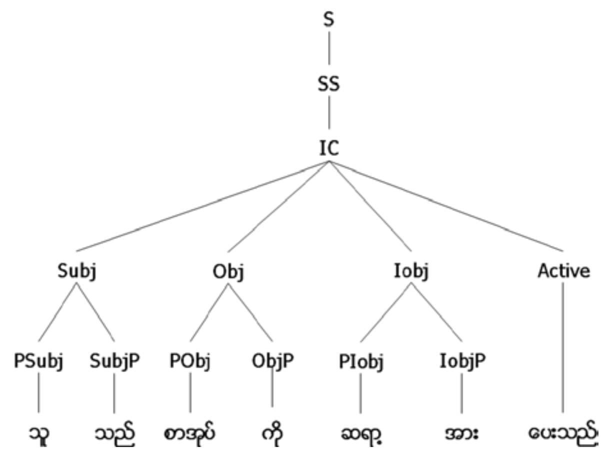


Fig 4: A parse tree for simple sentence

7.2 Complex Sentences

7.2.1 Complex Sentence joined with postpositions

Consider a complex sentence that is joined with postposition (ကို), "ကလေးများ သစ်ပင်အောက်တွင် ကစားနေသည်ကို ကျွန်တော် မြင်သည်" (I see that children are playing under the tree). This sentence is described as a sequence of function-tags as"Subj[ကလေးများ]#PPla[သစ်ပင်] # PlaP[အောက်တွင်]# Active[ကစားနေသည်]# CCP[ကို]# Subj[ကျွန် တော်]#Active[မြင်သည်]".

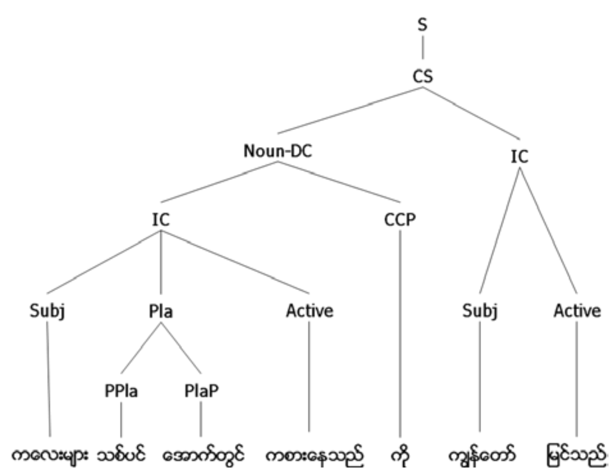


Fig 5: A parse tree for complex sentence (Noun_DC) + (IC)

7.2.2 Complex Sentence joined with particles

Consider a complex sentence that is joined with particle (သော), "ကျွန်တော် ဖတ်နေ သော စာအုပ် ကို အဖေ ဝယ်ခဲ့သည်" (I am reading the book that is bought by my father). This sentence is described as a sequence of function-tags as "Subj[ကျွန်တော်]#Active[ဖတ်နေ]#CCA[သော]#PObj[စာအုပ်]#ObjP[ကို]#Subj[အဖေ]#Active[ဝယ်ခဲ့သည်]".

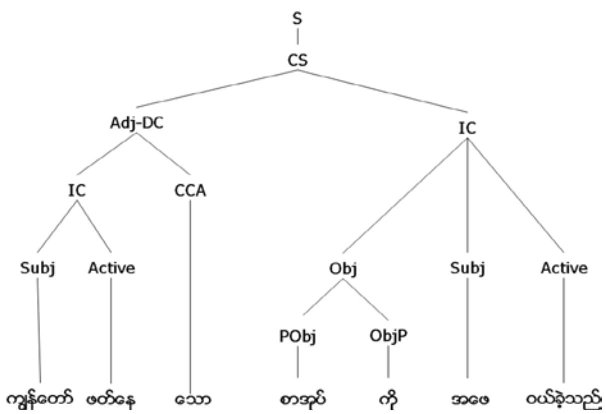


Fig 6: A parse tree for complex sentence (Adj_DC) + (IC)

7.2.3 Complex Sentence joined with conjunctions

Consider a complex sentence that is joined with conjunction (သောကြောင့်), "မောင်မောင် ကြိုးစား သောကြောင့် ဂုဏ်ထူး ရသည်" (Mg Mg gets the distinction because he tried). This sentence is represented as a sequence of function-tags as "Subj[မောင်မောင်]#Active[ကြိုးစား]#CCS[သောကြောင့်]#Obj[ဂုဏ်ထူး]#Active[ရသည်]".

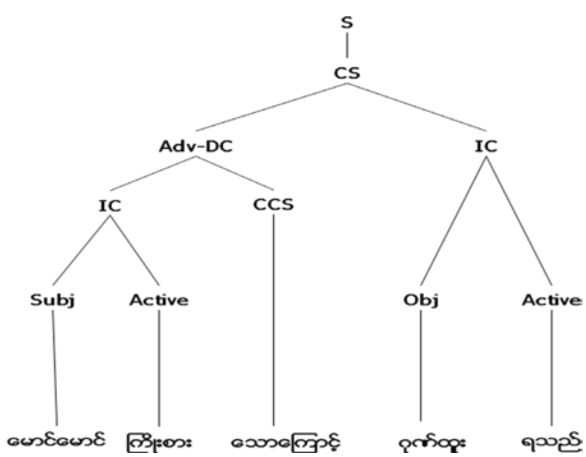


Fig 7: A parse tree for complex sentence (Adv_DC) + (IC)

7.2.4 Complicated Complex Sentence

The unrecognized grammatical relations occurs, are the problem that were caused by the DC that are in the middle of IC and do not has a fixed format. DC may exist between the subject phrase and verb phrase of IC. Consider a complex sentence "မောင်ဘ က ကျွန်တော် စာကျက်နေသည် ဟု ပြောသည်" (Mg Ba says that he is studying). This sentence is described as a sequence of function-tags as "PSubj[မောင်ဘ]#SubjP[က]#Subj[ကျွန်တော်]# Active[စာကျက်နေသည်] #CCP[ဟု]#Active[ပြောသည်]".

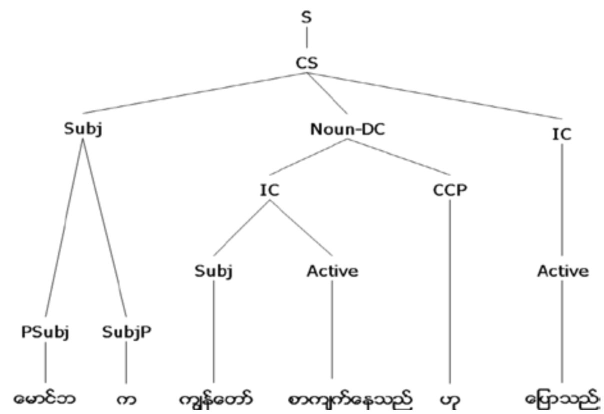


Fig 8: A parse tree for complex sentence Subj+ (Noun_DC) + (IC)

8. Performance Evaluation

Evaluation is based on the performance evaluation by comparing between the system's outputs with the manual parse tree of the sentence. By using the way of assessing the quality of grammatical relations is to assign scores to the output sentences. That is affected by POS tagging and function tagging errors. The evaluation steps describe the evaluation methodology:

- Run the system on the selected test case
- Compare the original parse tree with the system's output
- Classify the criteria that arise from the mismatches between the two grammatical relations of the sentences or parse trees
- Assign a suitable score for each criterion. A range of score between 0 and 3 determines the correctness of the relations. While 0 indicates absolutely incorrect grammatical relations and 3 indicates absolutely correct grammatical relations
- When a situation belongs to multiple problems compute its score average
- Determine the correctness of the test case by computing the percentage of the total scores

Table 5: Accuracy scoring criteria

No	Criterion	Score
1	if the output parse tree is completely wrong format	0
2	if each Myanmar word can generate correct function tag but the grammatical relations are false	1
3	if each Myanmar word cannot generate correct function tag but the grammatical relations are true	1.5
4	if the output sentence is quite well in function tagging and there are some errors in grammatical relations	2
5	if the output parse tree is completely true	3

To the best of our knowledge, there has been no Myanmar-English machine translation before so that there is no standard test set for evaluating Myanmar-English MT system. The data set is derived from the Myanmar textbooks of middle school and Myanmar grammar books, Ministry of Education. The data set consists of 65 sentences for simple sentence, 54 sentences for complex sentence joined with postpositions, 37 sentences for complex sentence joined with particles, 44 sentences for complex sentence joined with conjunctions and 29 sentences for complicated complex sentence.

The system produces 94.36% score for simple sentences while 68.39% score for complicated complex sentences, as shown in Table 6.

Table 6: The result of the score for each sentence type from data set

No	Sentence Types	No. of sentences	Total Score	Score (%)
1	Simple	65	184	94.36
2	Complex (Noun_DC)+(IC)	54	141	87.04
3	Complex (Adj_DC) +(IC)	37	96.5	86.94
4	Complex (Adv_DC) + (IC)	44	121	91.67
5	Complicated Complex	29	59.5	68.39
	Total	229	602	87.63

Fig 9 depicts the relation accuracy for each sentence type. Table 7 shows detailed expression of the score for each sentence type. It can be seen that the proposed system generates 63.5% accuracy for all sentence types, as shown in Table 7.

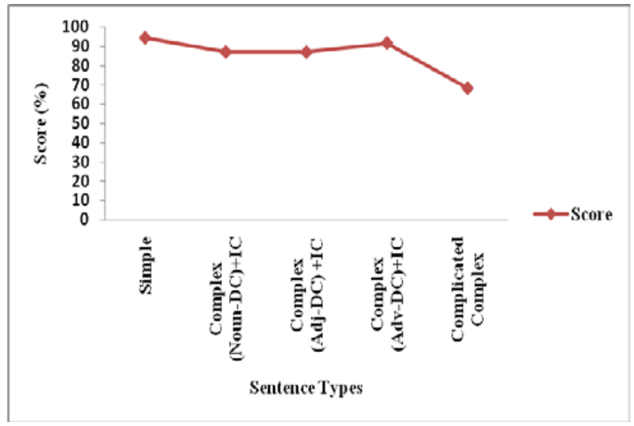


Fig 9: The result of the grammatical relations accuracy for each sentence type

Table 7: The result for each sentence type from the score's point of view

Sentence Types	Score 3	Score 2	Score 1.5	Score 1	Score 0
Simple	74.0%	8.5%	12.1%	5.4%	0.0%
Complex (Noun_DC) +(IC)	67.9%	0.0%	29.4%	2.7%	0.0%
Complex (Adj_DC) +(IC)	62.2%	6.3%	17.4%	14.1%	0.0%
Complex (Adv_DC) +(IC)	81.6%	8.2%	0.0%	10.2%	0.0%
Complicated Complex	32.4%	1.8%	19.5%	46.3%	12%
Accuracy	63.5%	4.7%	15.6%	15.7%	2.4%

Fig 10 to 14 shows the accuracy of grammatical relations for simple and complex sentences. Fig 15 shows the total result of the grammatical relation accuracy from the score point of view.

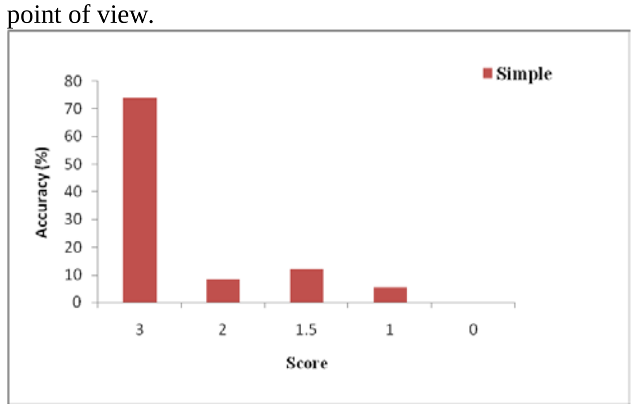


Fig 10: Accuracy for Simple Sentence

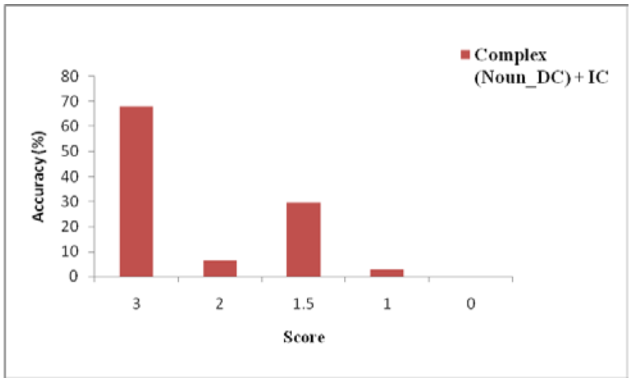


Fig 11: Accuracy for Complex Sentence (Noun_DC) + IC

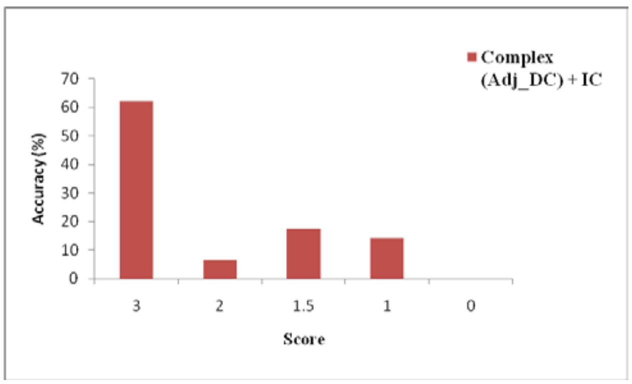


Fig 12: Accuracy for Complex Sentence (Adj_DC) + IC

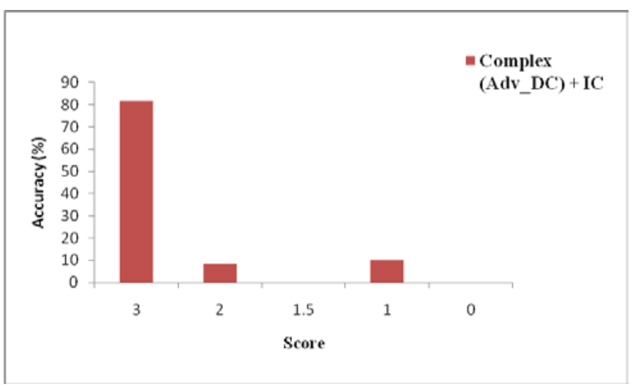


Fig 13: Accuracy for Complex Sentence (Adv_DC) + IC

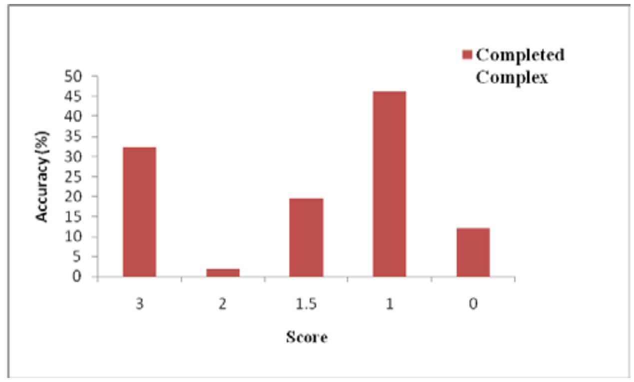


Fig 14: Accuracy for Complicated Complex Sentence

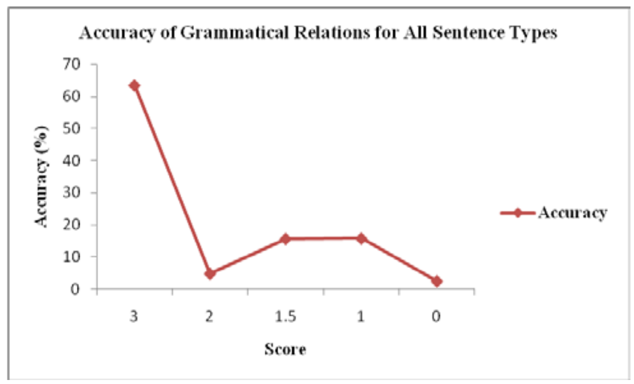


Fig 15: Grammatical relation accuracy for all sentence types from the score point of view

9. Conclusion

We demonstrated the use of TBL for function tagging for Myanmar language. Using TBL method further improved accuracy and produced correct function tags that could not be produced by previous method. Once studied the results and analyzed the mistakes, it must be said that a correct identification of the function tag is crucial in order to obtain a good analysis. If the function tagging fails in this process, the error is dragged throughout the analysis and the result is a badly parse tree. The more accuracy for function tagging increase, the more convenient for grammatical relations of simple sentences and complex sentences of Myanmar language are.

From our experience we have noted that development in natural language processing for Myanmar language is very slow. The main reason for this includes non-availability of large scale data resources and also due to the inherent complexities of the language. The performance of the proposed system can be improved by incorporating more syntactical information by increasing more and more sentence types and well-formed large corpus.

Appendix

Table 8: Function Tagset

Tag	Description	Example
Active	Verb	စားသည်
Subj	Subject	သူ
PSubj	Subject	သူ
SubjP	PPM of Subject	သည်
Obj	Object	ကော်ဖီ
PObj	Object	ကော်ဖီ
ObjP	PPM of Object	ကို
PIobj	Indirect Object	မလှ
IobjP	PPM of Indirect Object	အား
Pla	Place	ရန်ကုန်
PPla	Place	ရန်ကုန်
PlaP	PPM of Place	သို့
Tim	Time	မနက်
PTim	Time	မနက်
TimP	PPM of Time	တွင်
PExt	Extract	ကျောင်းသားများ
ExtP	PPM of Extract	အနက်
PSim	Similie	မင်းသမီး
SimP	PPM of Similie	ကဲ့သို့
PCom	Compare	သူ့ဦးလေး
ComP	PPM of Compare	နှင့်အတူ
POwn	Own	သူ
OwnP	PPM of Own	၏
Ada	Adjective	လှ
PcomplS	Subject Complement	သူသည်**ဆရာ**ဖြစ် သည်
PcomplO	Object Complement	ရွှေကို**လက်စွပ်**လုပ်သည်
PPcomplO	Object Complement	ထွန်းထွန်း
PcomplOP	PPM of Object Complement	ဟု
PUse	Use	တုတ်
UseP	PPM of Use	ဖြင့်
PCau	Cause	မိုး
CauP	PPM of Cause	ကြောင့်
PAim	Aim	အမေ့
AimP	PPM of Aim	အတွက်
CCS	Join with conjunctions	လျှင်
CCM	Join the meanings	ထို့ကြောင့်
CCC	Join the words	နှင့်
CCP	Join with postpositions	ကို
CCA	Join with particles	မည့်

Table 9: Chunk

Chunk Type	Example
Noun Chunk	NC[ခွေး/n.animal]
Postpositional Chunk	PPC[ကို/ppm.obj]
Adjectival Chunk	AC[လှ/adj.dem]
Adverbial Chunk	RC[ပျော်ရွှင်စွာ/adv.manner]
Conjunctional Chunk	CC[နှင့်/cc.chunk]
Verb Chunk	VC[ဖြစ်/v.common]
Sentence Final Chunk	SFC[၏/sf.declarative]

Table 10: POS tags

Description	POS Tag Name
Noun	n
Pronoun	pron
Postpositional Marker	ppm
Adjective	adj
Adverb	adv
Conjunction	cc
Particle	part
Verb	v
Sentence Final	sf

Table 11: Categories

Category	Example
Noun Categories	n.animal, n.food, n.body, n.person, n.group, n.time, n.common, n.building, n.location, n.objects, n.congnition,
Pronoun Categories	pron.person, pron.distplace, pron.disttime, pron.possessive
Postpositional Categories	ppm.subj, ppm.obj, ppm.time, ppm.cause, ppm.use, ppm.sim, ppm.aim, ppm.compare, ppm.accept, ppm.place, ppm.extract,
Adjectival Categories	adj.dem, adj.distobj
Adverbial Categories	adv.manner, adv.state
Conjunctional Categories	cc.sent, cc.mean, cc.chunk, cc.part, cc.adj

Particle Categories	part.type, part.eg, part.number
Verb Categories	v.common, v.compound
Sentence Final Categories	sf.declarative, sf.question, sf.negative,

Acknowledgments

We would like to thank Ministry of Science and Technology, Department of Myanmar, Department of English and the Republic of the Union of Myanmar, for promoting a project on Myanmar to English Machine Translation System, where this part of the work was carried out. Large part of this work was carried out at University of Computer Studies, Yangon and our thanks go to all members of the project for their encouragement and support.

References

[1] D. Blaheta, and M. Johnson," Assigning function tags to parsed text". In Proceedings of the 1st Annual Meeting of the North American Chapter of the Association for Computational Linguistics, 234–240, 2000.

[2] W. W. Thant, T. M. Htwe, and N. L. Thein, "Function Tagging for Myanmar Language", Inernational Journal of Computer Applications, Vol. 26, No. 2, July, 2011

[3] Y. Park and H. Kwon, "Korean Syntactic Analysis using Dependency Rules and Segmentation ", Proceedings of the Seventh International Conference on Advanced Language Processing and Web Information Technology(ALPIT2008), Vol.7, pp.59-63, China, July 23-25, 2008

[4] M. Nederhof and G. Satta, "Parsing Non-Recursive Context-Free Grammars". In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL ANNUAL'02), July 7-12, Pages 112-119, Philadelphia, Pennsylvania, USA, 2002.

[5] J. Okell, A Reference Grammar of Colloquial Burmese, London: Oxford University Press, 1969.

[6] Myanmar Thudda, vol. 1 to 5 in Bur-Myan, Text-book Committee, Basic Edu., Min. of Edu., Myanmar, ca. 1986.

[7] S. P. Soe, Aspects of Myanmar Language, Myanmar Department, University of Foreign Language, 2010.

[8] K. Lay, Construction of Myanmar Thudda. Ph.D. Dissertation, Myanmar Department, University of Educaion, 2003.

[9] P. M, Tin, Some Features of the Burmese Language. Myanmar Book Centre & Book Promotion & Service Ltd, Bangkok, Thailand, 1954.

[10] E. Brill, and P. Resnik, "A transformation-based approach to prepositional phrase attachment disambiguation." In Proceedings, Fifteenth International Conference on Computational Linguistics (COLING-1994), Kyoto, Japan.

[11] E. Brill, "Transformation-based error driven learning and natural language processing: A case study in part-of-speech tagging". Computational Linguistics, 1995.

[12] E. Charniak, "Statistical parsing with a context-free grammar and word statistics". In Proceedings of the Fourteenth National Conference on Artificial Intelligence, pages 598-603, Menlo Park, 1997.

[13] P. H. Myint, "Assigning automatically Part-of-Speech tags to build tagged corpus for Myanmar language", The Fifth Conference on Parallel Soft Computing, Yangon, Myanmar, 2010.

[14] P. H. Myint, "Chunk Tagged Corpus Creation for Myanmar Language". In Proceedings of the ninth International Conference on Computer Applications, Yangon, Myanmar, 2011.

Win Win Thant is a Ph.D research student. She received B.C.Sc (Bachelor of Computer Science) degree in 2004, B.C.Sc (Hons.) degree in 2005 and M.C.Sc (Master of Computer Science) degree in 2007. She is now an Assistant Lecturer of U.C.S.Y (University of Computer Studies, Yangon). She has written one local paper for Parallel and Soft Computing (PSC) conference in 2010, one international paper for International Conference on Computer Applications (ICCA) conference in 2011 and one journal paper for International Journal of Computer Applications (IJCA) in July 2011. Her research interests include Natural Language Processing and Machine Translation.

Tin Myat Htwe is an Associate Professor of U.C.S.Y. She obtained Ph.D degree of Information Technlogy from University of Computer Studies, Yangon. Her research interests include Natural Language Processing, Data Mining and Artificial Intelligence. She has published papers in International conferences and International Journals.

Ni Lar Thein is a Rector of U.C.S.Y. She obtained B.Sc. (Chem.), B.Sc. (Hons) and M.Sc. (Computer Science) from Yangon University and Ph.D. (Computer Engg.) from Nanyang Technological University, Singapore in 2003. Her research interests include Software Engineering, Artificial Intelligence and Natural Language Processing. She has published papers in International conferences and International Journals.

Asynchronous Hybrid Kogge-Stone Structure Carry Select Adder Based IEEE-754 Double-Precision Floating-Point Adder

Abhijith Kini G.

Department of Electronics and Communication Engineering, National Institute of Technology Karnataka, NITK-Surathkal. Surathkal, Karnataka 575025, India.

Abstract

In this paper, the design and implementation of a generic fast asynchronous Hybrid Kogge-Stone Structure Carry Select based Adder (HKSS-CSA) is described in detail and its application in the design of asynchronous Double Precision Floating-Point Adder (DPFPA) is presented and the improved latency performance it provides is discussed. A detailed analysis in terms of maximum combinational delay, number of logic levels and logic resources used by both these adders is provided. The proposed HKSS-CSA adder's performance is compared with a generic reference Carry Look-Ahead Adder (CLA) in terms of the above parameters. For the same set of inputs, the HKSS-CSA resulted in approximately 40% (32-bit) – 65% (128-bit) reduction in the number of logic levels, thereby improving the overall latency by a factor of 2 (32-bit) – 6 (128-bit) times compared to a CLA. A 64-bit instance of this HKSS-CSA was made use of in the design of an asynchronous DPFPA and its performance compared with a reference DPFPA which makes use of a CLA in the intermediate stage. The reference DPFPA had a maximum combinational delay of 36.25ns while the newly suggested DPFPA had a delay of 18.60ns for the same set of inputs, giving about 50% improvement in overall latency performance, which can be mainly attributed to the latency improvement provided by the HKSS-CSA.

Keywords: *Double Precision Floating-Point, Hybrid, Kogge-Stone, Carry Select, Carry Look-Ahead, Adder.*

1. Introduction

Floating-point adders (FPAs) are one of the most frequently used components in modern microprocessors, digital signal processors (DSPs) and graphic processing units (GPUs). These adders must not only be fast enough to accommodate the ever increasing speed requirements of designs but also small enough for designs which make use of a number of these adders in parallel; while maintaining the accuracy at the output. The main bottlenecks in the design of FPAs are latency, area and power. Both floating-point addition and subtraction make use of FPA, hence the latency and throughput of FPA is critical in improving the overall performance of a Floating-Point Unit (FPU). Therefore, a lot of effort was spent on reducing the FPA latency [1-4].

In [1], a DPFPA design which makes uses of flagged prefix addition is proposed and the improvement got in latency is described in detail. A two-path FPA design is proposed to improve the overall latency in [2]. In [3], a design which reduces the overall latency by reducing latency at each sub-module by improvement in structural level by using synthesis method is presented. A variable latency algorithm is proposed and made use of in the design of FPA and the improvements got are detailed in [4]. The design tradeoff analysis of FPAs in FPGAs is presented in [5]. The IEEE standard for binary floating-point arithmetic [6] provides a detailed description of the floating-point representation and the specifications for the various floating-point operations. It also specifies the method to handle special cases and exceptions. Nowadays, most floating point units are IEEE compliant and are capable of handling both single precision and double precision floating-point operands.

The main objective of this paper is to present the design of an asynchronous DPFPA which makes use of newly designed HKSS-CSA and is conformable with the latest draft of IEEE-754 standard, its implementation using Very high-speed integrated-circuit Hardware Description Language (VHDL) and its synthesis for a Xilinx Virtex-V FPGA using Xilinx's Integrated Software Environment (ISE) 9.1i. Asynchronous adders offer many advantages, most important is that they do not use a clock signal, hence not constrained to a global timing constraint. The designed DPFPA accepts normalized double precision floating-point numbers as input and the output is also in the same format. It can also handle special cases and exceptions described in the IEEE standard. This design is compared against a reference DPFPA and analyzed in terms of latency, number of logic levels and logic resources used. This design can be easily extended to support operations on single precision floating-point numbers though better delay performance is got for double precision numbers.

The organization of this paper is as follows: In section 2 the standard algorithm for double precision floating-point addition is presented and design improvements in sub-modules which can lead to increase in overall performance of DPFPA are suggested. The design and implementation of generic asynchronous HKSS-CSA used in the proposed

DPFPA design is discussed in section 3 and its performance compared with a standard CLA. In section 4, the design of the suggested DPFPA is described. Also the performance analysis of the proposed DPFPA adder is done and compared with the reference DPFPA in this section. The main conclusions are elucidated in section 5.

2. Standard DPFPA Algorithm

The double precision floating-point addition is one of the most complex operations in a FPU and needs more logical resources compared to 64-bit adder mainly due to normalization, rounding logic and exception handling. The main steps involved in performing an IEEE double precision floating-point addition are summarized below:

- The first step is to extract the sign, exponent and mantissa part of the operands as per the IEEE representation and check whether the inputs are normalized or any of the special types like NaN, infinity or zero.
- Next step is pre-normalization where the two exponents are compared to identify the larger exponent and the smaller operand is aligned by right shifting it by the absolute difference of the exponents.
- Addition of the two aligned mantissa is performed in the next step.
- The post-normalization step involves detection of carry out and right shifting mantissa by 1 and incrementing the exponent in case of carry-out.
- Rounding is done based on certain internal signals, the exponent and mantissa are updated appropriately.
- The final step involves exception handling where checks are done on the output for any special types.

The algorithm used is shown in the form of a flow chart in Fig. 1. Operand A is unpacked based on IEEE double precision floating-point representation into sign (SA), exponent (EA) and mantissa (MA). Similarly SB, EB and MB are got from operand B. The output sign (SO), exponent (EO) and mantissa (MO) are packed in the last stage of the DPFPA and given out in IEEE double precision floating point format.

A detailed delay analysis of the different sub-modules used in the design of DPFPA revealed that the 53-bit mantissa adder lies in the critical path and is also the largest functional unit block in the data path. Improvements in the design of this adder to have better delay performance should lead to an improvement in the overall latency performance of the DPFPA. Parallel prefix addition is a general technique for speeding up binary addition and has a flexible area-time tradeoff. Parallel prefix adders are derived from the family of CLA. A CLA adder improves the speed by reducing the amount of time required to determine the carry bits.

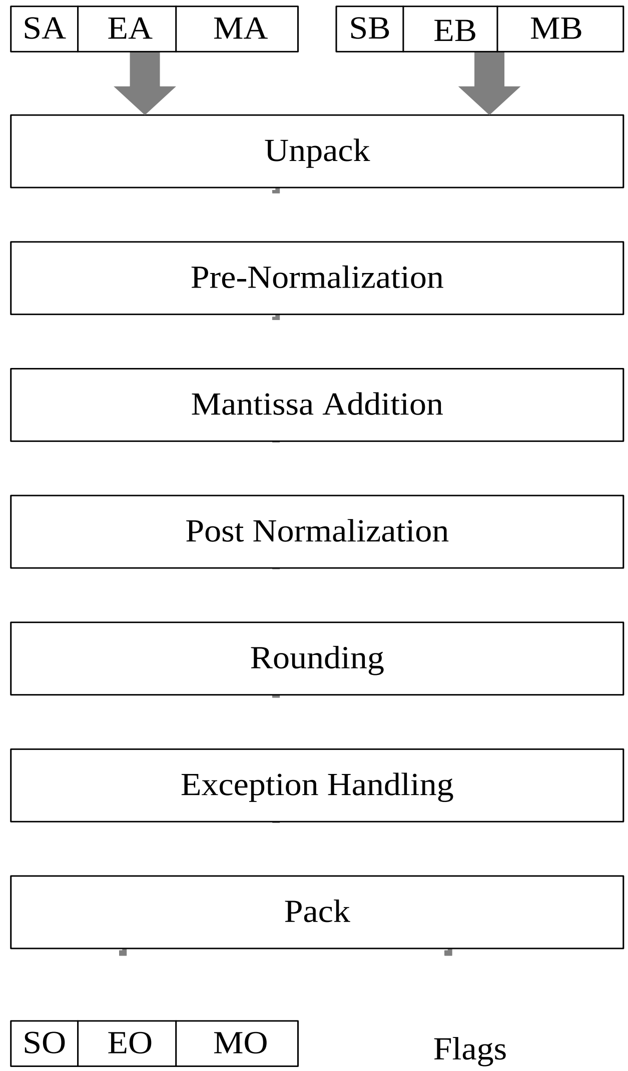


Fig. 1 Asynchronous DPFPA Algorithm.

3. Asynchronous Hybrid Kogge-Stone Structure-Carry Select Based Adder

A general overview of different addition techniques has been presented in [7]. A number of fast adders like carry-skip adder, carry-select adder and carry look-ahead adders have been proposed in the past [8]. The CLA uses the concepts of generating and propagating the carries. The addition of two 1-bit inputs A and B is said to *generate* if the addition will always carry, regardless of whether there is an input carry. The addition of two 1-bit inputs is said to *propagate* if the addition will carry whenever there is an input carry. The equations (1) and (2) are used to find generate and propagate terms respectively.

$$G(A,B) = A \, . \, B \quad (1)$$
$$P(A,B) = A \oplus B \quad (2)$$

The parallel prefix addition can be viewed as a 3-stage process shown in Fig. 2. The improvements can be made primarily in the carry generation stage which is the most intensive one. These adders make use of a tree structure for calculating the carry and hence reduce the latency of the output. Among all trees, the Kogge-Stone tree structure is the most commonly used parallel prefix topology in high performance data paths. The main features are minimum logic depth, regular structure and uniform fan-out. The main disadvantages are large number of wires and high power dissipation. The Kogge-Stone Structure (KSS) based adder is a parallel prefix form of CLA and generates the carry signals in $O(\log_2(N))$ time (where N represents number of input bits) and is considered the fastest adder design widely used in industry.

These designs are better suited for adders with wider word lengths and there is a minimal fan-out of 2 at each node, hence faster performance. The algorithm developed by Kogge-Stone [9] has both optimal depth and low fan-out but produces massively complex circuit realizations and also account for large number of interconnects. Brent-Kung adder [10] makes use of minimal number of computation nodes, hence results in reduced area, but the structure has maximum depth which results in increased latency compared with other structures. The Han-Carlson adder [11] combines Brent-Kung and Kogge-Stone structures to achieve a balance between logic depths and interconnect count. An algorithm for generating parallel prefix adders with variable parameters has been discussed in [12]. A matrix representation for the gate level design of parallel prefix adders has been presented in [13].

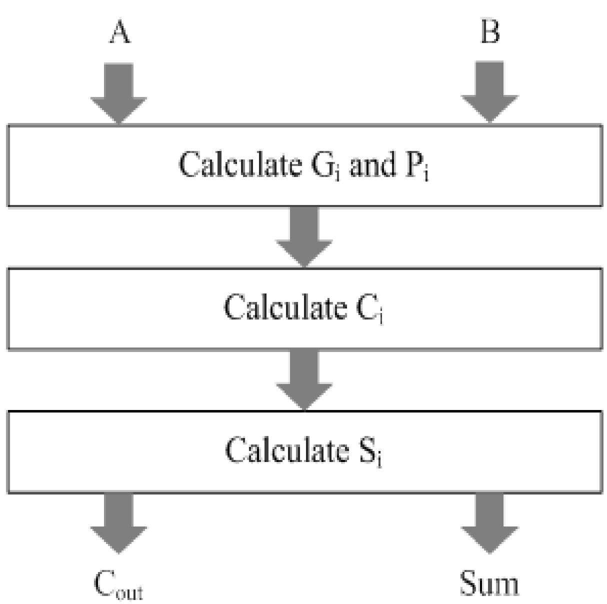


Fig. 2 Parallel-Prefix 3-stage Addition Process.

3.1 Design and Implementation

The parallel prefix adders and CLA adders differ in the way their carry generation block is implemented. In Stage I, generate and propagate components are found for each bit combination of the inputs A and B. The block level representations for each of the cell modules with its corresponding inputs and outputs are shown in Fig. 3. Each of these cell modules were written in VHDL and verified individually. The top module consists of instances of these sub-modules and is verified finally as a single block.

White cell module is used to form the initial generate and propagate components G_i and P_i with A_i and B_i as its inputs. The expression to calculate G_i and P_i are given by (3) and (4) respectively.

$$G_i = A_i \cdot B_i \tag{3}$$
$$P_i = A_i \oplus B_i \tag{4}$$

In Stage II, the calculation of the carries C_i is done by making of use of the Kogge-Stone structure. This stage makes use of black cell, grey cell and buffer cell modules, each having their unique functionalities. The Black cell module takes (G_{in1}, P_{in1}) and (G_{in2}, P_{in2}) as its inputs and forms the (G_{out}, P_{out}) using (5) and (6).

$$G_{out} = G_{in1} + G_{in2} \cdot P_{in1} \tag{5}$$
$$P_{out} = P_{in1} \cdot P_{in2} \tag{6}$$

The Grey cell module outputs only the generate component G_{out} using (5) taking (G_{in1}, P_{in1}) and G_{in2} as its input. The buffer cell acts as a buffer passing its inputs (G_{in1}, P_{in1}) as its outputs (G_{out}, P_{out}).

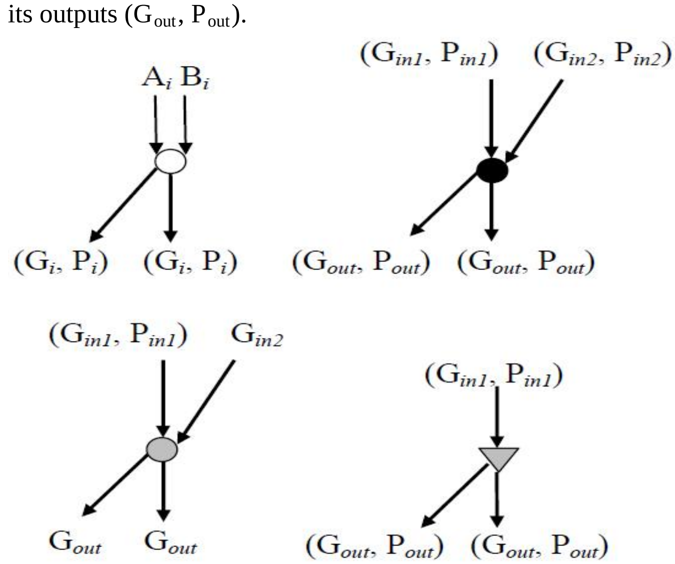


Fig. 3 (Clockwise)
White cell, Black cell, Grey cell, Buffer cell.

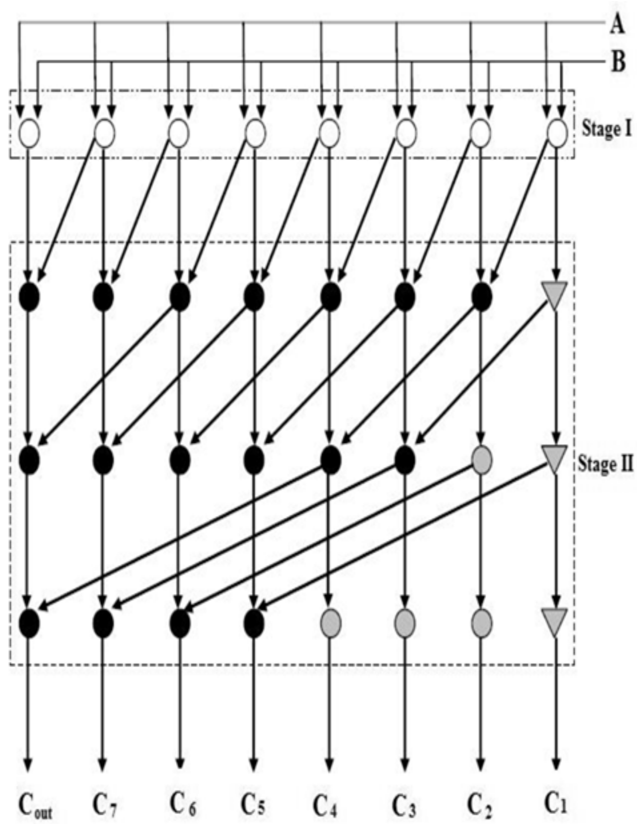


Fig. 4 Kogge-Stone Structure for 8-bit Adder.

The Kogge-Stone tree structure for an 8-bit adder is shown in Fig. 4. The height of a Kogge-Stone tree for an N-bit adder is given by $\log_2$ (N) logic stages. Hence for an 8-bit adder, we have a height of 3 logic stages. The *radix* of the adder refers to how many results from the previous level of computation are used to generate the next one. In this design Radix-2 implementation is done. Higher radix levels can be made use of to reduce the number of stages, but this increases the power and delay. The *sparsity* of the adder refers to how many carry bits are generated by the carry-tree. In this design every carry bit is generated, hence making it Sparsity-1 design.

Kogge-Stone structure has minimal depth but high node count, hence occupies more area in comparison to other tree structures like Brent-Kung, Han-Carlson, and Ladner-Fischer, but has a minimal fan-out of 2 at each given stage, hence giving a better delay performance.

The carries C_i generated in stage II is made use of to calculate the sum S_i in stage III. These carries are used as the carry-in inputs for much shorter RCAs or some other adder design, which generates the final sum bits. But by making use of a Carry Select Adder (CSA) in the last stage, the delay can be further reduced since the sum bits will already be calculated. In a CSA, we compute two results in parallel, each for different carry input assumption. A CSA consists of two Ripple Carry Adders (RCAs) and a multiplexer. Adding two N-bit numbers with a CSA is done with two RCAs in order to perform the calculation twice, one time with the assumption of the carry being zero (C_i = '0') and the other assuming one (C_i = '1'). These two calculations can be done in parallel. After the two results are calculated, the correct sum is selected with the multiplexer once the carry is known. The carries C_i from stage II are fed as inputs to the CSA, so as to choose the appropriate S_i which has already been calculated, thereby reducing the overall delay of the final output. Fig. 5 shows the block diagram of CSA used in the design. The block diagram of the RCA implemented is shown is Fig. 6.

In order to compare the performance of the proposed adder with the CLA, a generic CLA was designed in a structural manner. The block diagram of a generic N-bit CLA is shown in Fig 11. Each of the GPB blocks contains logic which calculates the G_i, P_i and C_j for i = N-2, N-1 … 2, 1, 0 and j = N-1, N-2…, 2, 1 using equations (3), (4) and (7).

$$C_j = G_i + P_i \cdot C_i \tag{7}$$

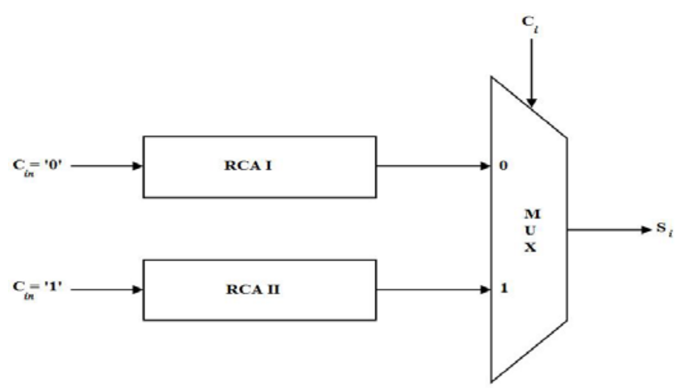


Fig. 5 Block diagram of Carry Select Adder (CSA).

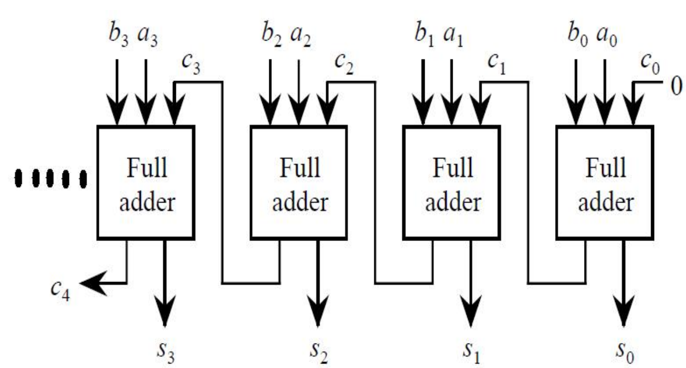
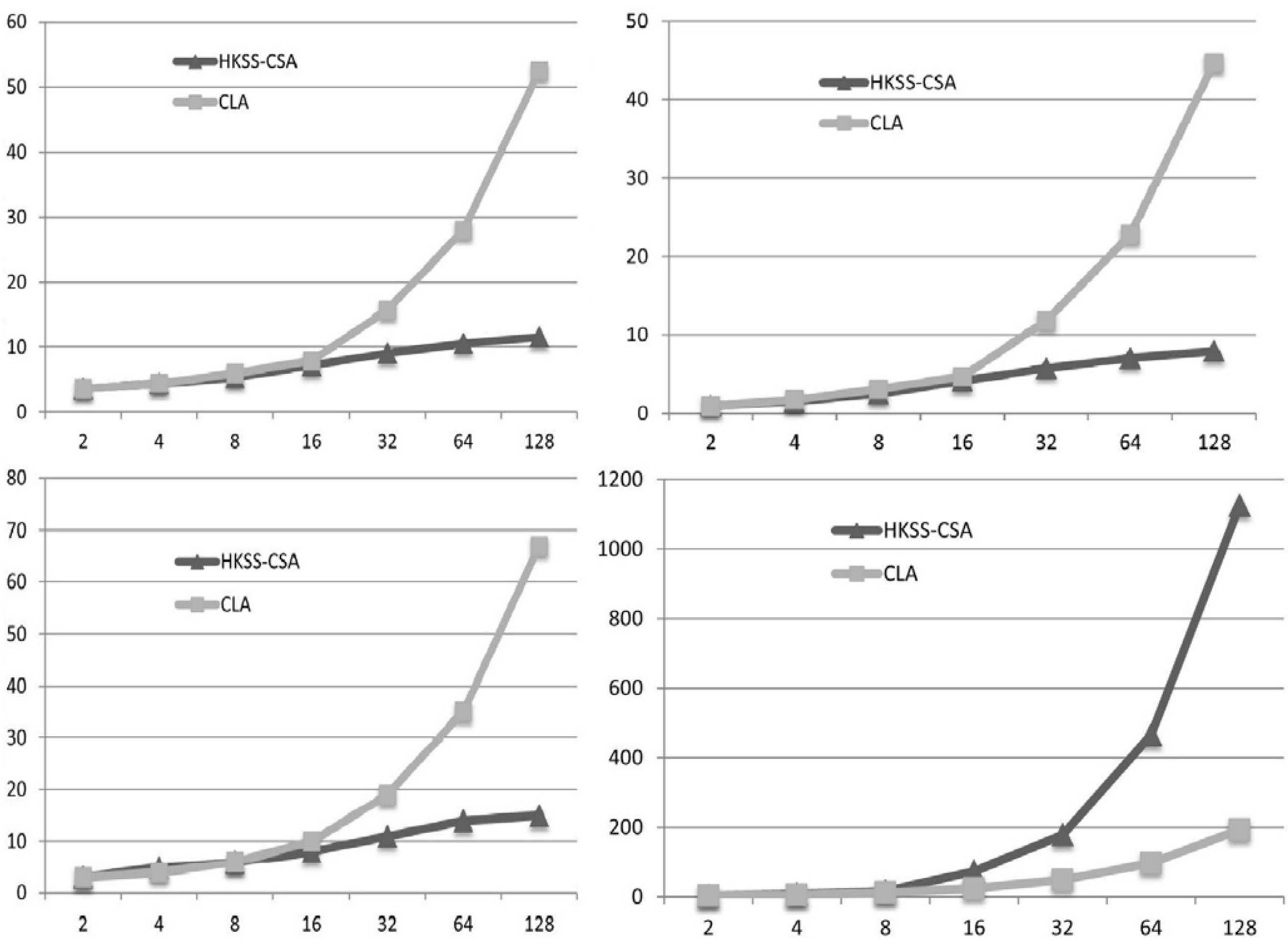


Fig. 6 Block diagram of Ripple Carry Adder (RCA). [7]

Table 1: Delay and device utilization summary.

N (No. of bits)	*Type of Adder*	*Maximum combinational path delay (ns)*	*Route Delay (ns)*	*Logic Delay (ns)*	*Levels of Logic*	*No. of Slices*
2	HKSS-CSA	3.600	0.970	2.630	3	3
	CLA	3.600	0.970	2.630	3	3
4	HKSS-CSA	4.418	1.515	2.903	5	8
	CLA	4.433	1.723	2.710	4	6
8	HKSS-CSA	5.367	2.497	2.870	6	16
	CLA	5.962	3.092	2.870	6	12
16	HKSS-CSA	7.155	4.125	3.030	8	73
	CLA	7.880	4.690	3.190	10	24
32	HKSS-CSA	9.030	5.760	3.270	11	178
	CLA	15.754	11.844	3.910	19	49
64	HKSS-CSA	10.561	7.051	3.510	14	466
	CLA	27.986	22.796	5.190	35	97
128	HKSS-CSA	11.540	7.951	3.590	25	1125
	CLA	52.451	44.701	7.750	67	193



(Clockwise)
Fig. 7 Maximum combinational delay (in ns) as a function of N (bits).
Fig. 8 Routing delay (in ns) as a function of N (bits).
Fig. 9 Levels of logic as a function of N (bits).
Fig. 10 No. of slices as a function of N (bits).

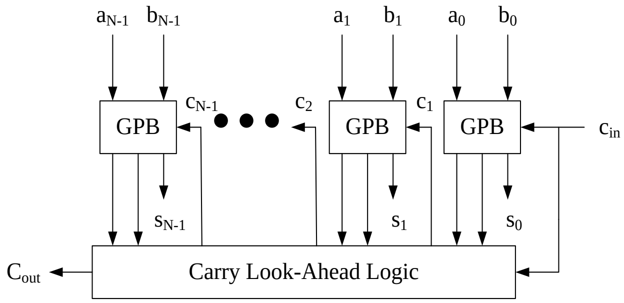


Fig. 11 Block diagram of N-bit CLA.

3.2 Performance Results

The design of the proposed HKSS-CSA and CLA was modeled in VHDL by making use of a module based structural approach. Both the adders where designed and synthesis was done using Xilinx-ISE EDA software on Virtex5 device. Functional testing was done on corner cases, as well as large number of test cases were written to verify the functionality of the adders. The simulation of the waveforms was done using ModelSim-SE.

The performance comparison of the two designs is done in Table 1, which provides the delay, number of logic levels and device utilization summary of HKSS-CSA and CLA. In Fig. 7, a graph of maximum combinational path delay as a function of number of input bits (N) is shown. From this we can observe that there is a drastic increase in the maximum combinational delay for CLA as N increases compared to HKSS-CLA. Hence HKSS-CSA design is more suitable for adders with higher input widths.

As evident from Fig. 8, in CLA the routing delay increases for higher values of N. The advantage of the HKSS-CSA becomes more apparent for higher values of N; delay is reduced by 6 times (N=128) and by 3 times (N=64). The routing delay increases almost in a linear manner for HKSS-CSA while its almost exponential for CLA.

As can be observed from Fig. 9, there is a reduction of about 40% (32-bit) to 65% (N=128) in the number of logic levels when compared with a CLA of same width, thus leading to a better overall delay for the proposed HKSS-CSA when compared to CLA. Also on comparing with Fig. 7, we can see that the number of logic levels has a definite relationship to the maximum combinational path delay.

Also as expected the HKSS-CSA makes use of more number of logic resources when compared to the CLA. CLA uses lesser logic resources in comparison to HKSS-CSA and this difference increases considerably for higher values of N. Hence as the number of required logic resources increases, area also increases. In Fig. 10, a plot of number of slices used as a function of N is shown.

4. Asynchronous Hybrid Kogge-Stone Structure Carry Select Based DPFPA

This section will review the DPFPA algorithm architecture and the hardware sub-modules used in implementing it. The block diagram of DPFPA is shown in Fig. 12.

4.1 Design and Implementation

The standard DPFPA algorithm presented in section 2 is designed and implemented here. The two 64-bit inputs A and B are both in IEEE double precision floating-point format and the output sum will also be in the same format. A 64-bit double precision floating point number defined as per the IEEE-754 standard consists of the 1- bit sign ([63]), 11- bit exponent ([62:52]) and 52- bit mantissa ([51:0]). The design consists of mainly three blocks: pre-adder, adder and post-adder. Each of these blocks contains some Functional Unit Blocks (FUBs) which perform specific functions.

- Pre-Adder Block

The FUBs in the pre-adder block perform operation on the operands to be done before the mantissa addition is done. The *Unpack* FUB basically extracts and signs the sign, exponent and mantissa parts of A (SA, EA, MA) and B (SB, EB, MB) as per IEEE double precision floating point format. The check for special formats is also done in this FUB and flags (EI) sent to the post-adder block. Pre-normalization is done in the *Exp* FUB, in which the two exponents EA and EB are compared to identify the larger exponent. The right shifting of the smaller operand by an amount equal to the absolute difference of the exponents is done in the *Align* FUB. The sign bit of the final sum is got from the *Sign* FUB.

- Adder Block

A 53-bit adder is needed to add the mantissas MAF and MBF coming in from the pre-adder block. The improving the latency performance of this adder we can improve the overall latency performance of the DPFPA. As shown in section III, the newly proposed 64-bit KSS-CSA gives good latency performance in comparison to the CLA. A 64-bit adder is chosen here instead of a 53-bit adder so as to take advantage of the well-ordered tree structure for carry generation, minimal fan-out at each stage, reduced number of logic levels, and the carry select adder in the last stage. Hence the inputs MAF and MBF to the adder block are zero padded before addition and the final result is sent to the post-adder block.

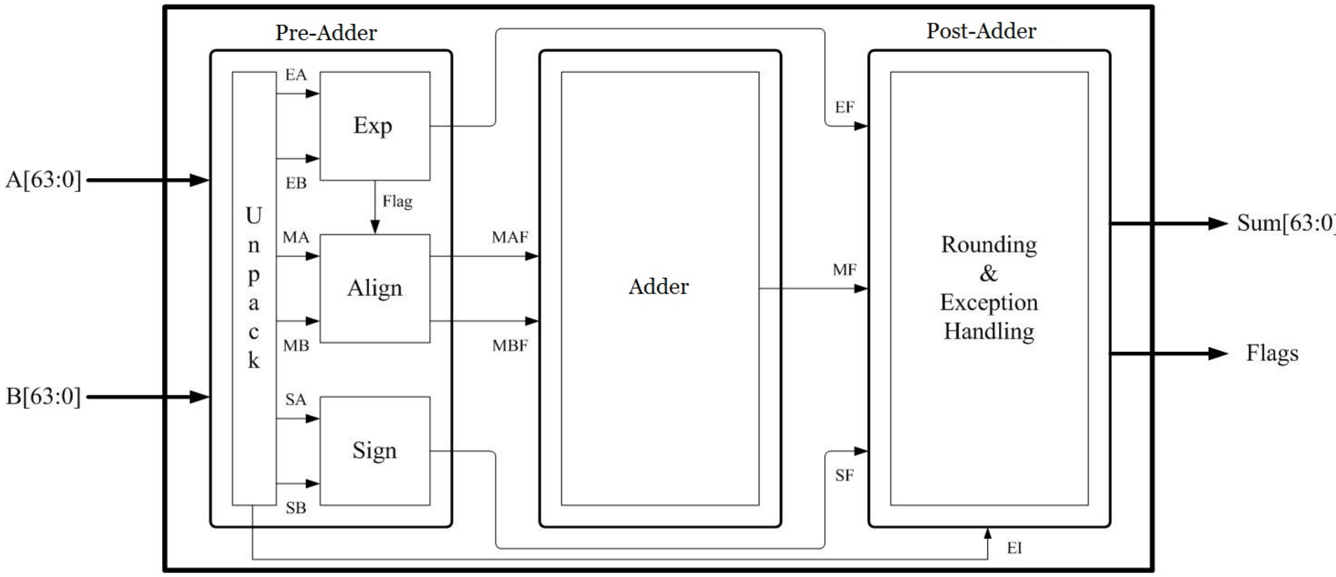


Fig. 12 Block diagram of DPFPA.

- Post-Adder Block

The FUBs in the post-adder block perform operations on the results of mantissa addition. Post-normalization of the output mantissa is done in this FUB. Rounding and exception handling is done in the last stage of the FUB. The result sign, exponent and mantissa are stitched together to form the 64-bit floating point output SUM. Additional flags are also generated at the output to detect special cases.

4.2 Performance Results

In order to compare the performance of the proposed HKSS-CSA based DPFPA against a reference CLA based DPFPA, both the designs were modeled in VHDL and synthesized using Xilinx-ISE EDA software on Virtex5 device. The functionality of the designs was tested by applying a large number of test vectors and Modelsim-SE was used to verify the design. Various corner and special cases were also covered and the performance statistics in terms of maximum combinational delay, route and logic delays, number of levels of logic and slice logic utilization were collected. Table 2 provides the performance comparison data with respect to the proposed and reference DPFPA. Here again we can see a definitive relationship between the number of logic levels and maximum combinational path delay.
The suggested design led to a reduction in the number of logic levels by about 50% compared to the reference DPFPA for the same set of inputs. The route and logic delays which comprise the total delay for the two designs can be seen in Fig. 13. The maximum combinational path delay was reduced from 36.25ns to 18.60ns, so the proposed HKSS-CSA based DPFPA is almost twice as fast as the CLA based DPFPA.
Also as expected the proposed design makes use of more amount of logical resource while giving a better delay performance which can be attributed mainly to the improved latency performance of the intermediate adder. The delay of an adder depends on how fast the carry bit can be generated and used in sum generation. Thus by making use of HKSS-CSA in the mantissa addition stage, the suggested DPFPA gave a better overall delay performance in contrast to the DPFPA which uses the CLA.

Table 2: Delay and device utilization summary of DPFPA.

DPFPA based on	*Maximum combinational path delay (ns)*	*Route Delay (ns)*	*Logic Delay (ns)*	*Levels of Logic*	*Slice Logic*
HKSS-CSA	18.600	13.756	4.844	28	1850
CLA	36.253	29.889	6.364	51	1556

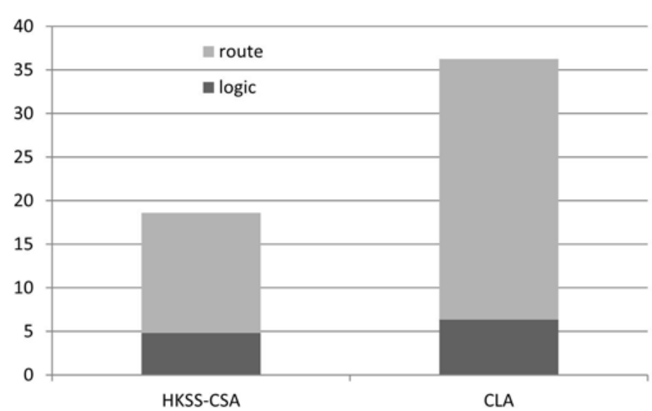


Fig. 13 Delay performance of DPFPA based on HKSS-CSA and CLA.

5. Conclusions

In this paper, the design and implementation of generic HKSS-CSA is proposed and its performance compared with a generic CLA. We have analyzed the delay, number of logic levels and logic resources needed by the HKSS-CSA and CLA for different input widths. The results show that the HKSS-CSA is faster in comparison to CLA because of a number of factors like minimal fan-out at each stage, reduced number of logic levels, well ordered tree structure for carry generation and the carry select adder in the last stage. But with regard to the use of logic resources, the CLA performs better. The new design led to a reduction by 40%-65% in the levels of logic and thus a latency improvement by a factor of 2-6 times when compared to CLA for higher values of N (32, 64, 128 bit). Also this paper proves the tradeoff that exists between maximum combinational delay and amount of logic resources used, and hence the area i.e. area-timing tradeoff. In addition to this, we can see that lower delay advantage of HKSS-CSA can be got for adders with higher input widths.

An application of this HKSS-CSA in the design of fast asynchronous DPFPA is investigated and the performance of this adder compared with a DPFPA which makes use of CLA in the intermediate stage. The suggested design led to a reduction by about 50% in the number of levels of logic compared to one with CLA thereby leading to an overall decrease in the maximum combinational delay. For future work we plan to make use of this HKSS-CSA in the design of single precision and double precision floating point multipliers, to add the mantissa part.

References

[1] A. Beaumont-Smith, N. Burgess, S. Lefrere, and C. Lim, “Reduced Latency IEEE Floating-Point Standard Adder Architectures”, Proc.14th IEEE Symp. Computer Arithmetic, pp. 35-43, 1999.

[2] Peter-Michael Seidel and Guy-Even, “On the Design of Fast IEEE Floating-Point Adders”, Proc. International Symposium on Computer Arithmetic, 2001.

[3] Huang, C., X. Wu, J. Lai, C. Sun and G. Li, “A design of high speed double precision floating point adder using macro modules”, Proc. of the Asia and South Pacific Design Automation Conference, pp: D11-D12, 2005.

[4] S. Oberman, H. Al-Twaijry, and M. Flynn, “The SNAP Project: Design of Floating Point Arithmetic Units,” Proc. 13th IEEE Symp. Computer Arithmetic, pp. 156-165, 1997.

[5] Ali Malik, Lee Moon Ho, Dongdoung Chen, Younhee Choi, Seok Bum Ko, “Design Tradeoff Analysis of Floating-point Adders in FPGAs”, Canadian Journal of Electrical and Computer Engineering, vol. 33, pp. 170 - 175, 2008.

[6] The Institute of Electrical and Electronic Engineers, Inc. IEEE Standard for Binary Floating-point Arithmetic. ANSI/IEEE Std 754-1985.

[7] Prof. Vojin G. Oklobdzija, “VLSI Arithmetic”, http://www.ece.ucdavis. edu/acsel

[8] B. Parhami, Computer Arithmetic–Algorithm and Hardware Designs, Oxford University Press, 2000.

[9] P.Kogge and H.Stone, “A Parallel Algorithm for the efficient solution of a general class of recurrence relations,” IEEE Transactions on Computers, vol. C-22, no.8, August 1973, pp.786-793.

[10] R.Brent and H.Kung, “A Regular Layout for Parallel adders,” IEEE Transaction on Computers, vol. C-31, no.3, March 1982, pp. 260-264.

[11] T. Han and D. Carlson, “Fast Area Efficient VLSI adders,” Proceedings of the 8th Symposium on Computer Arithmetic, September 1987, pp. 49-56.

[12] A. Beaumont Smith, C. C. Lim, "Parallel Prefix Adder Design", Proceedings of 15th IEEE Symposium on Computer Arithmetic, June 2001, pp. 218-225.

[13] Youngmoon Choi, Earl E. Swartzlander Jr, "Parallel Prefix Adder Design with Matrix Representation", Proceedings of 17th IEEE Symposium on Computer Arithmetic, June 2005, pp. 90 - 98.

Abhijith Kini G.: holds a Bachelor’s degree (2006-2010) in Electronics and Communication Engineering from National Institute of Technology Karnataka, NITK-Surathkal, India. His areas of interest are VLSI systems, signal processing and communication systems. He has an IEEE publication in the field of wireless sensor networks and a journal in the field of VLSI signal processing. He is presently working as a component design engineer at Intel India Technology Pvt. Ltd since August 2010.

3D Model Classification and Retrieval Based on Semantic and Ontology

My Abdellah Kassimi [1] and Omar El beqqali [2]

**GRMS2I, Sidi Mohamed Ben Abdellah University,
Fez Morocco**

Abstract

Classifying 3D models into classes is an important step in 3D model retrieval process. However, most classification system does not include semantic information. In this paper, a new method has been proposed to classify and to retrieve 3D model using semantic concepts and ontology. First, we use the machine learning methods to label 3D models by k-means algorithm in shape indexes space. Second, classification of 3D models into classes based on semantic annotation is applied. Finally, ontology is constructed with instances for each class, which contain spatial relationships, shape indexes and measures. Therefore, the 3D model retrieval system is comprised of low-level and high-level visual features. We interpret our results using the Princeton Shape Benchmark Database and our prototypical 3D Model Search System (3DMSS). The results show the performance of the proposed new approach to classify and to retrieve 3D models.

Keywords: *3D Model, Classification, 3D retrieval, shape indexes, semantic, ontology.*

1. Introduction

The size of 3D models that is used on the web or stored in databases is becoming increasingly high. To increase the identification rate and decrease the time to search for items, different methods have been developed for the classification of 3D models, using geometrical characteristics. However, most of these methods do not include semantic information. Therefore, 3D model retrieval system has been affected by the similarity gap between the lower and the higher level features. In this paper, two ways are used to reduce the semantic gap: First, unsupervised learning method is used to create the link between shape indexes features and the semantic concepts. Then, the classification of 3D models which is based on semantic annotation is applied. Second, using OWL ontology to define the concepts of 3D models, the spatial relationships are applied to disambiguate among models of similar appearance. 3D model retrieval system which is based on the semantic and ontology is developed through the use of 3D shape indexes and spatial relationships represented by concepts in ontology. There are two motivations for using shape indexes: For the first motivation, shape indexes, which are all normalized, are frequently used to quantify different aspects of 3D model shape. Concerning the second motivation is to extract visual concepts easily, and semantic information can be extracted using unsupervised learning method.

2. Related work

Several systems and approaches for the classification of 3D models have been proposed in the literature. Chin-Chia Wu and al. [1] proposed the new approach for classifying 3D models in points clouds based on geometric graph representation. The approach uses a RIMLS technique and spin image signature to calculate the geometric characteristic. Based on the spatial clustering ontology, the authors in [3] developed ontology-based spatial clustering and reasoning system. This system integrates domain knowledge and user goals into clustering. Maria and Mihai [2] proposed the classification method based on clustering algorithm for body shape recognition. For the 3D model-semantic problem, many approaches have been proposed. The work presented in [4] introduces the classification and retrieval 3D model by integrating shape features and semantic information. The paper proposes a new type of shape feature based on 2D views and use Gaussian processes as supervised learning to mode the mapping from low-level features to query concepts. In the paper [5], the author explores a new framework for 3D model retrieval based on an off-line learning of the most salient features of the shapes. The proposed approach uses a large set of features, which can be heterogeneous, to capture the high-level semantic concepts of different shape classes. Hou and al, in [6] Support Vector Machine (SVM) is used to cluster 3D models with respect to semantic information for the organization a database of shapes. Therefore, the classification and retrieval 3D model system is integrated

[7]. Also the semantic description of an object based on the ontology, the matching of this description with the low-level features such as color, texture, shape and spatial relationships [8] [9] 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.

In this paper, we suggest reducing the semantic gap in two steps during 3D model retrieval process: the first one aims at classifying 3D models based on semantic concepts. To label 3D models, in this step, the machine learning methods is applied in shape indexes space. Second, we use SPARQL engine to question the information displayed in OWL ontology using spatial relationships.

3. System overview

In this paper, our content-based search (3DMSS) is used to test the proposed classification and retrieval. The proposed 3DMSS is illustrated in figure ("Fig 1").

In the inline process, the user can navigate in the database and sends a 3D request to the server. The 3D model retrieval procedure is a three-step process: first, 3DMSS defines the class membership which is semantically classified. Second, sends the request to ontology by SPARQL engine using spatial relationships. Finally, compares its descriptor with the descriptors of all models selected in second step using geometrical characteristics.

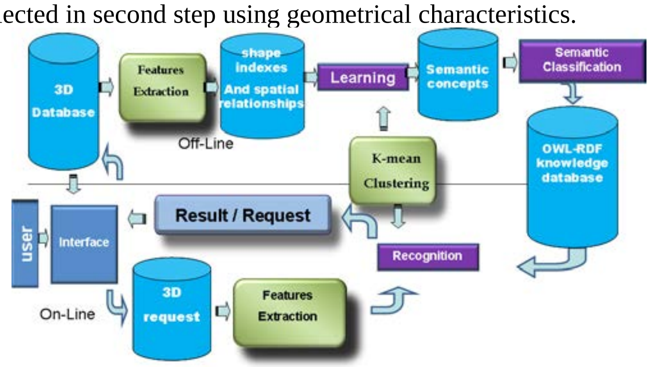


Fig 1. Overview of the proposed system

Our 3D Database is composed of Princeton Shape Benchmark 3D models [12] that are stored in a format (*.off) which represents the shape of 3D models by polygonal mesh.

4. Descriptor Extraction

The size and shape of 3D model have been used to describe 3D model. There are several ways to describe a 3D model shape. However, no single shape descriptor is appropriate for all models. Therefore, such a way of characterizing shape is required. Shape indexes that provide Sphericity, Compactness, Convexity and elongation, are frequently used to quantify different aspects of 3D model shape ("Fig 2"). In addition, motivation for using shape indexes is extracting visual concepts easily and tells a lot about the semantics of the 3D model. Therefore, the association of shape indexes with semantic concepts should be used.

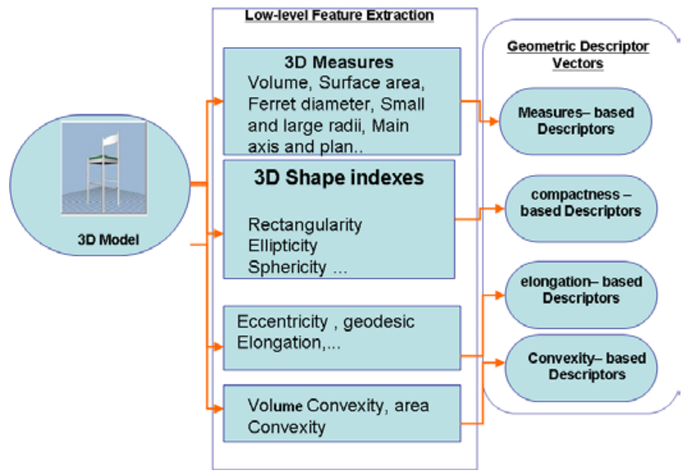


Fig 2. Descriptor Extraction

To describe 3D model size, various measures from this 3D model are calculated. The most frequently used are the diameter and the length measure of the three representative main axes. The measure equivalent spherical diameter (ESD) can be used to define 3D model size. It is the diameter of a sphere of equivalent volume that gets larger or smaller as the model does. Let S_{ch} the convex hull surface area. The ESD is defined by:

$$ESD = (4/3) * \pi * \sqrt{(S_{ch} / \pi)^3} \qquad (1)$$

To compute 3D shape indexes, the most important 3D measures are: surface area and volume. With 3D polygonal model representation, these measures [15] can be readily obtained from the models and are used directly for calculating 3D shape indexes. For other 3D measures, we use the 2D measures transformed by modifying the calculation. In practice, we used the following measures: Volume, Surface area, diameter, Ferret diameter, Small and large radii, main axis, plan and ESD. Similar measures are calculated from bounding box and convex hall that is the minimum enveloping boundary. These measures are used as semantic concepts in ontology and allow to define the spatial relationships.

From these basic measures, multiple shape indexes are calculated for each 3D model and subdivided into four groups: Compactness, Sphericity, Convexity and Elongation. For others shape indexes see our works in [14].

4.1 Compactness

From ESD measure, new shape index that provides compactness indicator can be calculated [9] as follows:

$$C1 = ESD / D_{ch} \qquad (2)$$

Where D_{ch} is the convex hall diameter.
Based on volume (V) and surface area (S) ratio, other shape index characterizing the Compactness of 3D model have been calculated as follows:

$$C2 = (V^2 * 36 * \pi) / S^3 \qquad (3)$$

4.2 Sphericity

Sphericity and roundness have been used to represent the 3D model shape and are indications of compactness. Sphericity is a shape index that provides how spherical an object is. Roundness is related to angularity and represents the curvature of model's corners. New shape index characterizing the Sphericity of 3D model has been calculated [24] as follows:

$$S1 = a * V^{2/3} / S \qquad (4)$$

Where a = 62/3 π1/3 ≅ 4.84 (which makes the S1 equal to 1 for a sphere).
The Sphericity index is a dimensionless constant with values ranging from zero for a laminar disc to unity for a sphere and it is most sensitive to elongation.

4.3 Elongation

Elongation provides an indication of 3D model overall form by comparing the strength of the major axis and the strength of the minor axis of a 3D model.
It is defined as (1-[width/length]) and has values in the range of zero to one. We are using main axes and radii to compute elongation. We can use also moments to compute them as in [16].

4.4 Convexity

This shape indexes group computes and provides the surface roughness of a 3D model and is calculated by dividing the convex hull surface area by a surface area of 3D model [14].

$$C_s = S_{ch} / S \qquad (5)$$

We can also calculate convexity shape index using 3D model volume as follows:

$$C_v = V_{ch} / V \qquad (6)$$

A smooth shape, regardless of form, has a convexity of 1 while a very 'spiky' or irregular object has a convexity closer to 0.
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, the important idea is extracting semantic concepts easily from shape indexes to classify 3D models. Measures and shape indexes are considered descriptors in this paper and others definitions are detailed in [13] [14].

5. Semantic-based classification

The next step after features extracting is to classify 3D models semantically. The problem is assigning an appropriate class to the query model. The proposed 3D model classification contains a semantic labeling step and a classifying step. In first one, we exploit shape indexes for semantic labeling and we use machine learning methods to associate shape indexes with semantic concepts as shown in figure 3. In this paper, shape indexes are used to represent 3D model visual concepts [17].

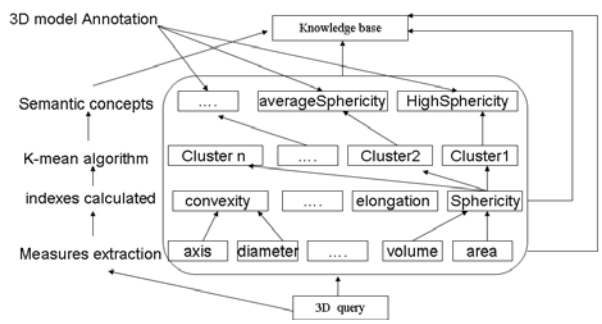


Fig 3. Definition of semantic concepts and Knowledge base augmented and guided by a 3D Shape index ontology to describe the 3D Models

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 figure ("Fig 3"). 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.
The second step is classifying 3D models based on these semantic concepts. In general, classification is done after training. In this paper, k-nearest neighbor algorithm (k-NN) is used as a method for classifying 3D models. Based on closest training examples in the feature space ("Table 1"), k-NN classifies an object by a majority vote of its

neighbors, with the object being assigned to the class most common amongst its k nearest neighbors. Euclidean distance is used as the distance metric. For example, Sphericity and Ellipticity IDs are based on the following table:

Table 1. Semantic concepts IDs

ID	Ellipticity	Sphericity
0	"High Ellipticity"	"High Sphericity"
1	"Average Ellipticity"	"Average Sphericity "
2	"small Ellipticity"	"small Sphericity "
3	"smaller Ellipticity"	"smaller Sphericity "

In this example, the semantic concept is assigned to ID in semantic labeling step and applied to all semantic concepts. We should create a database to describe 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-based retrieval

6.1 Ontology

After building the classes based on semantic, the next step to reduce the semantic gap is constructing the ontology for each class. Ontology is employed to organize semantic concepts (e.g. Sphericity, elongation, convexity...) and other concepts such as semantic entities (e.g. lines, points, surface, and plan). This Ontology also comprises set of spatial relations and some axioms (transitivity, reflexivity, symmetry). The proposed ontology is represented in Ontology Language OWL [18].
SPARQL is used to request this ontology and the result is considered as the second method for the classification and the selection of 3D model in one class membership. Finally, to evaluate the similarity between two 3D models in all models selected based on visual concepts; appropriate distance by numeric value is required. In this paper, measures, shape indexes and moments (section 6.2) are used as numeric value to compute similarity between two 3D models using Euclidean distance.

6.2 Spatial relationships

To describe spatial relationships that are usually defined according to the location of the measure in the 3D model, we calculated the local characteristics from convex hull and bounding box. Therefore, four points from convex hull are considered ("Figure 4"): centroid (P1), the closest point to centroid (P2), the farthest point to centroid (P3), and the farthest point to P3 (P4). In the first step, the center of gravity G (or P1) is located. Then, we calculate other points using Euclidean distance.

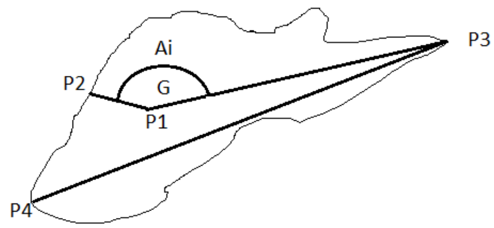


Fig 4. Four points to describe spatial relationships

The maximum P1P3 and minimum distance P1P2 (respectively Small and large radii) are used to define the bounding box and there ratio is used as a dimensionless shape index.
From these points, we can compute polygons [23], lines, plans. Then, ratios computed between surface area and volumes from polygons are inserted in ontology as relationships. For each point Pi we calculate also the angle Ai, this angle allow to compute various directions. This set of features allows the description of model independently of their size, rotation, translation or line type [23]. During the process of calculating the four sets of point distances (P1, P2, P3, P4), the moments of distributions are calculated (12 moments of the resulting four distributions) [22] and are used to compute similarity via numeric value.
From bounding box three main axes are considered to describe position, distances and orientation of an entity in the 3D model. Therefore, several relationships are described ("Figure 5"):

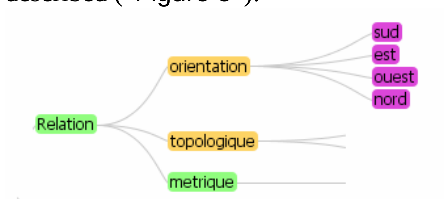
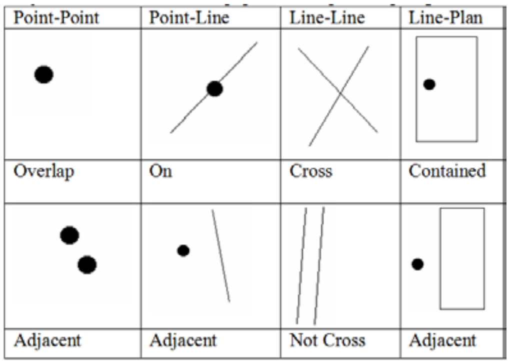


Fig 5. Partial hierarchy of relationships

The distances can be computed from a point to point, line to line, point to line, point to plan and line to plan. Also we are 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 ("Table 2"). The RCC-8 [19] [20] relations can be used for taking into account spatial relations. RCC (Region Connection Calculus) is a logic-based formalism symbolically to represent and reason with topological properties of objects [21].

Table 2. Topological relations implemented in our system



Based on the spatial relationships and their properties, we build the ontology using the web ontology language (OWL).

7. 3D model retrieval

The third step in our 3D model retrieval system is to compute de similarity between two 3D models as shown in figure ("Figure 6").

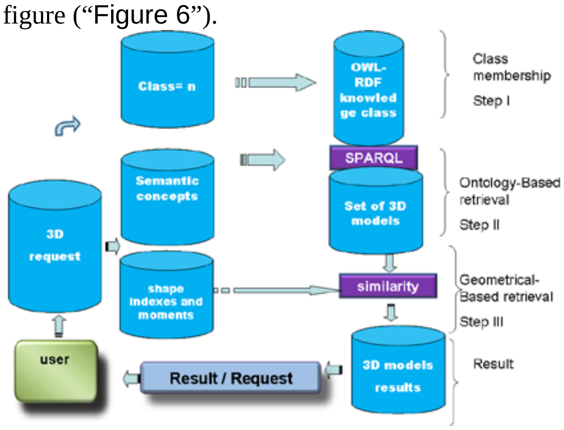


Fig 6. semantic and numeric query to evaluate the similarity

The similarity between two models is measured through the use of a distance between their measures, 3D shape indexes and the moments of distributions calculated during the process of calculating the four sets of point distances (P1, P2, P3, and P4).
In our system, the Euclidean distance is used to measure the similarity between 3D models. Therefore, to provide the best results, it is necessary to combine shape indexes, measures and moments to compute the most relevant. A simple approach for combining these descriptors is to compute the weighted sum of the distances [14].

8. Experimental Results

We are using Java language to develop our content-based retrieval systems for 3D models [14]. The average times that are used to compute all measures, shape indexes, moments and relationships is 0,114 seconds for a model, using the Princeton Shape Benchmark Database ("Table 3").

Table 3. Time to compute all descriptors

Example model	Number of edges	number of polygons	Number of vertices	Time
1	1638	546	341	0,128405697
2	648	228	100	0,131520481
3	1476	492	298	0,085323701
4	1224	408	216	0,234916242
5	40626	13542	7056	0,80845978
6	3336	1112	557	0,222504762
7	50925	16975	8469	0,774157548
8	5637	1879	781	0,24777793

During the process on line, all features are computed, and we can directly retrieval models as shown in figure 7 and 8 ("Fig 7, Fig8").
The 12 most similar models are extracting and return to do a user by 2D images. To visualize the 3D models in the 3D space, the user clicks the button or image.

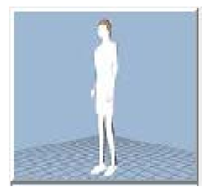


Fig 7. query model

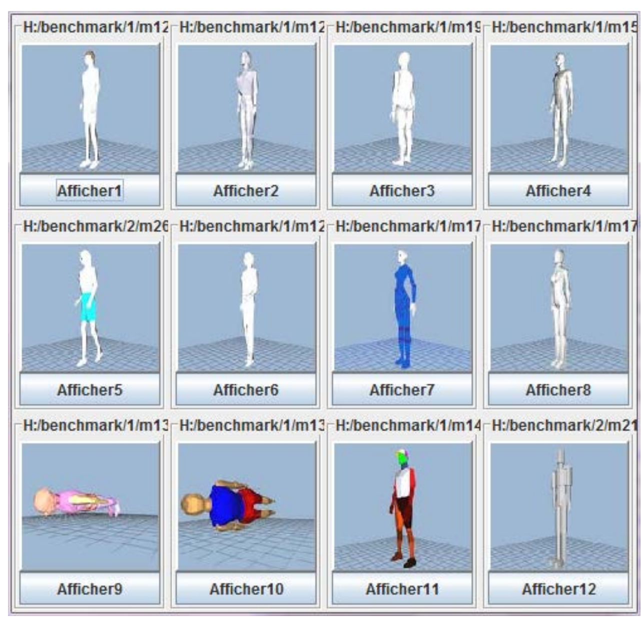


Fig 8. Models found with our visual descriptor

Fig 9. query model

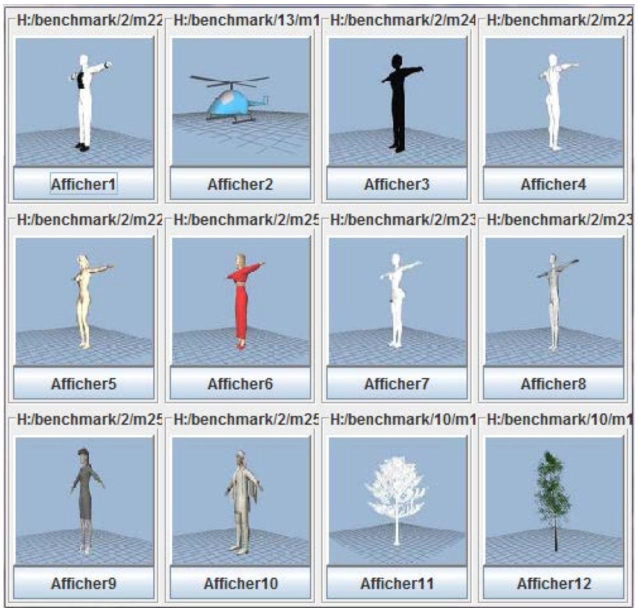


Fig 10. Models found without classification without ontology

Fig 11. Models found with classification and without ontology

To classify 3D models by introducing the semantic descriptor ("Fig 10, Fig 11, Fig 12"), 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.

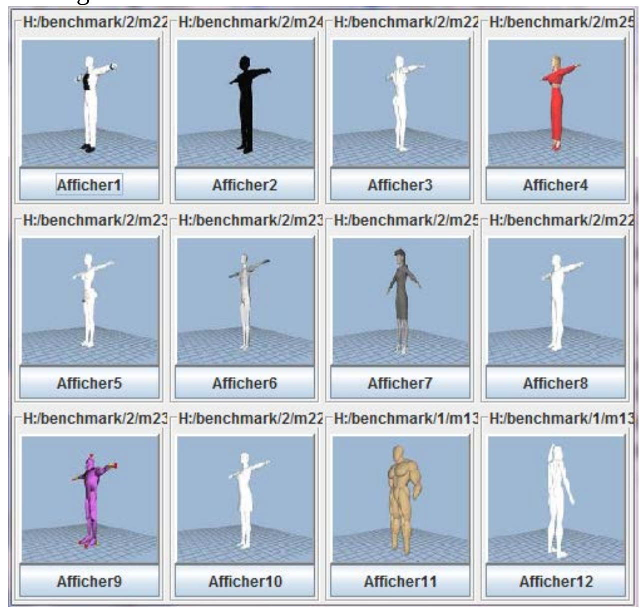


Fig 12. Models found with classification and ontology

Figure ("Fig 11") and other examples show that our 3D classification based on semantic concepts is the important method to classify 3D model justifying the importance of using the shape indexes and measures as visual concepts.

The evaluation of our system consists of two main steps: shape indexes effectiveness and 3D model retrieval. For the first one, we evaluated the effectiveness of integrating new shape indexes to label 3D model and we compared the retrieval performances of the shape indexes at different groups. In this experiment, combining Sphericity, convexity and elongation gives the most reliable results.

Concerning second step, 3D Harmonics and Moments are implemented. We used the Recall and Precision to compare different descriptors. Figure ("Fig 13"). shows that our proposed descriptor performs well.

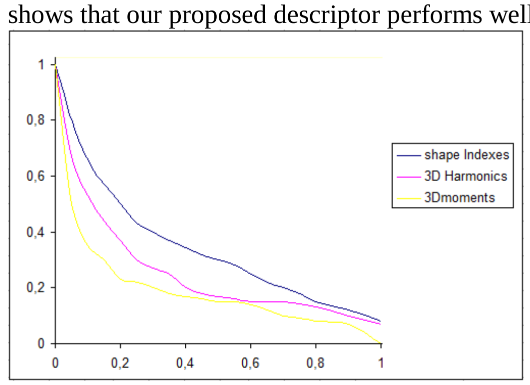


Fig 13. The precision-recall curves of our system

The classification based on semantic concepts developed reduces the similarity gap, and the retrieval method by introducing the ontology retrieval is considered as the second method to classify 3D model in one class membership. Both methods to classify 3D models allow our system more efficient. This performance is also linked to the combination of shape indexes and semantic concepts structured in ontology.

9. Conclusion

A new method for 3D models classification and retrieval is introduced in this paper. First, the 3D classification that has been based on semantic concepts is proposed. Second, the method combines semantic concepts and geometrical characteristics which are structured in ontology to 3D model retrieval. The new approach is tested with a large 3D database using the search engine developed, which allows us to show the relevance of our method. The results are promising and show the interest of our approach.

References

[1] Chin-Chia Wu and Sheng-Fuu Lin “Content-based 3D Model Classification using Geometric Graph Representation” (IJAEST) Vol No. 5, Issue No. 2, 116 - 122 April, 2011

[2] Aileni Raluca Maria and Ciocoiu Mihai “3D shape recognition software used for classification of the human bodies” 5 th International Conference on Virtual Learning ICVL 2010.

[3] Wei Gu , Xin Wang , and Danielle Ziébelin “An Ontology-Based Spatial Clustering Selection System” In Proceedings of Canadian Conference on AI'2009.

[4] Boyong GAO, Sanyuan ZHANG and Xiang PAN “Semantic-Oriented 3D Model classification and Retrieval Using Gaussian Processes” in Journal of Computational Information Systems 7:4 2011.

[5] Hamid Laga “3D Shape Classification and Retrieval Using Heterogenous Features and Supervised Learning” InTech, January 2009.

[6] HOU S., LOU K., RAMANI K.: “Svm-based semantic clustering and retrieval of a 3d model database”. In Journal of Computer Aided Design and Application 2005.

[7] J. Pu, K. Ramani, An integrated 2D and 3D shape-based search framework and applications, Computer-Aided Design and Applications 4 (6) 2007.

[8] 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.

[9] 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,

[10] R. Ohbuchi, K. Osada, T. Furuya, T. Banno, “Salient local visual features for shape-based 3D model retrieval”, Proc. SMI ‘08, 93-102, (2008).

[11] 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 Repulic, p. 25-28, February, 2008.

[12] P. Shilane, P.Min,M. Kazhdan, et T. Funkhouser. “The princeton shape benchmark“. In Shape Modèleing International, June 2004.

[13] 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.

[14] A. Kassimi and O. El beqqali. “3D Model Retrieval Based on Semantic and Shape Indexes” IJCSI, Vol. 8, Issue 3, No. 1, May 2011

[15] 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.

[16] Milos Stojmenovic, Amiya Nayak. Measuring the Related Properties of Linearity and Elongation of Point Sets. In Proceedings of CIARP' 2008.

[17] Boder, M.: “Mpeg-7 visual shape descriptors“. IEEE Transactions on Circuits and Systems For Video Technology 716-71911 2001

[18] OWL web ontology language guide. http://www.w3.org/tr/2004/rec-owl-guide-20040210/, w3C Recommendation (Feb 2004).

[19] D. A. Randell, Z. Cui et A. G. Cohn. œ “A Spatial Logic based on Regions and Connection”. In : 3rd International Conference on Principles of Knowledge Representation and Reasoning. pp. 165œ176. œ Morgan Kaufmann Publishers, 1992.

[20] Cohn AG, Hazarika SM “Qualitative Spatial Representation and Reasoning: An Overview“. Fundamenta In-formaticae 46(1-2):1629. 2001

[21] P. L. Rosin. “Computing global shape measures”. Handbook of Pattern Recognition and Computer Vision, 3rd edition, page 177:196, 2005.

[22] Pedro J.Ballester and W. Graham Richards “Ultrafast Shape Recognition to Search Compound Databases for Similar Molecular Shapes” in Wiley InterScience Published online 6 March 2007.

[23] Manuel J. Fonseca, Alfredo Ferreira, and Joaquim A. Jorge “Generic Shape Classification for Retrieval” In Proceedings of GREC' 2005

[24] P. B. CANHAM and ALAN C. BURTON “Distribution of Size and Shape in Populations of Normal Human Red Cells” in Circulation Research 1968

My abdellah Kassimi is a PhD student at Sidi Med Ben AbdEllah University (GRMS2I 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 'GRMS2I' research group since 2005 (Information Systems engineering and modeling) of USMBA and the Research-Training PhD Unit 'SM3I'. 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).

Dual Transform based Feature Extraction for Face Recognition

Ramesha K[1], K B Raja[2]

[1]**Department of Telecommunication Engineering, Vemana Institute of Technology, Koramnagala Bangalore-560034, India**

[2]**Department of Electronics and Communication Engineering, University Visvesvaraya College of Engineering, Bangalore University, K R Circle, Bangalore-560001, India**

Abstract

Face recognition is a biometric tool for authentication and verification, has great emphasis in both research and practical applications. Increased requirement on security, fully automated biometrics on personal identification and verification has received extensive attention over the past few years. In this paper Dual Transform based Feature Extraction for Face Recognition (DTBFEFR) is proposed. The images from database are of different size and format, and hence are to be converted into standard dimension, which is appropriate for applying DT-CWT. Variation due to expression and illumination are compensated by applying DWT on an image and also DWT reduces image dimension by decomposition. The DT-CWT is applied on LL subband, which is generated after two-level DWT, to generate DT-CWT coefficients to form feature vectors. The feature vectors of database and test face are compared using Random Forest, Euclidian Distance and Support Vector Machine matching algorithms. The correct recognition rate, false reject rate, false acceptance rate and efficiency are better in the case of proposed method as compared to existing techniques.

Key Words: *Face Recognition, Histogram Equalization, Discrete Wavelet Transform, Random Forest, Dual Tree Complex Wavelet Transform, Support Vector Machine.*

1. INTRODUCTION

Biometrics refers to a science of analyzing human body parts for security purposes and the word is derived from the Greek words [1] *bios* (life) and *metrikos* (measure). Most of the biometric system employed in real-world application is unimodal; they depend on the evidence of a single source of information for authentication. Biometric identification is becoming more popular now a day's, due to the existing security requirements of society in the field of information, business, military, e-commerce and etc. In the mid-19th century, criminal identification division of the police department in Paris [2], developed and practiced the idea of using many body characteristics to identify criminals. Since then fingerprint recognition technique emerged rapidly in law enforcement to identify the criminals. The different techniques for recognition of a person is based on (i) physiological characteristics such as fingerprint, face, iris, retinal blood vessel patterns, hand geometry, vascular pattern, and DNA, and (ii) behavioral characteristics such as voice, signature and keystroke. The verification of a person using biometrics is more secured since, biometric parameters are the parts of human body hence cannot be stolen and/or modified, compared to traditional systems such as Personal Identification Number (PIN), passwords, smartcards etc. Face recognition is a nonintrusive method, and facial images are the most common biometric characteristic used by humans to make a personal recognition. The popular approaches for face recognition are based on either: (i) The location and shape of facial attributes such as the eyes, eyebrows, nose, lips and chin, and their spatial relationships, or (ii) The overall analysis of the face image that represents a face as a weighted combination of a number of canonical faces. Some of the face recognition systems are commercially available and their performance is reasonably good but they impose some restrictions on variation such as illumination, expression, pose, occlusions.

In general the face recognition system consists of *(i) Preprocessing unit*: In which the color image may be converted into gray scale image, the image may be resized to the required size, illumination normalization [3, 4, 5, 6] may be done using Histogram Equalization (HE), expression normalization using higher decomposition of Discrete Wavelet Transform (DWT) [7], pose normalization [8, 9], *(ii) Feature extraction:* In which, facial features are extracted using edge detection techniques [10], Principal Component Analysis (PCA) technique [11], Discrete Cosine Transform (DCT) coefficients [12, 13], DWT coefficients, Complex Wavelet Transform [14], Dual Tree Complex Wavelet Transform (DT-CWT) [15], fusion of different features [16], *(iii) Matching*: In which Euclidean Distance (ED), Hamming Distance, Support Vector Machine (SVM), Neural Network [17] and Random Forest (RF) [18, 19] may be used for matching.

2. RELATED WORK

Sidra Batool Kazmi et al., [20] presented an automatic recognition of facial expressions from face images. The features are extracted by performing three level 2-D discrete wavelet decomposition of image and further dimensionality reduction is achieved by performing Principal Component Analysis (PCA) and the features are supplied to a bank of five binary neural networks, each trained for a particular expression class using one-against-all approach. Neural network outputs are combined using a maximum function. Taskeed Jabid et al., [21] presented a novel local feature descriptor, the Local Directional Pattern (LDP), for recognizing human face. A LDP features are obtained by computing the edge response values in all eight directions at each pixel position and generating a code from the relative strength magnitude. Each face is represented as a collection of LDP codes for face recognition process.
Xiaoyang Tan and Bill Triggs [22] introduced image preprocessing stage based on gamma correction, Difference of Gaussian filtering and robust variance normalization. Local Binary Pattern (LBP) proves to be the best of these features, and we improve on its performance in two ways: by introducing a 3-level generalization of LBP, Local Ternary Patterns (LTP), and by using an image similarity metric based on distance transforms of LBP image slices.

Vaishak Belle et al., [23] presented a system for detecting and recognizing faces in images in real time which is able to learn new identities in instants. To achieve real-time performance used random forests for both detection and recognition, and compared with well-known techniques such as boosted face detection and SVM's for identification. Bo Du et al., [24] presented comparisons of several preprocessing algorithms such as Histogram Equalization, Histogram Specification, Logarithm transform, Gamma Intensity Correction and Self Quotient Image for illumination insensitive face recognition. Analysis is done on face databases CMU-PIE, FERET and CAS-PEAL. Zhi-Kai Huang et al., [25] presented a method for Multi-Pose Face Recognition in color images, to addresses the problems of illumination and pose variation. The color Multi-Pose faces image features were extracted based on Gabor wavelet with different orientations and scales filters, then the mean and standard deviation of the filtering image output are computed as features for face recognition. These features were fed into SVM for face recognition. Harin Sellahewa and Sabah A Jassim [26] presented the fusion strategy for a multi-stream face recognition scheme using DWT under varying illumination conditions and facial expressions. Based on experimental results, argue for an image quality-based, adaptive fusion approach to wavelet-based multi-stream face recognition.
K Jaya Priya and R S Rajesh [27] presented face recognition method on local appearance feature extraction using DT-CWT. Two parallel DWT with different low pass and high pass filters in different scales were used to implement DT-CWT. To generate six different directional subbands with complex coefficients, the linear combination of subbands generated by two parallel DWT is used. Face is divided into several non-overlapped parallelogram blocks. The local mean, standard deviation and energy of complex wavelet coefficients are used to explain face image. K Jaya Priya and R S Rajesh [28] proposed multi resolution and multi direction method for expression and pose invariant face recognition based on local fusion of magnitude of the DT-CWT detailed subband at each levels. Multi orientation information on each image is obtained by a subset of band filtered images containing coefficients of DT-CWT to characterize the face textures. The overall texture features of an image at each resolution are obtained through fusion of the magnitude of the detailed subbands. Fused subbands are divided into small subblocks, from these extracted compact and meaningful feature vectors using simple statistical measures. Yue-Hui Sun and Ming-Hui Du [29] proposed face detection using DT-CWT on spectral histogram. Laplacian of Gaussian filter and DT-CWT filter are used to capture spatial and frequency properties of human faces at different scales and different orientations. Then, the responses convolved with the filters are summarized to multi dimensional histograms. The histogram matrix is fed to the trained SVM for classification.

3. BACKGROUND

In this section necessary background for the work such as HE, DCT, DWT and DT-CWT are discussed.

3.1 Histogram Equalization (HE)

In image processing, the idea of equalizing a histogram [30] is to stretch and redistribute the original histogram using the entire range of discrete levels of the image to achieve good contrast image. HE modifies the dynamic range and contrast of an image by altering the image such that its intensity histogram has a desired shape. This uses a monotonic, non-linear mapping to re-assign the intensity values of pixels in the input image such that the output image contains a uniform distribution of intensities. This corresponds to a brightness distribution where all the values are equi-probable. For image *I(x, y)* with discrete *k* gray values histogram is given by Eq. (1)

$$P(i) = \frac{n_i}{N} \quad \text{.............................(1)}$$

where $i = 0, 1, 2, \ldots k-1$ grey level and N is total number of pixels in the image.

Transformation to a new intensity value is defined by Eq. (2)

$$I_{HE} = \sum_{i=0}^{k-1} \frac{n_i}{N} = \sum_{i=0}^{k-1} P(i) \text{.....................................(2)}$$

Output values are from domain of [0, 1]. To obtain pixel values in to original domain, it must be rescaled by the $k-1$ value. The enhanced image using HE is shown in Figure 1.

Fig. 1: (a) Original Image (b) HE Image.

3.2 Discrete Cosine Transform

The DCT [31] is the illumination normalization approach for face recognition under varying lighting conditions. The main idea of the method is to reduce illumination variation by eliminating low frequency coefficients in the logarithm DCT domain. The 2D *M*N* DCT is defined using Eq. (3)

$$C(u, v) = \alpha(u)\alpha(v) \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x.y)$$

$$\cos\left[\frac{\prod(2x+1)u}{2M}\right]\cos\left[\frac{\prod(2y+1)v}{2N}\right] \text{...................(3)}$$

and the inverse transform is given in Eq. (4)

$$f(u, v) = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} \alpha(u)\alpha(v)\cos(u, v)$$

$$\cos\left[\frac{\prod(2x+1)u}{2M}\right]\cos\left[\frac{\prod(2y+1)v}{2N}\right] \text{.....................(4)}$$

and the coefficients are given in Equations (5) and (6)

$$\alpha(u) = \begin{cases} \frac{1}{\sqrt{M}}, & u = 0 \\ \sqrt{\frac{2}{M}}, & u = 1,2,3,\ldots, M-1 \end{cases} \quad \text{.......................(5)}$$

$$\alpha(v) = \begin{cases} \frac{1}{\sqrt{N}}, & v = 0 \\ \sqrt{\frac{2}{N}}, & v = 1,2,3,\ldots, N-1 \end{cases} \quad \text{.......................(6)}$$

Normally illumination lies in the low frequency band, and it can be reduced by removing low frequency components. This is achieved by setting to zero, this works like an ideal high pass filter. The DCT coefficient determines the overall illumination of a face image, so, it sets to the same value to obtain the desired uniform illumination as in Eq. (7)

$$C(0.0) = \log\mu_0\sqrt{MN} \text{.......................................(7)}$$

The value μ_0 is chosen near the middle level of the original image. This approach, by discarding the DCT coefficients of the original image, only adjusts the brightness, whereas, by discarding the DCT coefficients of the logarithm image, adjusts the illumination and recovers the reflectance characteristic of the face.

3.3 2D-Discrete Wavelet Transform

The 2D-DWT of a signal *x* is implemented by iterating the 2D analysis filter bank on the lowpass subband image. Here, at each scale there are three subbands instead of one. There are three wavelets which are associated with the 2D wavelet transform. The Figure 2 illustrates three wavelets as gray scale images. The first two wavelets are oriented in the vertical and horizontal directions and the third wavelet does not have a dominant orientation. The third wavelet mixes two diagonal orientations and gives rise to the checkerboard artifact.

In spite of efficient computational algorithm and sparse representation, the DWT [32] having four fundamental disadvantages compared to DT-CWT. (i) *Oscillations*: Since wavelets are bandpass functions, the wavelet coefficients oscillate between positive and negative around singularities. This complicates wavelet-based processing and making singularity extraction and signal modeling. It is quite possible for a wavelet overlapping a singularity to have a small or even zero wavelet coefficient. (ii) *Shift Variance*: A small shift in input signal makes larger shifts in the wavelet coefficient oscillation pattern around singularities. Shift variance complicates wavelet-domain processing and developed algorithms must be made to cope up with the wide range of possible wavelet coefficient patterns caused due to shifted singularities. (iii) *Aliasing*: The wide spacing of the wavelet coefficient samples computed using iterated discrete-time down sampling operations with non ideal low-pass and high-pass filters, results in aliasing. The inverse DWT may cancel this aliasing, provided if the wavelet and scaling coefficients are not changed. Any wavelet coefficient processing such as filtering, thresholding and quantization upsets the delicate balance between the forward and inverse transforms, lead to reconstructed signal artifacts. (iv)*Lack of Directionality*: Fourier sinusoids in higher dimensions correspond to highly directional plane waves, the standard tensor product construction of M-D wavelets produces a checkerboard pattern that is simultaneously oriented along several directions. This lack of directional selectivity complicates modeling and processing of geometric image features such as ridges and edges.

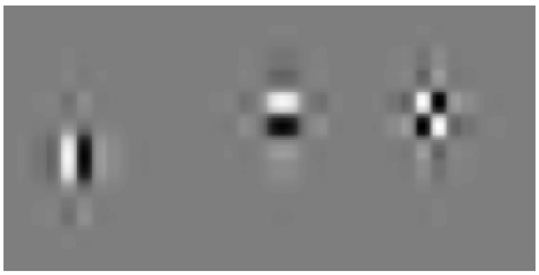


Fig. 2: 2D-Wavelets

3.4 Dual-Tree Complex Wavelet Transform (DT-CWT)

Fortunately, we have solutions to the four DWT shortcomings. Dual Tree Complex Wavelet Transform [33], a form of discrete wavelet transform which generates complex coefficients by using a dual tree of wavelet filters to obtain their real and imaginary parts. DT-CWT has the following properties to overcome the drawbacks of DWT:
1. Approximate shift invariance; 2. Good directional selectivity in 2-dimensions (2-D) with Gabor like filters also true for higher dimensionality: m-D); 3. Perfect reconstruction; 4.Limited redundancy: 2× redundancy in 1-D (2d for d-dimensional signals), this is less than the $\log_2 N\times$ redundancy of a perfectly shift-invariant DWT; 5. Efficient order N computation. DT-CWT introduces limited redundancy (2m:1 form-dimensional signals) and allows the transform to provide approximate shift invariance and directionally selective filters by preserving the properties of perfect reconstruction and computational efficiency with balanced frequency responses. The only drawback is a moderate redundancy:
The dual-tree complex DWT of a signal *x(n)* is implemented using two critically-sampled DWTs in parallel on the same data, as in Figure 3. To gain advantage over DWT, the filters designed in the upper and lower DWTs are different and are designed to interpret the subband signals of the upper DWT as the real part of a complex wavelet transform, and lower DWT as the imaginary part. When designed in this way the DT-CWT is nearly shift invariant, in contrast to the classic DWT. The DT-CWT is used to implement 2D wavelet transforms where each wavelet is oriented, and useful for image processing such as image denoising and enhancement applications.

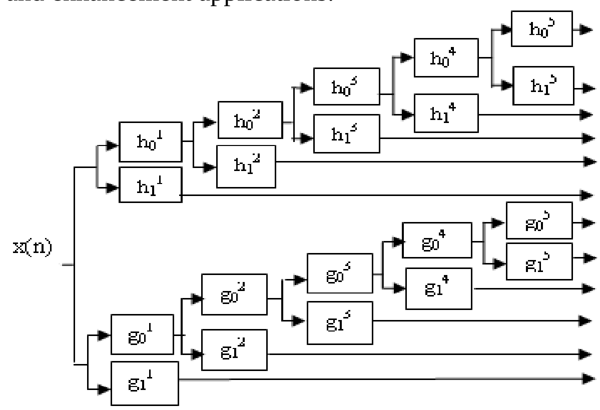


Fig. 3: DT-CWT structure

There are two types of the 2D dual-tree wavelet transform, they are real and complex. The real 2-D dual-tree DWT is 2-times expansive and the complex 2-D dual-tree DWT is 4-times expansive, and they are oriented in six distinct directions.

Real 2-D Dual-Tree Wavelet Transform: The real 2-D dual-tree DWT of an image *x* is implemented using two critically sampled separable 2D-DWTs in parallel. Then we take the sum and difference, for each pair of subbands. The six wavelets associated with the real 2D dual-tree DWT and are oriented in a different direction as illustrated in the Figure 4. Each subband of the 2-D dual-tree DWT corresponds to a specific orientation.

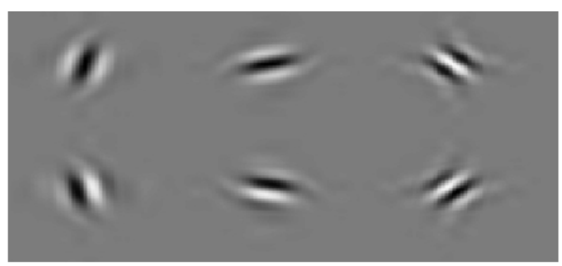


Fig. 4: Directional wavelets for reduced 2-D DWT

Complex 2-D Dual-Tree Wavelet Transform: The complex 2-D dual-tree DWT gives wavelets in six distinct directions and in this case there are two wavelets in each direction and become totally twelve wavelets as shown in Figure 5. In each direction, one of the two wavelets can be interpreted as the real part of a complex-valued 2D wavelet, and the other wavelet is interpreted as the imaginary part of a complex-valued 2D wavelet. The complex version is 4-times expansive due to the complex version have twice as many wavelets as the real version of the transform. Similar to real case, the sum and difference of subband images is performed to obtain the oriented wavelets.

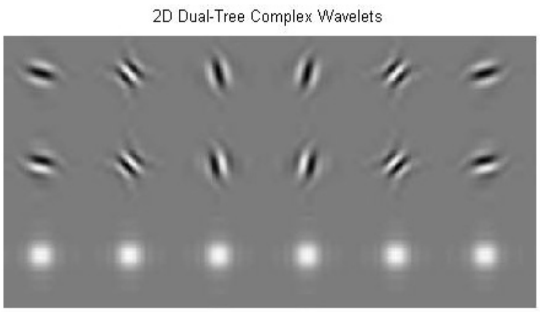


Fig. 5: 2D Dual-Tree Complex Wavelets

The six wavelets displayed on the first and second row are representing real and imaginary part of a set of six complex wavelets, and the third row represents the magnitude of the six wavelets. The magnitudes of the complex wavelets have bell-shaped behavior, rather than oscillatory behavior. The resulting complex wavelet is then approximately one sided in the frequency domain. It has the ability to differentiate positive and negative frequencies and produces six subbands oriented in $\pm 15^0$, $\pm 45^0$, $\pm 75^0$.

4. MODEL

In this section proposed DTBFEFR model is discussed. Face recognition using DWT and DT-CWT for illumination, expression and pose variations with different kind of databases is as shown in the Figure 6.

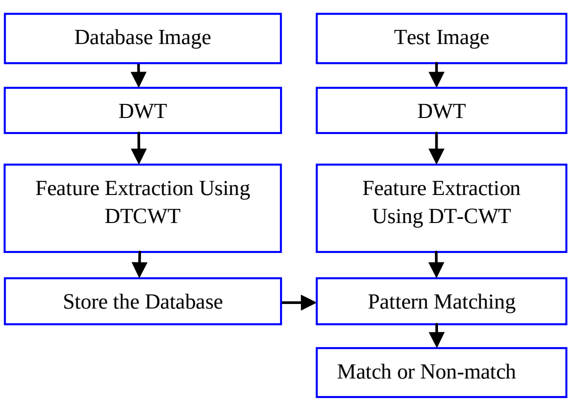


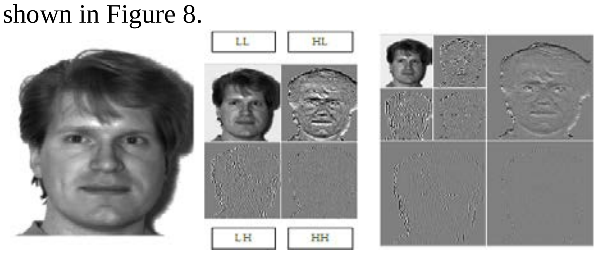
Fig. 6: Block diagram of the Proposed DTBFEFR model.

4.2 Proposed DTBFEFR Model

The face recognition using DWT and DT-CWT for illumination, expression and pose variations with different kinds of databases is as shown in the Figure 7.

Face Database: The available face Databases such as L-Spacek, ORL, Yale-B, JAFEE, and CMU-PIE are considered for experimental purpose. The image is preprocessed by resizing to $2^n * 2^n$, which is suitable for DT-CWT.

DWT: The DWT is used to decompose the original image into four subbands. The significant information of the original image is present in the low-low (LL) subband that represents the significant features of the original image. The two level DWT is applied on face image to achieve more reduction in significant information of an image is as shown in the Figure 7. The DWT removes expression and illumination variations and gives better contrast and resolution as shown in Figure 8.



(a) (b) (c)

Fig.7: (a) Original Image, (b) 1-level Wavelet Decomposition and (c) 2-level Wavelet Decomposition.

(a) (b)

Fig. 8: (a) Original Image (b) DWT Image

Feature Extraction by DT-CWT: DWT alone gives larger feature length and very low accuracy compared to DT-CWT. The DT-CWT overcome the drawbacks of DWT and provides more information with respect to the particular directions such as $\pm 15^0$, $\pm 45^0$, $\pm 75^0$.

The number of features and dimensions are reduces as the number of DT-CWT level increases. The number of features for level-1, level-2, level-3, level-4 and level-5 are 393216, 98304, 24576, 6144 and 1536 respectively, and corresponding reduction in image size is as shown in the Figure 9. In the proposed algorithm, 5-level DT-CWT is used to generate magnitude and phase features of length 1536. Reduction in number of features obviously reduces memory requirement and computational time.

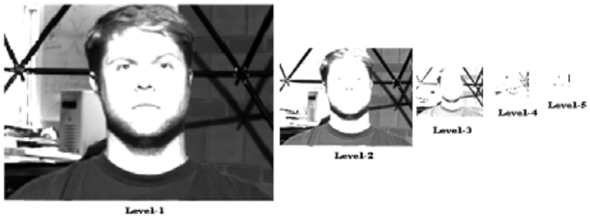


Fig. 9: DT-CWT images at different levels.

Face Classification: Matching techniques such as ED, RF and SVM are used for face classification to compute Correct Recognition Rate (CRR), False Acceptance Rate (FAR), False Rejection Rate (FRR) and Equal Error Rate (EER).

5. ALGORITHM

Prolem Definition: The biometric face recognition is verified by DWT and DT-CWT with multi-matching classifiers is proposed to identify a person.

The Objectives:

- The Face recognition using DTBFEFR.
- Reduce the value of FAR, FRR and EER.
- Increase the overall efficiency.

The DTBFEFR algorithm to verify a person using DWT and DT-CWT is given in the Table 1.

Table 1: Algorithm of DTBFEFR

- Input : Face Database, Test Face Image
- Output : Match/Mismatch Face Image

1. Read Face image from various databases
2. Resize the face images
3. Apply 2-level DWT and consider only LL subband
4. Use 5-level DT-CWT on LL subband to generate features
5. ED, RF and SVM are used as classifiers to verify a person.

6. RESULTS AND PERFORMANCE ANALYSIS

For performance analysis we used available five databases such as L-Spacek, JAFEE, ORL, Yale-B and CMU-PIE. The L-spacek database consists of 119 persons with 19 image samples per person. The database consists of 900 images with no background, light variation, hair style change and with minor variations in head turn, tilt and slant. To evaluate FRR and CRR, 100 persons with 18 image samples per person are considered to create database and one image per person is used as test image. To evaluate FAR, remaining 19 persons from the L-Spacek database is considered. The JAFEE face database consists of 10 persons, with 19 images per person. Here first 6 persons with 20^{th} image is considered to evaluate FRR and CRR and the remaining 4 persons images are considered to evaluate FAR. The ORL face database consists of 40 persons and each person with 9 images. First 30 persons with eight images per person are considered to evaluate FRR and CRR. To evaluate FAR remaining 10 persons are considered. The Yale-B database consists of 57 persons, each person with 30 images. To evaluate FRR and CRR first 30 persons are considered and to evaluate FAR remaining 27 persons are considered. The CMU-PIE face database consists of 68 persons with different modes such as illumination, lighting and talking conditions, each person with 45 images. To evaluate FRR and CRR first 35 persons are considered and to evaluate FAR remaining 33 persons are considered. Mixed (MX) face database is created by combining L-Spacek, ORL, JAFEE, Yale-B and CMU-PIE database images.

The percentage face recognition rate of the proposed algorithm is compared with other face recognition techniques using transformations such as DT-CWT, DT-CWT on Histogram Equalization and DT-CWT on DCT with different databases given in Table 2. It is seen that the proposed algorithm using DT-CWT on DWT gives 100% recognition compared to other transformation techniques.

Table 2: Comparison of Recognition Rate of the Proposed Algorithm with other Algorithms.

Data base	Algorithms			
	DT-CWT	HE+ DT-CWT	DCT+ DT-WT	DTBFEFR
L-Spacek	87.2%	96.7%	89.4%	100%
JAFEE	90.3%	92.3%	87.3%	100%
ORL	76.6%	95%	83.3%	91%
Yale-B	88.6%	96. 7%	88.3%	100%
CMU-PIE	99.7%	98.3%	97.6%	100%
MX	96.6%	98.3%	95%	100%

The performance evaluation parameters such as CRR, EER, Efficiency, FAR and FRR for different face database using DTBFEFR with different matching techniques viz., ED, RF and SVM are given in Tables 3, 4 and 5 respectively. For best performance the databases L-Spacek, Yale-B and Mixed images of sizes 64*128, 512 * 512 and 256 * 256 are considered respectively and also the databases JAFEE, ORL and CMU-PIE of image sizes 128 * 128 are considered for the proposed method DTBFR. In the case of RF classifier, the performance parameters CRR and FRR are better compared to ED and SVM. The CRR is 100% with all the three classifiers for databases such as L-Spacek, JAFEE, ORL and CMU-PIE, except Yale-B face database. It is observed that the CRR values are better in the case of RF compared to ED and SVM. The overall performance efficiency of the proposed algorithm is better in the case of ED compared to RF and SVM. The efficiency on different databases with three classifiers is given in the Table 6 and it is noticed that using ED classifier gives better recognition rate compared to RF and SVM.

Table 3: DTBFEFR using Euclidean Distance

Database	*CRR %*	*EER %*	*%η*	*FAR %*	*FRR %*
L-Spacek	100	0	100	0	0
JAFEE	100	0	100	0	0
ORL	100	9	91	9	9
Yale-B	83.3	0	100	0	0
CMU_Illum	100	0	100	0	0
CMU_Light	100	0	100	0	0
CMU_Talk	100	0	100	0	0
MX	100	0	100	0	0

Table 4: DTBFEFR using Random Forest

Database	*CRR %*	*EER %*	*%η*	*FAR %*	*FRR %*
L-Spacek	100	0.5	99.5	1	0
JAFEE	100	8.5	91.5	17	0
ORL	100	1.7	98.4	3.3	0
Yale-B	76.7	3.3	96.7	3.3	3.3
CMU_Illum	100	2.8	97.2	5.7	0
CMU_Light	100	0	100	0	0
CMU_Talk	100	1.8	98.2	2.9	0
MX	100	9	91	16	2

Table 5: DTBFEFR using Support Vector Machine

Database	*CRR %*	*EER %*	*% η*	*FAR %*	*FRR %*
L-Spacek	100	0	99	0	0
JAFEE	100	0	100	0	0
ORL	100	9	100	0	0
Yale-B	76.7	0	93.4	3.3	10
CMU_Illum	100	0	100	0	0
CMU_Light	100	0	100	0	0
CMU_Talk	100	0	100	0	0
MX	100	0	100	0	0

The variations of FAR, FRR and Efficiency with respect to the threshold for different face databases viz., L-Spacek, JAFEE, ORL, Yale-B, CMU-PIE_Illumination, CMU-PIE_Lights, CMU-PIE_Talking and Mixed databases are given in the Figures 10, 11, 12, 13, 14, 15, 16, and 17 respectively. It is noticed that the value of overall efficiency is less in the case of ORL face database as few face images are occluded by spectacles and other components.

The percentage recognition of the proposed DTBFEFR algorithm is compared with existing face recognition algorithms such as Dual Tree Complex Wavelet Transform based Face Recognition [34, 35, 36], Anisotropic Dual-Tree Complex Wavelet Packets (ADT-CWP) [34], Dual Tree Complex Wavelet Transform based Face Recognition with Single View [35] (ADT-CWT (μ, σ)), Face Detection using DT-CWT on Spectral Histogram (DT-CWTSH) [36] and Local Fusion DT-CWT [37] and it is given in Table 7.

It is observed that recognition rate in the proposed algorithm is about 100% for five databases except on ORL database.

Table 6: The Overall Experimental Efficiency of the DTBFEFR with Multi-matching Classifiers.

Database	**Classifiers**		
	ED	**RF**	**SVM**
L-Spacek	100%	99.5%	99%
JAFEE	100%	91.5%	100%
ORL	91%	98.44%	100%
Yale-B	100%	96.67%	93.4%
CMU-PIE	100%	98.44%	100%
MX	100%	91%	100%

Table 7: Comparison of the Proposed Algorithm with Other Algorithms

Algorithms	*% Recognition on Databases*				
	Yale-B	*CMU*	*ORL*	*Yale*	*MX*
DT-CWT [34,35, 36]	90.2	99.7	76.6	88.6	96.6
ADT-CWP[36]	97.5	99.8		95.3	92.4
Blocked Based DT-CWT(μ, σ)[35]	90.3	-	78.4	90.3	-
DT-CWTSH[36]	-	88	-	-	-
Local Fusion DT-CWT[37]	-	-	82.2	93.3	-
Proposed DTBFEFR	100	100	91	100	100

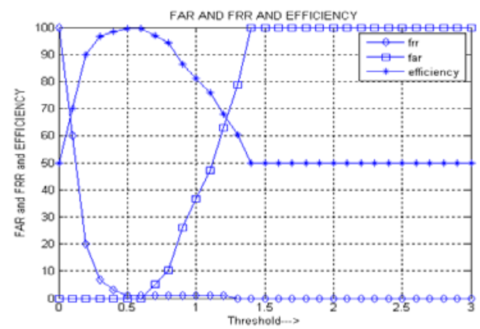


Fig.10: FAR, FRR and Efficiency versus Threshold using L-Spacek Database.

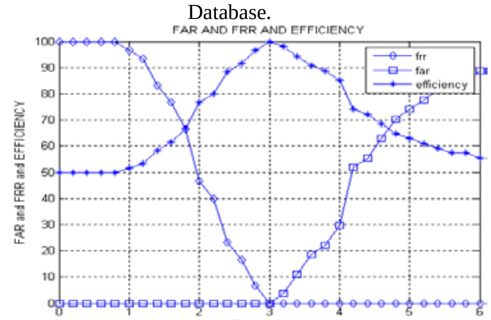


Fig.11: FAR, FRR and Efficiency versus Threshold using Yale-B Database.

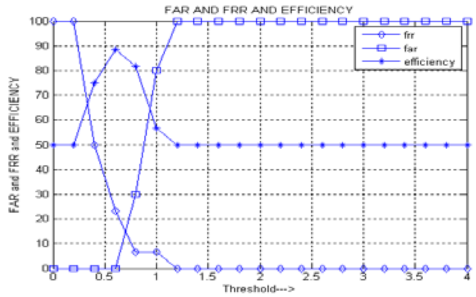


Fig.12: FAR, FRR and Efficiency versus Threshold using ORL Database.

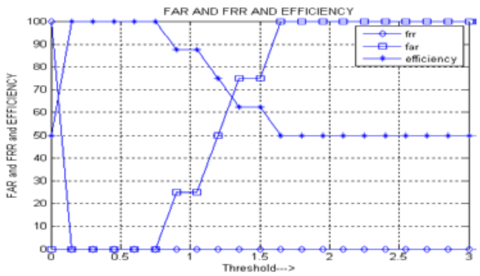


Fig.13: FAR, FRR and Efficiency versus Threshold using JAFEE Database.

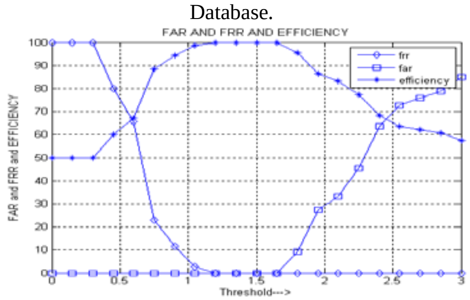


Fig.14: FAR, FRR and Efficiency versus Threshold using CMU-PIE Illumination Database.

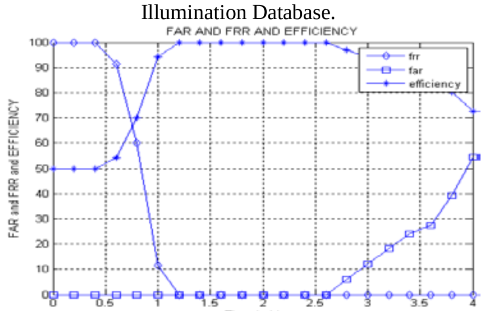


Fig.15: FAR, FRR and Efficiency versus Threshold using CMU-PIE Lights Database

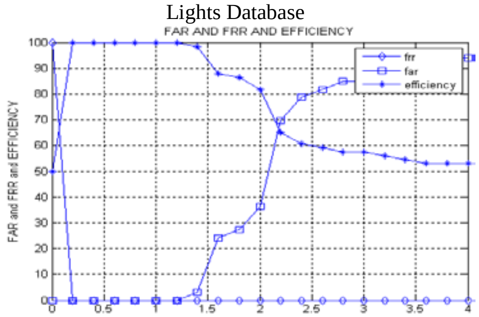


Fig.16: FAR, FRR and Efficiency versus Threshold using CMU-PIE Talking Database.

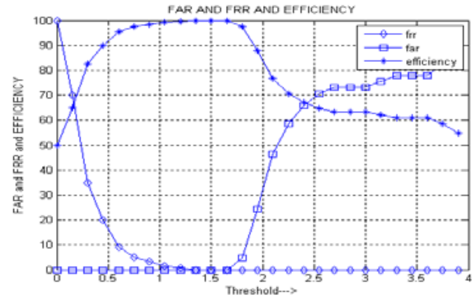


Fig.17: FAR, FRR and Efficiency versus Threshold using Mixed Database.

7. Conclusion

In this paper dual transform based feature extraction method is proposed for face recognition. The L-Spacek, ORL, Yale-B, JAFEE, CMU-PIE and Mixed database images are used to test the proposed algorithm. DWT is applied on Facial images to normalize expression and illumination variations, which greatly reduces image dimension by retaining visually significant components of an image. Further reduction in dimension is obtained by using DT-CWT, which gives out feature vectors of face image. Using feature vectors face image is verified using ED, SVM and RF. It is observed that the performance parameters are improved in the case of proposed algorithm compared to the existing algorithms. In future the algorithm may be tested by fusing features of individual transformations used and also fusion at matching level.

References

[1]. Marcos Faundez-Zanuy, "Biometric Security Technology," *Encyclopedia of Artificial Intelligence*, 2009, pp. 262-264.

[2]. Anil K Jain, Arun Ross and Salil Prabhakar, "An Introduction to Biometric Recognition," IEEE Transaction on Circuits and Systems for Video Technology, 2004, vol. 14, no.1, pp. 4-20.

[3]. A Abbas, M I Khalil, S Abdel-Hay and H M Fahmy, "Expression and Illumination Invariant Preprocessing Technique for Face Recognition," *Proceedings of the International Conference on Computer Engineering and System*, 2008, pp. 59-64.

[4]. Jukka Holappa, Timo Ahonen and Matti Pietikainen, "An Optimized Illumination Normalization Method for Face Recognition," *Proceedings of the IEEE Second International Conference on Biometrics: Theory, Applications and Systems*, 2008, pp. 1-6.

[5]. Sun Yuehui and Du Minghui, "Robust Face Recognition for Illumination Removal using DT-CWT and EMD," *Proceedings of the Nineteenth*

International Conference on Computational Science, 2008, pp. 357-361.

[6]. S Shan, W Gao, B. Cao, and D. Zhao, "Illumination Normalization for Robust Face Recognition against Varying Lighting Conditions," *Proceedings of the IEEE Workshop on Analysis and Modeling of Face Gestures*, 2003, pp. 157–164.

[7]. Shan Du and Rabab Ward, "Wavelet based Illumination Normalization for Face Recognition," *Proceedings of the IEEE International Conference on Information Processing*, 2005, vol. 2, pp. 954 - 957.

[8]. Shan Du and Rabab Ward, "Component-Wise Pose Normalization for Pose-Invariant Face Recognition," *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2009, pp. 873-876.

[9]. Yali LI, Shengjin WANG and Xiaoqing DING, "Person-Independent Head Pose Estimation based on Random Forest Regression," *Proceedings of IEEE Seventeenth International Conference on Image Processing*, 2010, pp. 1521-1524.

[10]. Ramesha K, K B Raja, Venugopal K R and L M Patnaik, "Feature Extraction based Face Recognition, Gender and Age Classification," *International Journal on Computer Science and Engineering*, 2010, vol. 02, no.01S, pp. 14-23.

[11]. Ramesha K and K B Raja, "Face Recognition System Using Discrete Wavelet Transform and Fast PCA," *Proceedings of the International Conference on Advances in Information Technology and Mobile Communication*, CCIS 157, 2011, pp. 13-18.

[12]. Marios Savvides, Jingu Heo, Ramzi Abiantun, Chunyan Xie and Vijay Kumar, "Class Dependent Kernel Discrete Cosine Transform Features for Enhanced Holistic Face Recognition in FRGC-II," *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 2006, vol. 5, no. 2, pp. 185-188.

[13]. Derzu Omaia, Jankees V D Poel and Leonardo V Batista, "2D-DCT Distance based Face Recognition using a Reduced Number of Coefficients," *Proceedings of Twenty Second Brazilian Symposium on Computer Graphics and Image Processing*, 2009, pp. 291-298.

[14]. Alaa Eleyan, Huseyin Ozkaramanli and Hasan Demirel, "Complex Wavelet Transform based Face Recognition," *Eurasip Journal on Advances in Signal Processing*, 2008, vol. 2008, pp. 1-13.

[15]. Chao-Chun Liu and Dao-Qing Dai, "Face Recognition using Dual-Tree Complex Wavelet Features," *IEEE Transactions on Image Processing*, 2009, vol. 18, issue 11, pp. 2593-2599.

[16]. Gunawan Sugiarta Y B, Riyanto Bambang, Hendrawan and Suhardi, "Feature Level Fusion of Speech and Face Image based Person Identification System," *Proceedings of the Second International Conference on Computer Engineering and Applications*, 2010, pp. 221-225.

[17]. S S Ranawade, "Face Recognition and Verification using Artificial Neural Network," *International Journal of Computer Applications*, 2010, vol.1, no. 14, pp. 21-25.

[18]. Albert Montillo and Haibin Ling, "Age Regression from Faces using Random Forests," *Proceedings of the IEEE International Conference on Image Processing*, 2009, pp. 2465-2468.

[19]. Gerard Biau, Luc Devroye and Gabor Lugosi, "Consistency of Random Forests and Other Averaging Classifiers," *Journal of Machine Learning Research*, 2008, vol. 9, pp. 2015-2033.

[20]. Sidra Batool Kazmi, Qurat-ul-Ain and M Arfan Jaffar, "Wavelets based Facial Expression Recognition using a Bank of Neural Networks," *Fifth International Conference on Future Information Technology*, 2010, pp. 1-6.

[21]. Taskeed Jabid, Md. Hasanul Kabir, and Oksam Chae, "Local Directional Pattern for Face Recognition," *International Conference on Consumer Electronics*, 2010, pp. 329 – 330.

[22]. Xiaoyang Tan and Bill Triggs, "Preprocessing and Feature Sets for Robust Face Recognition," *Proceedings of the IEEE International Conference on Computer Vision and Pattern Analysis*, 2007, pp. 1-8.

[23]. Vaishak Belle, Thomas Deselaers and Stefan Schiffer, "Randomized Trees for Real-Time One-Step Face Detection and Recognition," *Proceedings of the Nineteenth International Conference on Pattern Recognition*, 2008, pp.1-4.

[24]. Bo Du, Shiguang Shan, Laiyun Qing and Wen Gao, "Empirical Comparisons of Several Preprocessing Methods for Illumination Insensitive Face Recognition,"*IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2005, vol. 2, pp. 981-984.

[25]. Zhi-Kai Huang, Wei-Zhong Zhang, Hui-Ming Huang and Ling-Ying Hou, "Using Gabor Filters Features for Multi-Pose Face Recognition in Color Images," *Proceedings of the IEEE Second International Conference on Intelligent Information Technology Application*, 2008, pp. 399-402.

[26]. Harin Sellahewa and Sabah A Jassim, "Illumination and Expression Invariant Face Recognition: Toward Sample Quality-based Adaptive Fusion," *Proceedings of the IEEE Second International Conference on Biometric Theory, Applications and Systems*, 2008pp. 1-6.

[27]. Priya K J and Rajesh R S "Local Statistical Features of Dual-Tree Complex Wavelet Transform on Parallelogram Image Structure for Face Recognition with Single Sample," *Proceedings of the International Conference on Recent Trends in Information, Telecommunication and Computing*, 2010, pp. 50-54.

[28]. K Jaya Priya and R S Rajesh, "Local Fusion of Complex Dual-Tree Wavelet Coefficients based Face Recognition for Single Sample Problem," *Proceedings of the International Conference and Exhibition on Biometrics Technology*, 2010, vol. 2, pp. 94-100.

[29]. Yue-Hui Sun and Minghui Du, "DT-CWT Feature Combined with ONPP for Face Recognition," *Published by Springer-Verlag Berlin, Heidelberg in the book of Computational Intelligence and Security*, 2007, pp. 1058 – 1067.

[30]. Mariusz Leszczyński, "Image Preprocessing for Illumination Invariant Face Verification," *Journal of Telecommunication and Information Technology*, 2010, vol. 4, pp. 19-25.

[31]. Yuehui Chen and Yaou Zhao, "Face Recognition using DCT and Hierarchical RBF Model," *IDEAL 2006, LNCS 4224*, 2006, pp. 355-362.

[32]. Sathesh and Samuel Manoharan, "A Dual Tree Complex Wavelet Transform Construction and its Application to Image Denoising," International Journal of Image Processing, vol. 3, issue 6, pp. 293- 300.

[33]. Ivan W Selesnick, Richard G Baraniuk and Nick G. Kingsbury, "The Dual-Tree Complex Wavelet Transform," *IEEE Signal Processing Magazine*, 2005, pp. 123-151.

[34]. Yigang Peng, Xudong Xie, Wenli Xu and Qionghai Dai, "Face Recognition using Anisotropic Dual-Tree Complex Wavelet Packets," *Proceedings of the Nineteenth International Conference on Pattern Recognition*, 2008, pp. 1-4.

[35]. K Jaya Priya and R S Rajesh, "Dual Tree Complex Wavelet Transform based Face Recognition with Single View," *Proceedings of the International Conference on Computing, Communications and Information Technology Applications*, 2010, vol. 5, pp. 455-459.

[36]. Yue-Hui Sun and Ming-Hui Du, "Face Detection using DT-CWT on Spectral Histogram," *Proceedings of the IEEE International Conference on Machine Learning and Cybernetics*, 2006, pp. 3637 – 3642.

[37]. K Jaya Priya and R S Rajesh, "Local Fusion of Complex Dual-Tree Wavelet Coefficients based Face Recognition for Single Sample Problem," *International Journal by Elsevier Ltd.*, 2010,vol. 2, pp. 94-100.

Ramesha K awarded the B.E degree in E & C from Gulbarga University and M.Tech degree in Electronics from Visvesvaraya Technological University. He is pursuing his PhD. in E & C Engineering at JNTU Hyderabad, under the guidance of Dr. K. B. Raja, Assistant Professor, Department of Electronics and Communication Engineering, University Visvesvaraya College of Engineering. He has over 8 research publications in refereed International Journals and Conference Proceedings. He is currently an Assistant Professor, Dept. of Telecommunication Engineering, Vemana Institute of Technology, Bangalore. His research interests include Image processing, Computer Vision, Pattern Recognition, Biometrics, and Communication Engineering. He is a life member of Indian Society for Technical Education, New Delhi.

K B Raja is an Assistant Professor, Dept. of Electronics and Communication Engineering, University Visvesvaraya college of Engineering, Bangalore University, Bangalore. He obtained his BE and ME in Electronics and Communication Engineering from University Visvesvaraya College of Engineering, Bangalore. He was awarded Ph.D. in Computer Science and Engineering from Bangalore University. He has over 75 research publications in refereed International Journals and Conference Proceedings. His research interests include Image Processing, Biometrics, VLSI Signal Processing, and Computer Networks.

Color and Texture Feature for Remote Sensing – Image Retrieval System: A Comparative Study

Retno Kusumaningrum[1] and Aniati Murni Arymurthy[2]

**[1] Department of Informatics, Diponegoro University
Semarang, 50275, Indonesia**

**[2] Faculty of Computer Science, University of Indonesia
Depok, 16424, Indonesia**

Abstract

In this study, we proposed score fusion technique to improve the performance of remote sensing image retrieval system (RS-IRS) using combination of several features. The representation of each feature is selected based on their performance when used as single feature in RS-IRS. Those features are color moment using L*a*b* color space, edge direction histogram extracted from Saturation channel, GLCM and Gabor Wavelet represented using standard deviation, and local binary pattern using 8-neighborhood. The score fusion is performed by computing the value of image similarity between an image in the database and query, where the image similarity value is sum of all features similarity, where each of feature similarity has been divided by SVD value of feature similarity between all images in the database and query from related feature. The feature similarity is measured by histogram intersection for local binary pattern, whereas the color moment, edge direction histogram, GLCM, and Gabor are measured by *Euclidean Distance*. The final result shows that the best performance of remote sensing image retrieval in this study is a system which uses the combination of color and texture features (i.e. color moment, edge direction histogram, GLCM, Gabor wavelet, and local binary pattern) and uses score fusion in measuring the image similarity between query and images in the database. This system outperforms the other five individual feature with average precision rates 3%, 20%, 13%, 11%, and 9%, respectively, for color moment, edge direction histogram, GLCM, Gabor wavelet, and LBP. Moreover, this system also increase 17% compared to system without score fusion, simple-sum technique.

Keywords: *Color Moment, Edge Direction Histogram, GLCM, Gabor Wavelet, Local Binary Pattern, Score Fusion.*

1. Introduction

Remote sensing image is a representation of an up-to-date part of the earth surface as seen from the space. The information of remote sensing images is close to the reality of earth surfaces. Hence, remote sensing images are widely used in various fields at the present time, such as agriculture, mineral exploration, military, forestry, fisheries, etc. The high advantage of using remote sensing images as reference not only makes increasing of sensor system technology, but also increases size or volume of remote sensing image. Therefore, it needs to develop a remote sensing – image retrieval system (RS-IRS) that has good performance and easy to use.

Recently, most of proposed RS-IRS automatically extract low-level features (e.g. color, texture, and shape) to measure similarity among images by comparing the feature similarity. Maheswary and Srivastava use combination of color and texture feature [1]. Color feature is represented by color moment of HSV color space, while texture feature is represented by Gray Level Co-occurrence Matrix. On the other hand, Long, et.al. said that the color moment perform better if it is defined by both CIE color spaces, L*a*b* and L*u*v*, as opposed to solely by the HSV color space [2]. Another system proposed by Ruan, et.al. uses combination of color feature and GLCM as texture feature [3]. The color feature vector is defined as combination of ratio between mean value of red and mean value of green and ratio between mean value of blue and mean value of green. Peijun, et.al. develop a prototype of content based remote sensing image retrieval which can be used to retrieve hyperspectral remote sensing images by spectral feature (Normalized Difference of Vegetation Index – NDVI, Normalized Difference of Building Index – NDBI, and Normalized Difference of Water Index – NDWI), color feature (RGB space represented by histogram), texture feature (GLCM, fractal or wavelet), or spatial features (spatial location – coordinate) [4].

However, their approach suffer from a number of weakness. First, they use various combination of features according to the interest of each authors. For example, most of systems use color feature and GLCM as texture representation, whereas there are many other texture representations that probably more suitable for remote sensing images domain. Second, there are some retrieval methods only focus on single low-level feature. Third, some systems define image similarity as simple sum of similarity features. They do not pay attention to the scale of each feature similarity values.

Therefore, this paper will discuss comparison of some representation of color features and texture features for RS-IRS using high-resolution images as domain. In addition, we also proposed a feature fusion technique to combine several features used. The feature representations that will be compared in this study are limited to color features and texture features. Selection of these two features is due to the several reasons. Firstly, the color feature is invariant to the rotation and scale. Secondly, the texture analysis offers the interesting possibilities to characterize the structural heterogeneity of each class. Thirdly, the image retrieval use global low-level model instead of objects-based semantic model. Hence, the shape feature become less appropriate, since this feature usually described after image have been segmented into regions or objects [2].

The rest of this paper follows the following structure. Section 2 describes all related feature analysis methods. Section 3 presents the mechanism of image retrieval system includes feature fusion. Section 4 discusses experimental result and conclusion of the study will be explained in section 5.

2. FEATURES ANALYSIS METHODS

In this section, we will briefly explain about the features used in this study and the techniques for extracting those features.

2.1 Color

Color is the most widely used visual content for image retrieval system. Two points should be considered when using color feature are color space and color description.

There are some color space used to represent an image. First, the most extensively used color space is RGB color space. This color space is called as "additives primaries" since a color in RGB space is produced by adding three color components, i.e. red channel, green channel, and blue channel.

Second, the color space derived from RGB color space is the HSV color space. This color space more intuitive to describe color, invariant to the changes of illumination and the direction of capture, and easy to transform from RGB to HSV and vice versa [2]. The first step of color space conversion from RGB to HSV is find maximum value of RGB triplet, *M*, and minimum value from the RGB triplet, *m*. Saturation, S, is then calculated using this following equation.

$$S = \frac{(M - m)}{M} \tag{1}$$

and Value, V, is equal to *M*. To calculate the Hue, H, we need to normalize the RGB triplet into ranges 0 to 1 as follow.

$$r = \frac{M - R}{M - m} \tag{2}$$

$$g = \frac{M - G}{M - m} \tag{3}$$

$$b = \frac{M - B}{M - m} \tag{4}$$

and Hue, H, is

$$H = \begin{cases} 60(b - g) \;\; if\; R = M \\ 60\,(2 + r - b) \;\; if\; G = M \\ 60(4 + g - r) \;\; if\; B = M \\ H - 360 \;\; if\; H \geq 360 \\ H + 360 \;\; if\; H < 0 \end{cases} \tag{5}$$

Third, the color spaces defined by the Commission Internationale de L'Éclairage (CIE) are L*a*b* (CIELab) and L*u*v* (CIELuv). Both of these color spaces classify color according to the human visual system, so that CIELAB and CIELuv are device independent but suffer from being quite un-intuitive despite the L parameter having a good correlation with perceived lightness [5].

The foundation of CIELab and CIELuv is the XYZ color space, the CIE standard color system, which has a linear relationship with non-gamma corrected RGB. See the equation below :

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 3.240479 & -1.53715 & -0.498535 \\ -0.969256 & 1.875992 & 0.041556 \\ 0.055648 & -0.204043 & 1.057311 \end{bmatrix} \times \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{6}$$

$$x = \frac{X}{(X + Y + Z)} \tag{7}$$

$$y = \frac{Y}{(X + Y + Z)} \tag{8}$$

Subsequently the lightness, denoted by L*, is defined as:

$$L^* = \begin{cases} 116\left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} - 16, & if \ \frac{Y}{Y_n} > 0.008856 \\ \left(\frac{Y}{Y_n} - \frac{4}{29}\right)\frac{108}{841}, & if \ \frac{Y}{Y_n} \leq 0.008856 \end{cases} \tag{9}$$

Where X_n = 0.950456, Y_n = 1, and Z_n = 1.088754. The formula to compute u*, v* and a*, b* can be seen in the following equation.

$$u^* = 13(L^*)(u' - u_n') \tag{10}$$

$$v^* = 13(L^*)(v' - v_n') \tag{11}$$

Where :

$$u' = u = \frac{2x}{(6y - x + 1.5)} \tag{12}$$

$$v' = 1.5v = \frac{4.5y}{(6y - x + 1.5)} \tag{13}$$

One color description that is able to provide efficiency and effectiveness in representing the distribution of colors of an image is color moment [6]. Color moment was introduced by Stricker and Orengo [6]. There are three color moments, i.e. *mean*, *standard deviation*, and *skewness*. These central moments are computed for each channel. Therefore, if an image has three channels, then the dimension of this feature is 9D (9 dimension). Mathematically, those moments for each channels are defined as follow [1].

- *MOMENT 1 – Mean:* The average color value in the image.

$$E_i = \sum_{N}^{j=1} \frac{1}{N} p_{ij} \tag{14}$$

- *MOMENT 2 – Standard Deviation:* The square root of the variance of distribution.

$$\sigma_i = \sqrt{\left(\frac{1}{N}\sum_{N}^{j=1}(p_{ij} - E_i)^2\right)} \tag{15}$$

- *MOMENT 3 – Skewness:* A measure of the degree of asymmetry in the distribution

$$s_i = \sqrt[3]{\left(\frac{1}{N}\sum_{N}^{j=1}(p_{ij} - E_i)^3\right)} \tag{16}$$

Where the p_{ij} is the pixel value of the i^{th} color channel at j^{th} image pixel, N is the number of pixels in the image.

2.2 Textures

Texture is a property to represent the surface and structure of an image which can be defined as a regular repetition of an element or pattern on a surface [7]. Basically, texture representation divide into two major categories, i.e. structural and statistical approach [2]. In this paper, we only discuss some of statistical approaches for texture analysis. In the statistical approach, the texture features are computed from the statistical distibution of observed combination of intensities at specified positions relative to each other position in the image. Based on the number of pixels defining the local feature, statistical approach can be futher classified into first-order (one pixel), second-order (two pixels), and higher-order (three or more pixels) statistics [8].

There are four texture features that will be used in this study, including gray level co-occurrence matrix, edge direction histogram, gabor wavelet, and local binary pattern. The following sub-sections describe the four texture features in detail.

2.2.1 Gray Level Co-occurrence Matrix (GLCM)

GLCM is the two dimensional matrix of joint probabilities between pair of pixels (one with gray level *i* and the other with gray level *j*), separated by a distance *d* in a given direction θ [3]. Hence, GLCM is included in the second-order statistical texture analysis.

The extraction process of GLCM features are divided into two main processes, i.e. the formation of co-occurrence matrix and the extraction of GLCM descriptors against the co-occurrence matrix. The following figure illustrates the formation of co-occurrence matrix.

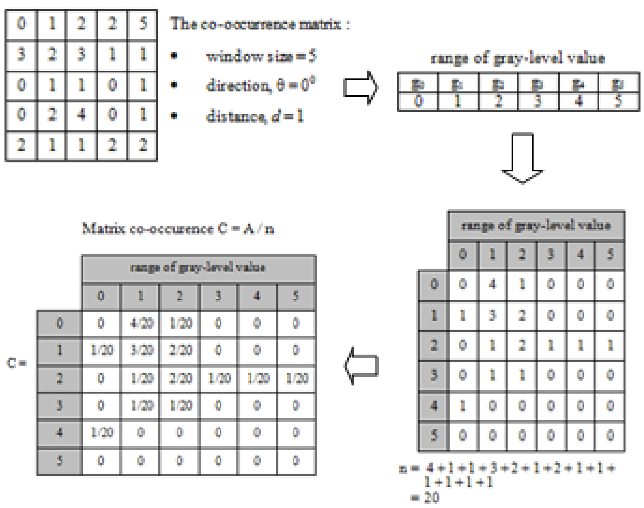


Fig. 3 Illustration of formation of matrix co-occurrence

Based on the co-occurrence matrix, the next step is computing the GLCM descriptors as follows [9]:

- *Angular Second Moment (ASM) / Energy:* shows the texture uniformity or texture homogenity. Energy value will be greater for a homogeneous texture.

$$Energy = \sum_{i=1}^{k}\sum_{j=1}^{k}(C_{ij})^2 \tag{17}$$

- *Entropy:* shows the degree of randomness. The maximum value of entropy will be reached when all elements C_{ij} has the same value. Inhomogeneous scenes have low entropy, while a homogeneous scene has a high entropy.

$$Entropy = -\sum_{i=1}^{k}\sum_{j=1}^{k} C_{ij}\log(C_{ij}) \tag{18}$$

- *Contrast / Second Order Element Difference Moment:* shows the contrast texture value. The calculation results in a larger figure when there is great contrast.

$$Contrast = \sum_{i=1}^{k}\sum_{j=1}^{k} C_{ij}(i-j)^2 \tag{19}$$

- *Cluster Shade:* shows the lack of symmetry in an image.

$$ClusterShade = \sum_{i=1}^{k}\sum_{j=1}^{k}(i-j-2m)^2 C_{ij} \tag{20}$$

- *Correlation:* A measure of gray level linear dependence between the pixels at the specified positions relative to each other.

$$Correlation = \sum_{i=1}^{k}\sum_{j=1}^{k}(i-m)(j-m)C_{ij} \tag{21}$$

- *Homogeneity:* shows the first order inverse element difference moment.

$$Homogeneity = \sum_{i=1}^{k}\sum_{j=1}^{k}\frac{C_{ij}}{n} \tag{22}$$

- *Maximum Probability*: shows the emergence of the gray-level value g_i adjacent to the gray-level value g_j more dominant in the image.

$$MaximumProbability = \max(C_{ij}) \tag{23}$$

- *Inverse Difference Moment (IDM)*: a low IDM value for inhomogeneous images, and a relatively higher value for homogeneous images

$$Inverse = \sum_{i=1}^{k}\sum_{j=1}^{k}\frac{C_{ij}}{n^2} \tag{24}$$

where:

C_{ij} is element of matrix co-occurence
$n = i - j$ if $i \neq j$ and $n = 1$ if $i = j$
m is mean value of matrix co-occurence

2.2.2 Edge Direction Histogram (EDH)

Initially, we will be performed the Gaussian smoothing against the image channel. After that, perform the edge point's detection using Canny filter. We calculate the gradient of each edge points by utilizing 5-type operators Sobel, i.e. horizontal edge, vertical edge, 45-degree edge, 135-degree edge, and non directional edge. The following figure define those 5 operators Sobel.

$$\begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

(a). Horizontal edge

$$\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

(b). Vertical Edge

$$\begin{bmatrix} -2 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix}$$

(c). 45-degree edge

$$\begin{bmatrix} 0 & 1 & 2 \\ -1 & 0 & 1 \\ -2 & -1 & 0 \end{bmatrix}$$

(d). 135-degree edge

$$\begin{bmatrix} -1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & -1 \end{bmatrix}$$

(e). Non directional edge

Fig. 4 Sobel Operators

Finally, the 5-dimensional edge histogram is calculated by counting the edge pixel in each direction.

2.2.3 Gabor Filter (Wavelet) [10]

For a given image *I(x,y)* with size *P* x *Q* and *s* and *t* are the filter mask size variables, the discrete Gabor wavelet transform is given by a convolution :

$$G_{mn}(x,y) = \sum_{s}\sum_{t} I(x-s, y-t)\, \psi^{*}_{mn}(s,t) \qquad (25)$$

where ψ^{*}_{mn} is the complex conjugate of ψ_{mn} which is a class of self-similar functions generated from dilation and rotation of the following mother wavelet :

$$\psi(x,y) =$$

$$\frac{1}{2\pi\sigma_x\sigma_y} exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)\right].\exp\,(j2\pi Wx) \qquad (26)$$

where *W* is called the modulation frequency.

The self-similar Gabor wavelet are obtained through the generating function :

$$\psi_{mn}(x,y) = a^{-m}\psi(\tilde{x},\tilde{y}) \qquad (27)$$

where *m* (*m = 0, 1, ... M-1*) and *n* (*n = 0, 1, ..., N-1)* are the scale and orientation of the wavelet respectively.

$$\tilde{x} = a^{-m}(x\cos\theta + y\sin\theta) \qquad (28)$$

$$\tilde{y} = a^{-m}(-x\cos\theta + y\cos\theta) \qquad (29)$$

where $a > 0$ and $\theta = {n\pi}/{N}$

After applying the Gabor filter on the image with different orientation at different scale, we obtained the energy content at different scale and orientation of the image.

$$E(m,n) = \sum_{x}\sum_{y} |G_{mn}(x,y)| \qquad (30)$$

2.2.4 Local Binary Patter (LBP)

The name of "Local Binary Pattern" reflects the functionality of the operator, i.e. a local neighborhood is thresholded at the gray value of centre pixels into a binary pattern [11]. Based on the labels, in the form of binary pattern, we can create histogram of labels as a texture descriptor. See the following figure for an illustration of the basic LBP.

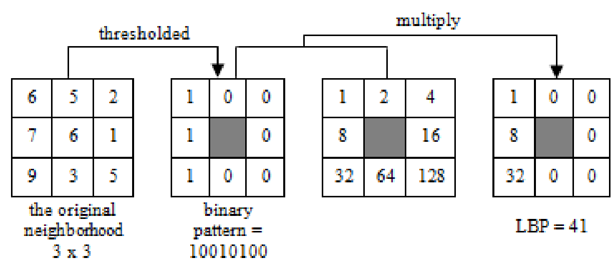


Fig. 5 Illustration of basic LBP8,1.

Matematically, it can be done as follow :

$$LBP_{P,R}(x_c,y_c) = \sum_{p=0}^{P-1} s\left(g_p - g_c\right).2^p \qquad (31)$$

where *s(x)* is thresholding function

$$s(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases} \qquad (32)$$

The variable in the Eq. (31) are defined as follows.

- *P* : number of neighborhood
- *R* : radius
- g_p : gray level value at neighborhood p^{th}
- g_c : gray level value at centre pixel

In practice, Eq. 18 means that the sign of the differences in a neighborhood are interpreted as a *P*-bit binary number, resulting in 2^P distinct value for LBP code and the local gray-scale distribution can thus be approximately described with 2^P-bin discrete distribution of LBP code [12].

3. The Image Retrieval System

We use the global low-level model for the image retrieval system. See the Appendix for the detail diagram of image retrieval system in this study.

Basically, the process of image retrieval system in this study is divided into four main steps, i.e. features extraction, similarity measurement, process of indexing (ranking), and display the results. In the following sub-section, we will explain those steps in detail.

3.1 Features Extraction

This step is divided into two processes, namely pre-processing and feature extraction process itself. The employed features in this study are color moment, EDH, GLCM, Gabor Wavelet, and LBP. For each of features, pre-processing using different approaches. Preprocessing for feature GLCM, Gabor and LBP is the conversion of images into gray-level image. While the color moment features, pre-processing is done by converting the image into 3 different color spaces, i.e. HSV, L*a*b* and L*u*v*. The use of three color spaces for color features is performed to get the most appropriate color space for color moment as a color descriptor, especially in the domain of high-resolution remote sensing images. And pre-processing for EDH is the conversion of images into the HSV color space and gray-level image.

As mentioned before, the result of GLCM extraction is matrix of each descriptor (e.g. energy, entropy, contrast, cluster shade, etc.), while Gabor wavelet is the matrix of Energy. Therefore, we implement two basic statistical analysis, mean and standard deviation, for those descriptor. The parameters used in the process of GLCM and Gabor features extraction are as follows:

- *GLCM:* The GLCM feature extraction uses moving kernel (window) with size 5x5 as recomended in [9]. And the angles used in this process are 0 (horizontal or east-west), 90 (vertical or south-north), 45 (diagonal or southwest-northeast), and 135 (diagonal or southeast-northwest), thus the counting process carried out for all possible of the objects direction.
- *Gabor Wavelet:* The representation of texture feature using Gabor wavelets will provide the best performance by using the scale value is 4 and orientation value is 6. It is recommended in [13].

3.2 Features Extraction

The following tables explain about the formal definition of variables and functions that is used in this study.

Table 1: Variable Formalization in Image Retrieval

Set	Symbol	Element	Size	Description
Image	I	i_j	N_I	Images in the database
Feature	F	f_n	N_F	Features that are extracted from an image
Query	Q	q	-	Feature vector of query

Table 2: Function Formalization in Image Retrieval

Name	Symbol	Mapping
Feature similarity	S_F	$I \times Q \times F \rightarrow [0, \infty]$
Image similarity	S_I	$I \times Q \rightarrow [0, \infty]$

The process of measure similarity between the query image and the images in the database is performed using image similarity function, S_I. The image similarity function is the sum of score fusion for each feature. Mathematically, it can be defined as follow :

$$S_I(i_j, q) = \sum_{f_n \in F} \left[\frac{S_F(i_j, q; f_n)}{svd(S_F(I, q; f_n))} \right] \qquad (33)$$

where :

$S_F(i_j, q; f_n)$: feature similarity between the image i_j and query q with the respect to feature f_n.

$S_F(I, q; f_n)$: all of the feature similarity value between the image i_j and query q with the respect to feature f_n. If we have N_I images in the database, then this value is column matrix of feature similarity with the size $N_I \times 1$.

$svd(S_F(I, q; f_n))$: a function to get singular value decomposition of $S_F(I, q; f_n)$

For example, if we have feature vector of an image i_j with the respect to feature f_n is $X_p = (x_1, x_2, \dots, x_D)$ and the feature vector of query q with the respect to feature f_n is $Y_p = (y_1, y_2, \dots, y_D)$, then the similarity feature between image i_j and query q with the respect to feature f_n is defined as follow.

$$S_F(i_j, q; f_n) = \begin{cases} 1 - \dfrac{\sum_{p=1}^{D} min(X_p, Y_p)}{min\left(\sum_{p=1}^{D} X_p, \sum_{p=1}^{D} Y_p\right)}, & if\ f_n\ is\ LBP \\ \left(\sum_{p=1}^{D} (X_p - Y_p)^2\right)^{1/2}, & otherwise \end{cases} \qquad (34)$$

In the other word, the feature similarity for local binary pattern is measured by using histogram intersection [14], while color moment, EDH, GLCM, and Gabor are measured by *Euclidean Distance*. We propose the use of SVD (Singular Value Decompostion) function in the score fusion because it is more robust and efficient. The *svd* function in Eq. (33) returns the singular value of all feature similarity value with the respect to feature f_n. This singular value is equal to the operator norm of feature similarity so that we can obtain the same scale for all of the feature similarity values.

3.3 Indexing (Ranking) and Display Result

The indexing process is performed by sorting the images in the database based on image similarity values in ascending. This means the image that has the zero value for image similarity is the most similar image to the query. And the system will display the results, i.e. top-n images with the lowest value of image similarity.

4. EXPERIMENTS AND RESULT

4.1 Dataset

This study uses remote sensing image database. This database contains 200 IKONOS images of resolution 256 x 256 pixel. These images are RGB images.

4.2 Experiment Environment

This study is implemented using Matlab R2010a with the operating system is Windows 7 – 64bit and the hardware specification is as follows:

- Intel Core i5-520M 2.40 GHz
- 4GB of memory (RAM)
- 500 GB of hard disk drive

4.3 Performance Measurement

The performance result of the RS-IRS is presented in the form of Precision-Recall graph (PR-graph) averaged over several queries. We use 30 queries by example in the form of remote sensing images, so that the value of precision in the PR-graph is average precision value.

4.4 Experimental Scenarios

1). Experimental Setup 1: The objective of this scenario is identify the appropriate representation for each features, 1 color descriptor and 4 texture descriptor. The meaning of appropriate representation is a representation which has the best average precision value. Therefore, we have five different comparisons. First, we will compare three types of color space (i.e. HSV, L*a*b, and L*u*v*) for color moment descriptor. Second, we will describe the comparison between gray-level image and each channel of HSV color space for EDH extraction. Third and Forth, we will compare the use of two basic statistical analysis, mean and standar deviation, for texture representation of GLCM and Gabor respectively. The final one is comparison of the appropriate number of neighborhood, 4-neighborhood and 8-neighborhood, for LBP extraction.

2). Experimental Setup 2: The aim of this scenario is compare the simple sum technique and the proposed score fusion technique. This comparison is carried out against two type features combination, i.e. combination of texture features (EDH, GLCM, Gabor, and LBP) and combination of color – texture features (Color Moment, EDH, GLCM, Gabor, and LBP).

4.5 Experiment Results

1). Result of Experimental Setup 1: As mentioned in the previous sub-section, we will compare HSV, L*a*b*, and L*u*v* color space for color moment descriptor. See the following figure.

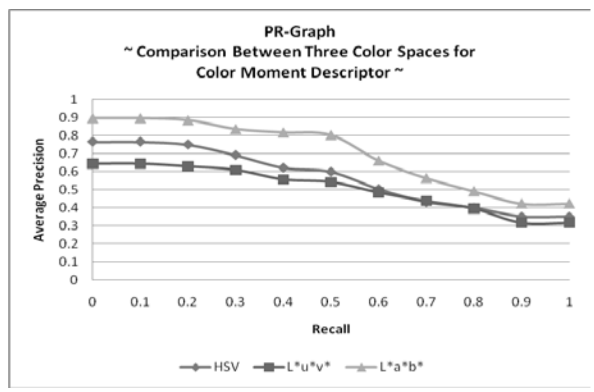


Fig. 6 PR-Graph, Comparison between Three Color Spaces for Color Moment Descriptor.

Based on the figure above, we can conclude that the best color space for color moment in this study is L*a*b*. It can be seen from the average precision of the color moment using the L*a*b* color space (green line) is higher 13% than the HSV color space (blue line) and higher 19% than the L*u*v* color space (red line). L*a*b* color space gives better performance since this space defines colors more closely to the human color perception. In addition, this space also uses three color coordinates includes L* – the lightness coordinate, a* – the red/green coordinate, and b* – the yellow/blue coordinate, thus it can be defined color as combinations of red and yellow, red and blue, green and yellow, and green and blue. Another interesting characteristic of the L*a*b* color space is that the distance can be calculated between two colors and proportional to the difference between the two colors as perceived by the human eye.

Furthermore, see the following figure to get the best pre-processing representation for process of EDH extraction.

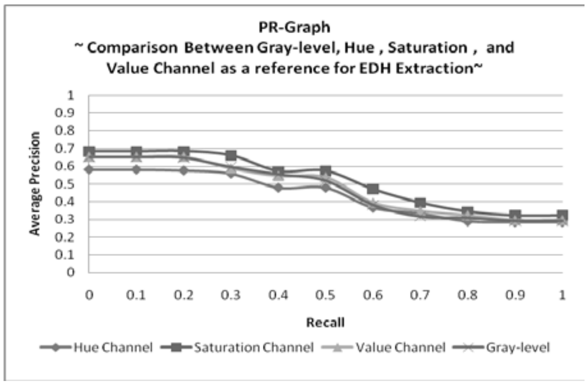
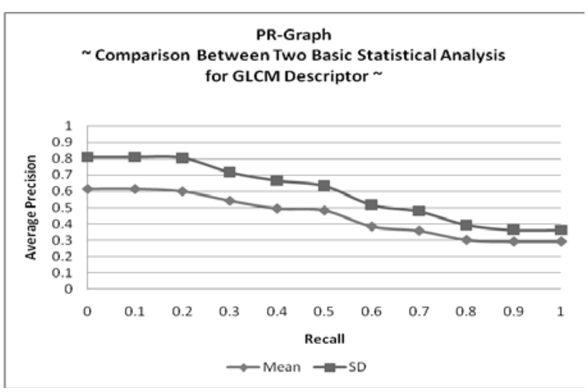
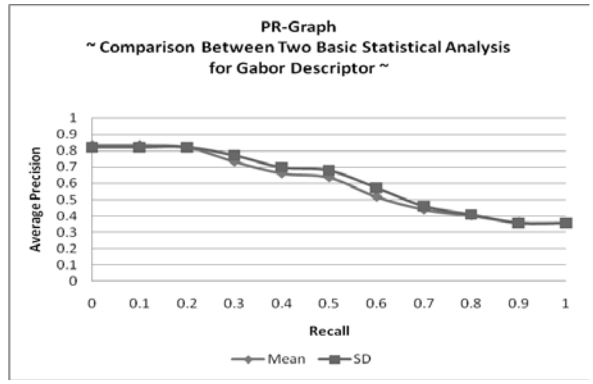


Fig. 7 PR-Graph, Comparison between Four Pre-processing Representation as a reference of EDH extraction.

The figure above shows that for each pre-processing representation, Saturation channel outperforms the other three representations with average precision rates 3%, 10%, and 3%, respectively, for gray-level, hue channel, and value channel representation. The saturation channel in HSV color space indicates the range of grey in the color space, thus it is appropriate to calculate the edge direction histogram. This color space has ranges from 0 to 100% or from 0 to 1. When the value is '0,' the color is grey and when the value is '1,' the color is a primary color.



(a). GLCM Descriptor



(b). Gabor Descriptor

Fig. 8 PR-Graph, Comparison between Two Basic Statistical Analysis for (a) GLCM Descriptor and (b) Gabor Descriptor.

Both figure above explain that standard deviation is the best representation for GLCM and Gabor. The performance of standard deviation exceeds the mean performance by 14% and 2% for GLCM and Gabor respectively. The standard deviation compared to the mean gives better performance since the standard deviation describe how spread out a set of values are around the mean of that set. A set of values that are closely clustered near the mean will have a low standard deviation, a set of numbers that are widely apart will have a higher standard deviation and a set of numbers that are all the same will have a standard deviation of zero.

The last comparison is compare 4-neighborhood and 8-neighborhood for LBP. The LBP with 8-neighborhood presents the better performace since this approach gives the complete information about the relationship between the centre pixel and all pixels surrounding. In the other word, the 8-neighborhood comparison carried out for all possible of the objects direction. See the following figure to see the comparison between LBP with 4-neigborhood and 8 neighborhood.

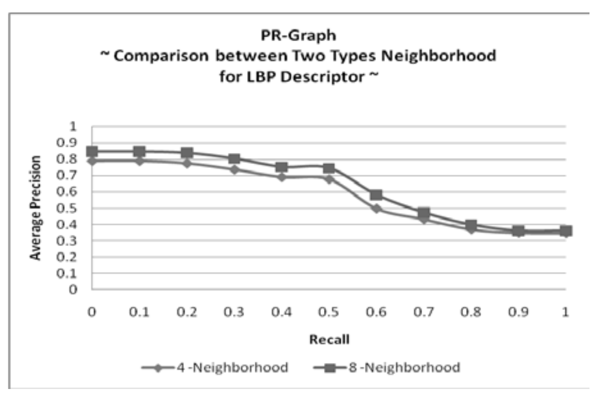


Fig. 9 PR-Graph, Comparison between LBP with 4-neighborhood and LBP with 8-neighborhood.

In summary, the best representation for each color and texture descriptor in this study is color moment descriptor with CIE L*a*b* color space, EDH which is extracted over the Saturation channel, GLCM and Gabor descriptor which is represented using standar deviation, and the LBP with 8-neighborhood.

2). *Result of Experimental Setup 2:* In this scenario, we will compare proposed score fusion technique and simple sum technique (without score fusion). Fig. 10 illustrates the comparison for combination of texture features, whereas Fig. 11 illustrates the comparison for combination of color – texture feature.

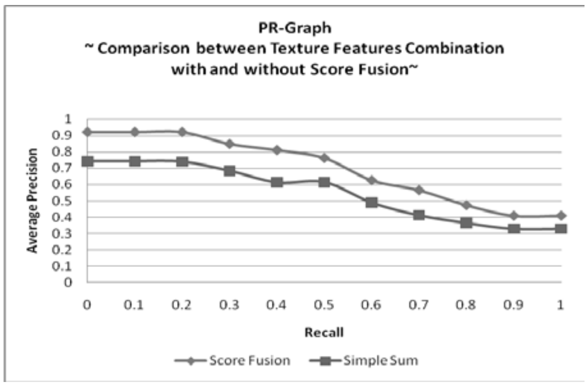


Fig. 10 PR-Graph, Comparison between Texture Feature Combination – with and without score fusion.

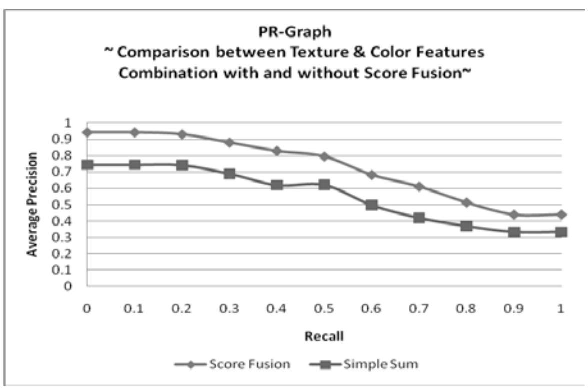


Fig. 11 PR-Graph, Comparison between Combination of Color and Texture Feature – with and without score fusion.

According to Fig. 10 and Fig. 11, score fusion always give better results than simple-sum technique, i.e. an increase of 15% for the combination of textures and 17% for the combination of colors - textures. The condition is due to the use of simple-sum technique is only adding up the value of all features similarity regardless the scale of each feature similarity. Hence, if there are types of features which have bigger feature similarity values than the other features, then these feature become dominant and the addition of other features will have no impact. In contrary, the use of score fusion technique will normalize the values of feature similarity, thus each features will contributes or influences on the image similarity value.

Moreover, Fig. 11 also shows that the addition of color feature gives better performance than simply using the combination of texture features. It is because the texture features can provide good performance on a textured area (heterogeneous), but tend to give unsatisfactory performance in a homogeneous area. On the other hand, color feature is able to distinguish objects in a homogeneous area. Therefore, a combination of both features could complement their respective advantages. In addition, the system which use combination of color and texture feature also exceeds the other five individual feature with average precision rates 3%, 20%, 13%, 11%, and 9%, respectively, for color moment, edge direction histogram, GLCM, Gabor wavelet, and LBP. In conclusion, the best performance of remote sensing image retrieval in this study is a system which uses the combination of color and texture features (i.e. color moment, edge direction histogram, GLCM, Gabor wavelet, and LBP) and uses score fusion in measuring the image similarity between query and images in the database.

4. Conclusions

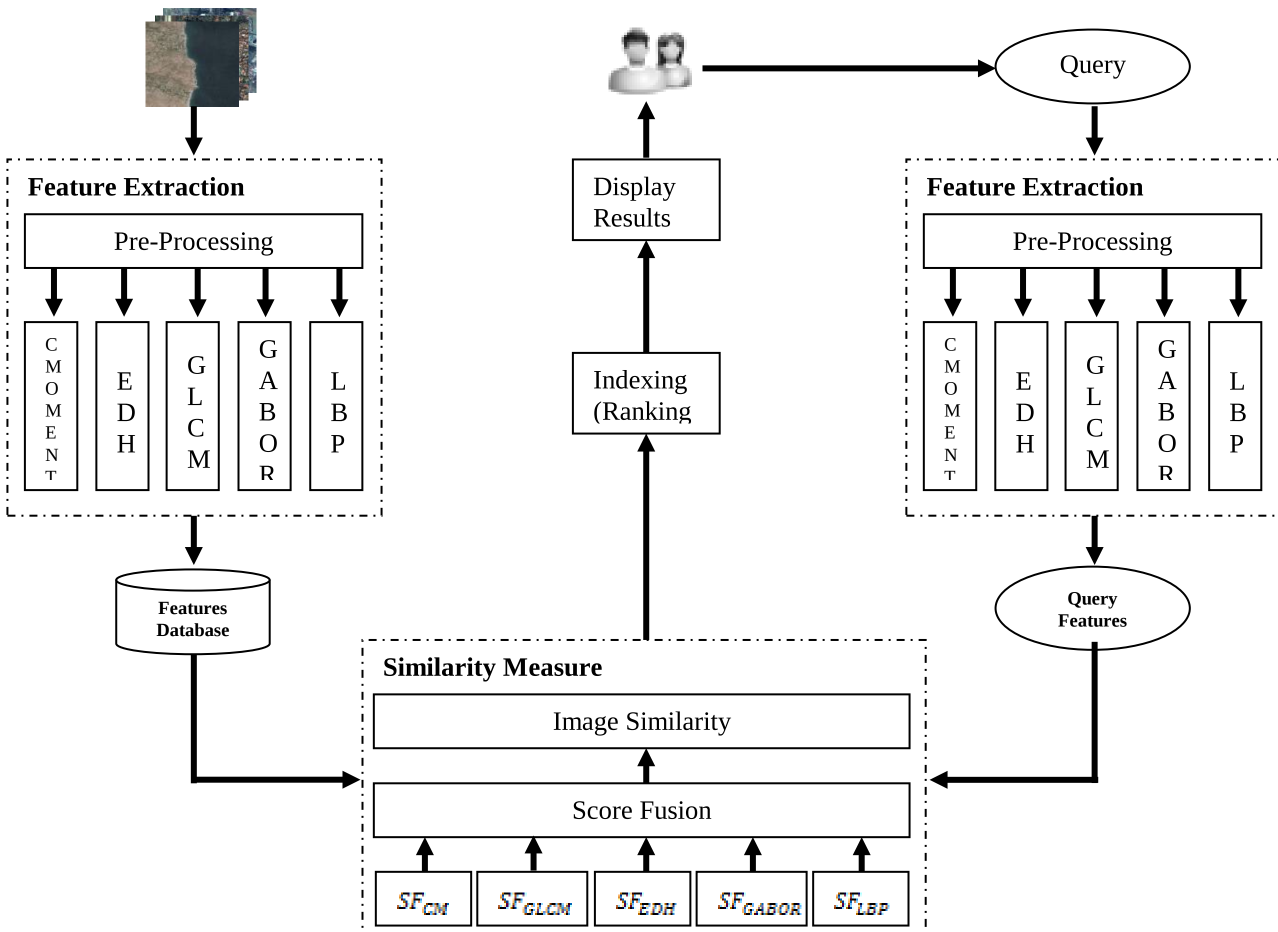
Nowadays, most of proposed RS-IRS use various combination of low-level feature according to the interest of each authors and some of them focus on single low-level feature.

In order to improve the performance of RS-IRS using combination of color and texture features we perform two steps. Firstly, we select the appropriate representation which give the best performance when used as single feature in RS-IRS. Secondly, we proposed score fusion technique to combine several features in the RS-IRS using multiple features.

Those selected features representation are color moment using L*a*b* color space, EDH extracted from Saturation channel, GLCM and Gabor wavelet represented using standard deviation, and LBP using 8-neighborhood. The score fusion is performed by computing the value of image similarity between an image in the database and query, where the image similarity value is sum of all features similarity, where each of feature similarity has been divided by SVD value of feature similarity between all images in the database and query from related feature. The feature similarity is measured by histogram intersection for LBP, whereas the color moment, EDH, GLCM, and Gabor are measured by *Euclidean Distance*.

The final result shows that the best performance of RS-IRS in this study is a system which uses the combination of color and texture features (i.e. color moment, EDH, GLCM, Gabor wavelet, and LBP) and uses score fusion in measuring the image similarity between query and images in the database. This system outperforms the other five individual feature with average precision rates 3%, 20%, 13%, 11%, and 9%, respectively, for color moment, edge direction histogram, GLCM, Gabor wavelet, and LBP. Moreover, this system also increase 17% compared to system without score fusion, simple-sum technique.

Appendix



Acknowledgments

The author would like to thank Indonesian National Institute of Aeronautics and Space - Lembaga Penerbangan dan Antariksa Nasional (LAPAN) which has provided some IKONOS images.

References

[1] P. Maheshwary, and N. Sricastava, “Prototype Sytem for Retrieval of Remote Sensing Images based on Color Moment and Gray Level Co-Occurrence Matrix”, IJCSI International Journal of Computer Science Issues, 2009, vol. 3, pp. 20-23

[2] F. Long, H. Zhang, and D.D. Feng, “Fundamentals of Content-Based Image Retrieval”, in Multimedia Information Retrieval, 2002

[3] N. Ruan, N. Huang, and. W. Hong, “Semantic-Based Image Retrieval in Remote Sensing Archive : An Ontology Approach”, in International Conference on Geoscience and Remote Sensing Symposium, 2006, pp. 2903 – 2906

[4] D. Peijun, C. Yunhao, T. Hong, and F. Tao, “Study on Content-Based Remote Sensing Image Retrieval”, in International Conference on Geoscience and Remote Sensing Symposium, 2005, 4 pp.

[5] A. Ford and A. Roberts, “Color Space Conversions”, 1998 [available online : http://debut.cis.nctu.edu.tw/~ching/Face/Articles/coloureq.pdf]

[6] M. Stricker and M. Orengo, "Similarity of Color Images," in IS&T and SPIE Storage and Retrieval of Image and Video Databases III, 1995, pp. 381-392

[7] G.N. Srinivasan, and G. Shobha, "Statistical Texture Analysis", in World Academy of Science, Engineering, and Technology, 2008, vol. 36, pp. 1264-1269

[8] T. Ojala and M. Pietikainen, "Texture Classification, Machine Vision, and Media Processing Unit", University of Oulu, Finland

[9] D. Chahyati, "Classification of Radar Images Based on Texture Features of Gray Level Co-occurrence Matrix Semivariogram and Wavelet Stationary", Master Thesis, Department of Computer Science, University of Indonesia , Depok, Indonesia, 2003

[10] D. Zhang, A. Wong, M. Indrawan, and G. Lu, "Content-based Image Retrieval Using Gabor Texture Features", IEEE Transaction on Pattern Analysis and Machine Intelligence – PAMI, 2000, pp. 13-15

[11] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns", IEEE Transaction on Pattern Analysis and Machine Intelligence – PAMI, vol. 24, no. 7, 2002, pp. 971-987

[12] T. Maenpaa and M. Pietikainen, "Texture Analysis with Local Binary Patterns in Handbook of Pattern Recognition and Computer Vision 3rd Edition", 2004, pp. 197-216

[13] B.S. Manjunath, and W.Y. Ma, "Texture Feature for Browsing and Retrieval of Image Data", IEEE Transaction on Pattern Analysis and Machine Intelligence – PAMI, 1996, pp. 837-842

[14] J. Smith and S.F. Chang, "Automated Image Retrieval Using Color and Texture Feature", IEEE Transaction on Pattern Analysis and Machine Intelligence – PAMI, 1995.

Retno Kusumaningrum achieved her undergraduate degree in Department of Mathematics from Diponegoro University, Semarang, Indonesia, where she is currently working toward as lecturer in Department of Informatics, in 2003. She earned her master degree in Faculty of Computer Science from University of Indonesia, Depok, Indonesia, in 2010. Currently, she is studying for her doctoral degree in Faculty of Computer Science, University of Indonesia, Jakarta, Indonesia. Her current research activities are in spatial pattern and image retrieval system, particularly feature extraction, relevance feedback, and objective evaluation, with main application in remote sensing domains.

Aniati Murni Arymurthy is a Professor in Faculty of Computer Science, University of Indonesia. She graduated from Department of Electrical Engineering, University of Indonesia, Jakarta, Indonesia. She earned her Master of Science in Department of Computer and Information Sciences, The Ohio State University (OSU), Columbus, Ohio, USA. She also holds Doktor from Department of Opto-Electronics and Laser Applications, University of Indonesia, Jakarta, Indonesia and a sandwich program at the Laboratory for Pattern Recognition and Image Processing (PRIP Lab), Department of Computer Science, Michigan State University (MSU), East Lansing, Michigan, USA. Her main research activities are image processing and pattern recognition.

On Tandem Communication Network Model with DBA and Modified Phase Type Transmission having NHP Arrivals for First Node and Poisson process arrivals for Second Node

M.V.Rama Sundari[1], K.Srinvas Rao[2] , P.Srinivasa Rao[3] and P.Suresh Varma[4]

[1] **Department of Computer Science and Systems Engineering, Andhra University**
Visakhapatnam-530 003, Andhra Pradesh, INDIA

[2] **Department of Statistics, Andhra University**
Visakhapatnam-530 003, Andhra Pradesh, INDIA

[3] **Department of Computer Science and Systems Engineering, Andhra University**
Visakhapatnam-530 003, Andhra Pradesh, INDIA

[4] **Department of Computer Science, Adikavi Nannaya University**
Rajahmundry-533105, Andhra Pradesh, INDIA

Abstract

Communication network models play a predominant role in performance evaluation of many communication systems. The packet arrival processes for data networks are not matching with the Poisson processes due to the nature of bursty and time dependent arrivals. In this paper, a three node communication network model with non homogeneous Poisson arrivals having dynamic bandwidth allocation under modified phase type transmission is introduced for performance evaluation and monitoring of several tele and satellite communications. The system performance measures of the network are derived explicitly. The sensitivity analysis reveals the dynamic bandwidth allocation strategy and non homogeneous Poisson arrivals can reduce the burstyness in buffers and delay in transmission. This model also includes some of the earlier models as particular cases for specific and limiting of parameters

Keywords: *Tandem Communication Network, Dynamic Bandwidth Allocation, non homogeneous Poisson process, Performance evaluation.*

1. Introduction

William Cock and Charles Wheatstone (1839) have pioneered the mathematical theory of communication. Thereafter in 1948 a tremendous revolution in communication network modelling is brought by Barden and Barttain. Later AT&T Bell labs, USA network group at MIT, IEEE communication society and other reputed organizations have put considerable efforts for invoking and designing efficient communication systems. (IEEE Communication Society, 2002). Conducting laboratory experiments under variable load conditions is highly complex and time consuming. Hence for efficient design and evaluation of communication networks, the network models are developed and analyzed with various assumptions on constituent processes of the model like arrival process, service process, flow control mechanisms, allocations, routing, etc,. To improve Quality of Service, packet switching gives better utilization over circuit or message switching and yields relatively shorter delay in statistical multiplexing in communication network can reduce the delay in packet switching. Many of communication networks which support tele processing applications are mixed with dynamic engineering skills and statistical multiplexing (Gaujal and Hyon (2002), Parthasarathy et al(2001), Srinivasa Rao et al (2000))

To reduce the congestion in buffers the dynamic bandwidth allocation is evolved as an alternative and efficient control strategy over bit dropping or flow control strategies (Sriram et al (1993), Suresh Varma et al (2007), Padmavathi et al (2009), Nageswara Rao et al (2010)). In DBA, a large portion of the unutilized bandwidth is utilized by changing the transmission rate of packets depending on the content of the buffer connected to it. Much work has been reported recently regarding communication network model with dynamic bandwidth (Nageswara Rao et al (2011)). In all these papers, the authors assumed that the arrival rate of packets is constant and follows a Poisson process.

But in many communication systems the assumption regarding Poisson process are seldom satisfied due to the time dependent nature of arrivals. Therefore, many phenomenon encountered in communication systems which reveal time dependent behaviour of arrival process due to the factors like work load fluctuations, initiating, failures, congestion and flow control, overload peaks, reconfigurations, adaptive routing and others. So to have an efficient performance evaluation entities problem of communication systems like adaptive isolated routing and load balancing, evaluating alternative buffer of changed sharing schemes and effects and to study the effects of flow and control strategies it is needed to develop communication network models with time dependent arrivals and time dependent analysis. In addition to this, Rakesh Singhai et al(2007) have showed that the packet arrival process of heavy tail distribution inter arrivals form a non homogeneous Poisson process and the mean packet arrival rate is not constant but it is time varying. This is also supported by the studies Feldmann (2000), who demonstrated the TCP connection arrival process is bursty and time dependent. The time dependent arrival process can be well characterized by NHP which follows Poisson process.

Very little work has been reported in literature regarding communication network models with direct arrivals to first two nodes having modified phase type transmission and dynamic bandwidth allocation with non homogeneous Poisson arrivals. Hence, in this paper a communication network model is developed and analyzed for these sort of situations. Here it is assumed that the messages arrive to the first and second buffers directly with time dependent arrival rate. After getting transmitted from first buffer the packets may join the second buffer connected to the second node in tandem with first node or get terminated with certain probability. In the second buffer the packets are from first node and directly from outside the network. After getting transmitted from the second node the packets may be routed to the third buffer connected to the third node or get transmitted with certain probability.

The direct arrival to the second node will have tremendous influence on congestion control of the communication system and reduce burstness in the first buffer. Here also it is assumed that transmission completion of each node follows Poisson process with different transmission rates. The dynamic bandwidth allocation strategy is adopted to utilize the ideal bandwidth at nodes and to improve transmission capabilities.

Using difference differential equations the joint probability generating function of the number of packets in each buffer is derived. The performance of the communication network is studied by deriving explicitly the performance measures of the network like the average number of packets in each buffer, the throughput of the node, the mean delay in buffers, the utilization of the transmitters, the variability of the buffer content etc,. The sensitivity of the model with respect to the parameters is also carried. A comparative study of the proposed model with that of Poisson arrivals is also presented. This model is much useful for evaluating several communication systems where the arrivals are time dependent.

2. Tandem Communication Network Model with DBA and Modified Phase Type Transmission having NHP Arrivals for First Node and Poisson process arrivals for Second Node:

In this section, a communication network model having three nodes in tandem is studied. The arrivals to the buffer connected at node one is assumed to follow a non-homogeneous Poisson process with mean arrival rate as a function of time t. It is of the form $\lambda(t) = \lambda + \alpha t$. The transmission process from node one to node two follows a Poisson process with parameter μ_1. It is also assumed that the packets arrive to the second buffer directly from out side of the network in a Poisson process with mean arrival rate ϵ. After getting transmitted from node one the packets are forwarded to the second buffer for transmission with probability θ or get terminated with probability $(1-\theta)$ i.e., the packets arrived at second buffer contains the packets received from first node and directly from outside. After getting transmitted from second node the packet are forwarded to the third buffer for transmission with probability π or get terminated with probability $(1-\pi)$. The transmission process of node two and three also follow Poisson process with parameters μ_2 and μ_3 respectively. The transmission rate of each packet is adjusted just before transmission depending on the content of the buffer connected to the transmitter. The packets are transmitted through the transmitters by the first in first out discipline. The schematic diagram representing the communication network model is shown in Figure 1

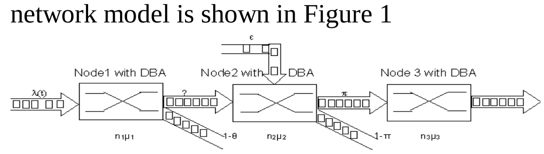


Figure 1: Schematic diagram of the Communication Network Model

Let $P_{n_1,n_2,n_3}(t)$ denote the probability that there are n_1 packets in the first buffer and n_2 packets in the second buffer and n_3 packets in the third buffer at time t

The difference-differential equations of the network are

$$\begin{aligned}\frac{\partial P_{n_1,n_2,n_3}(t)}{\partial t} = &-\left(\lambda(t)+\varepsilon+n_1\mu_1+n_2\mu_2+n_3\mu_3\right)P_{n_1,n_2,n_3}(t)+\lambda(t)P_{n_1-1,n_2,n_3}(t)+(n_1+1)\mu_1\theta P_{n_1+1,n_2-1,n_3}(t)\\&+(n_1+1)\mu_1(1-\theta)P_{n_1+1,n_2,n_3}(t)+\varepsilon P_{n_1,n_2-1,n_3}(t)+(n_2+1)\mu_2\,\pi\,P_{n_1,n_2+1,n_3-1}(t)\\&+(n_2+1)\mu_2(1-\pi)P_{n_1,n_2+1,n_3}(t)+(n_3+1)\mu_3P_{n_1,n_2,n_3+1}(t), \qquad n_1>0,n_2>0,n_3>0\end{aligned}$$

$$\frac{\partial P_{n_1,n_2,0}(t)}{\partial t} = -(\lambda(t)+\varepsilon+n_1\mu_1+n_2\mu_2)P_{n_1,n_2,0}(t)+\lambda(t)P_{n_1-1,n_2,0}(t)+(n_1+1)\mu_1\theta P_{n_1+1,n_2-1,0}(t)$$
$$+(n_1+1)\mu_1(1-\theta)P_{n_1+1,n_2,0}(t)+\varepsilon P_{n_1,n_2-1,0}(t)+(n_2+1)\mu_2(1-\pi)P_{n_1,n_2+1,0}(t)$$
$$+\mu_3 P_{n_1,n_2,1}(t), \qquad n_1>0, n_2>0$$

$$\frac{\partial P_{0,n_2,n_3}(t)}{\partial t} = -(\lambda(t)+\varepsilon+n_2\mu_2+n_3\mu_3)P_{0,n_2,n_3}(t)+\mu_1\theta P_{1,n_2-1,n_3}(t)+\mu_1(1-\theta)P_{1,n_2,n_3}(t)$$
$$+\varepsilon P_{0,n_2-1,n_3}(t)+(n_2+1)\mu_2\pi P_{0,n_2+1,n_3-1}(t)+(n_2+1)\mu_2(1-\pi)P_{0,n_2+1,n_3}(t)$$
$$+(n_3+1)\mu_3 P_{0,n_2,n_3+1}(t), \qquad n_2>0, n_3>0$$

$$\frac{\partial P_{n_1,0,n_3}(t)}{\partial t} = -(\lambda(t)+\varepsilon+n_1\mu_1+n_3\mu_3)P_{n_1,0,n_3}(t)+\lambda(t)P_{n_1-1,0,n_3}(t)+(n_1+1)\mu_1(1-\theta)P_{n_1+1,0,n_3}(t)$$
$$+\mu_2\pi P_{n_1,1,n_3-1}(t)+\mu_2(1-\pi)P_{n_1,1,n_3}(t)+(n_3+1)\mu_3 P_{n_1,0,n_3+1}(t), \qquad n_1>0, n_3>0$$

$$\frac{\partial P_{n_1,0,0}(t)}{\partial t} = -(\lambda(t)+\varepsilon+n_1\mu_1)P_{n_1,0,0}(t)+\lambda(t)P_{n_1-1,0,0}(t)+(n_1+1)\mu_1(1-\theta)P_{n_1+1,0,0}(t)$$
$$+\mu_2(1-\pi)P_{n_1,1,0}(t)+\mu_3 P_{n_1,0,1}(t), \qquad n_1>0$$

$$\frac{\partial P_{0,n_2,0}(t)}{\partial t} = -(\lambda(t)+\varepsilon+n_2\mu_2)P_{0,n_2,0}(t)+\mu_1(1-\theta)P_{1,n_2,0}(t)+\mu_1\theta P_{1,n_2-1,0}(t)$$
$$+\varepsilon P_{0,n_2-1,0}(t)+(n_2+1)\mu_2(1-\pi)P_{0,n_2+1,0}(t)+\mu_3 P_{0,n_2,1}(t), \qquad n_2>0$$

$$\frac{\partial P_{0,0,n_3}(t)}{\partial t} = -(\lambda(t)+\varepsilon+n_3\mu_3)P_{0,0,n_3}(t)+\mu_1(1-\theta)P_{1,0,n_3}(t)+\mu_2\pi P_{0,1,n_3-1}(t)$$
$$+\mu_2(1-\pi)P_{0,1,n_3}(t)+(n_3+1)\mu_3 P_{0,0,n_3+1}(t), \qquad n_3>0$$

$$\frac{\partial P_{0,0,0}(t)}{\partial t} = -(\lambda(t)+\varepsilon)P_{0,0,0}(t)+\mu_1(1-\theta)P_{1,0,0}(t)+\mu_2(1-\pi)P_{0,1,0}(t)+\mu_3 P_{0,0,1}(t)$$

(2.1)

Let $P(s_1,s_2,s_3;t)=\sum_{n_1=0}^{\infty}\sum_{n_2=0}^{\infty}\sum_{n_3=0}^{\infty} P_{n_1,n_2,n_3}(t)s_1^{n_1}s_2^{n_2}s_3^{n_3}$ be the joint probability generating function of $P_{n_1,n_2,n_3}(t)$. (2.2)

Multiplying the equation (5.2.1) with $s_1^{n_1} s_2^{n_2} s_3^{n_3}$ and summing over all n_1, n_2 and n_3 we get

$$\frac{dP}{dt} = -\lambda(t)P(s_1,s_2,s_3;t)+\lambda(t)s_1 P(s_1,s_2,s_3;t)-\varepsilon P(s_1,s_2,s_3;t)$$
$$+\varepsilon s_2 P(s_1,s_2,s_3;t)-\mu_1 s_1\frac{\partial P(s_1,s_2,s_3;t)}{\partial s_1}+\mu_1 s_2\theta\frac{\partial P(s_1,s_2,s_3;t)}{\partial s_1}$$
$$+\mu_1(1-\theta)\frac{\partial P(s_1,s_2,s_3;t)}{\partial s_1}-\mu_2 s_2\frac{\partial P(s_1,s_2,s_3;t)}{\partial s_2}+\mu_2 s_3\pi\frac{\partial P(s_1,s_2,s_3;t)}{\partial s_2}$$
$$+\mu_2(1-\pi)\frac{\partial P(s_1,s_2,s_3;t)}{\partial s_2}-\mu_3 s_3\frac{\partial P(s_1,s_2,s_3;t)}{\partial s_3}+\mu_3\frac{\partial P(s_1,s_2,s_3;t)}{\partial s_3}$$

(2.3)

After simplifying, we get

$$\frac{dP}{dt}=\lambda(t)(s_1-1)P(s_1,s_2,s_3;t)+\varepsilon(s_2-1)P(s_1,s_2,s_3;t)-\mu_1(s_1-1-\theta(s_2-1))\frac{\partial P(s_1,s_2,s_3;t)}{\partial s_1}$$
$$-\mu_2(s_2-1-\pi(s_3-1))\frac{\partial P(s_1,s_2,s_3;t)}{\partial s_2}-\mu_3(s_3-1)\frac{\partial P(s_1,s_2,s_3;t)}{\partial s_3}$$

(2.4)

Solving the equation (5.2.4) by Lagrangian's method, the auxiliary equations are

$$\frac{dt}{1}=\frac{ds_1}{\mu_1(s_1-1+\theta(1-s_2))}=\frac{ds_2}{\mu_2(s_2-1+\pi(1-s_3))}=\frac{ds_3}{\mu_3(s_3-1)}$$
$$=\frac{dP}{\lambda(t)P(s_1,s_2,s_3;t)(s_1-1)+\varepsilon P(s_1,s_2,s_3;t)(s_2-1)}$$

(2.5)

To solve the equations in (2.5) the functional form of λ(t) is required. Let the mean arrival rate of packets is λ(t) = λ + α t, where λ > 0, α > 0 are constants.

Solving the first and fourth terms in equation (2.5), we get

$$a=(s_3-1)e^{-\mu_3 t} \qquad (2.6a)$$

Solving the first and third terms in equation (2.5), we get

$$b=(s_2-1)e^{-\mu_2 t}+\frac{\pi\mu_2}{\mu_3-\mu_2}(s_3-1)e^{-\mu_2 t} \qquad (2.6b)$$

Solving the first and second terms in equation (2.5), we get

$$c=(s_1-1)e^{-\mu_1 t}+\frac{\theta\mu_1}{\mu_2-\mu_1}(s_2-1)e^{-\mu_1 t}+\frac{\pi\theta\mu_1\mu_2}{(\mu_3-\mu_1)(\mu_2-\mu_1)}(s_3-1)e^{-\mu_1 t}$$

(2.6c)

Solving the first and fifth terms in equation (2.5), we get

$$d=P(s_1,s_2,s_3;t)\exp\left\{-\left[\frac{(s_1-1)}{\mu_1}\left(\lambda+\alpha t-\frac{\alpha}{\mu_1}\right)+\frac{(s_2-1)}{\mu_2}\left(\varepsilon+\theta\left(\lambda+\alpha t-\frac{\alpha(\mu_1+\mu_2)}{\mu_1\mu_2}\right)\right)\right.\right.$$
$$\left.\left.+\frac{(s_3-1)\pi}{\mu_3}\left(\varepsilon+\theta\left(\lambda+\alpha t-\frac{\alpha(\mu_1\mu_2+\mu_2\mu_3+\mu_1\mu_3)}{\mu_1\mu_2\mu_3}\right)\right)\right]\right\}$$

(2.6d)

where, a, b, c and d are arbitrary constants.

Using the initial conditions $P_{000}(0)=1$, $P_{000}(t)=0 \;\; \forall\, t>0$. The general solution of (2.5) gives the probability generating function of the number of packets in the first, second and third buffers at time t, as

$$P(s_1,s_2,s_3;t)=\exp\left\{\frac{(s_1-1)}{\mu_1}\left(1-e^{-\mu_1 t}\right)\left(\lambda-\frac{\alpha}{\mu_1}\right)+\frac{(s_1-1)\alpha t}{\mu_1}+\left(\frac{(s_2-1)}{\mu_2}\right)\left(1-e^{-\mu_2 t}\right)\left(\varepsilon+\theta\left(\lambda-\alpha\left(\frac{1}{\mu_2}+\frac{1}{\mu_1}\right)\right)\right)\right.$$
$$+\left(\frac{(s_2-1)\theta}{\mu_2}\right)\alpha t+\left(\frac{(s_2-1)\theta}{\mu_2-\mu_1}\right)\left(e^{-\mu_2 t}-e^{-\mu_1 t}\right)\left(\lambda-\frac{\alpha}{\mu_1}\right)+\left(\frac{(s_3-1)\pi}{\mu_3}\right)\left(1-e^{-\mu_3 t}\right)\left(\varepsilon+\theta\left(\lambda-\alpha\left(\frac{1}{\mu_3}+\frac{1}{\mu_2}+\frac{1}{\mu_1}\right)\right)\right)$$
$$+\left(\frac{(s_3-1)\pi\theta\alpha t}{\mu_3}\right)+\left(\frac{(s_3-1)\pi\theta}{\mu_3-\mu_2}\right)\left(e^{-\mu_3 t}-e^{-\mu_2 t}\right)\left(\lambda-\alpha\left(\frac{1}{\mu_2}-\frac{1}{\mu_1}\right)\right)$$
$$\left.+(s_2-1)\pi\theta\mu_2\left(\frac{e^{-\mu_3 t}}{(\mu_2-\mu_3)(\mu_3-\mu_1)}+\frac{e^{-\mu_2 t}}{(\mu_3-\mu_2)(\mu_2-\mu_1)}+\frac{e^{-\mu_1 t}}{(\mu_1-\mu_3)(\mu_2-\mu_1)}\right)\left(\lambda-\frac{\alpha}{\mu_1}\right)\right\}$$

for λ < min{ μ1, μ2, μ3} (2.7)

3 PERFORMANCE MEASURE OF THE NETWORK:

In this section, we derive and analyze the performance measures of the communication network under transient conditions. Expanding P(s_1,s_2, s_3;t) given in equation (2.7) and collecting the constant terms, we get the probability that the network is empty as

$$P_{000}(t)=\exp\left\{-\left[\frac{1}{\mu_1}\left(1-e^{-\mu_1 t}\right)\left(\lambda-\frac{\alpha}{\mu_1}\right)+\frac{\alpha t}{\mu_1}+\frac{1}{\mu_2}\left(1-e^{-\mu_2 t}\right)\left(\varepsilon+\theta\left(\lambda-\alpha\left(\frac{1}{\mu_2}+\frac{1}{\mu_1}\right)\right)\right)\right.\right.$$
$$+\frac{\theta\alpha t}{\mu_2}+\frac{\theta}{\mu_2-\mu_1}\left(e^{-\mu_2 t}-e^{-\mu_1 t}\right)\left(\lambda-\frac{\alpha}{\mu_1}\right)+\frac{\pi}{\mu_3}\left(1-e^{-\mu_2 t}\right)\left(\varepsilon+\theta\left(\lambda-\alpha\left(\frac{1}{\mu_3}+\frac{1}{\mu_2}+\frac{1}{\mu_1}\right)\right)\right)$$
$$+\frac{\pi\theta\alpha t}{\mu_3}+\frac{\pi\theta}{(\mu_3-\mu_2)}\left(e^{-\mu_3 t}-e^{-\mu_2 t}\right)\left(\lambda-\alpha\left(\frac{1}{\mu_2}+\frac{1}{\mu_1}\right)\right)$$
$$\left.\left.+\pi\theta\mu_2\left(\frac{e^{-\mu_3 t}}{(\mu_2-\mu_3)(\mu_3-\mu_1)}+\frac{e^{-\mu_2 t}}{(\mu_3-\mu_2)(\mu_2-\mu_1)}+\frac{e^{-\mu_1 t}}{(\mu_1-\mu_3)(\mu_2-\mu_1)}\right)\left(\lambda-\frac{\alpha}{\mu_1}\right)\right]\right\}$$

(3.1)

Taking s_2 = 1, s_3 = 1 in equation (2.7), we get the probability generating function of the first buffer size as

$$P(s_1,t)=\exp\left\{\frac{(s_1-1)}{\mu_1}\left(\lambda-\frac{\alpha}{\mu_1}\right)\left(1-e^{-\mu_1 t}\right)+\frac{(s_1-1)}{\mu_1}\alpha t\right\}$$

for λ < μ1 (3.2)

Expanding P(s_1,t) and collecting the constant terms, we get the probability that the first buffer is empty as

$$P_{0..}(t)=\exp\left\{-\left[\frac{1}{\mu_1}\left(\lambda-\frac{\alpha}{\mu_1}\right)\left(1-e^{-\mu_1 t}\right)+\frac{1}{\mu_1}\alpha t\right]\right\} \quad (3.3)$$

Taking $s_1 = 1$, $s_2 = 1$ in equation (2.7), we get the probability generating function of the second buffer size as

$$P(s_2,t)=\exp\left\{\frac{(s_2-1)}{\mu_2}\left(1-e^{-\mu_2 t}\right)\left(\varepsilon+\theta\left(\lambda-\alpha\left(\frac{1}{\mu_2}+\frac{1}{\mu_1}\right)\right)\right)+\frac{(s_2-1)\theta}{\mu_2}\alpha t+\frac{(s_2-1)\theta}{\mu_2-\mu_1}\left(e^{-\mu_2 t}-e^{-\mu_1 t}\right)\left(\lambda-\frac{\alpha}{\mu_1}\right)\right\}$$

$$\text{for } \lambda < \min\{\mu_1,\mu_2\} \quad (3.4)$$

Expanding P(s_2,t) and collecting the constant terms, we get the probability that the second buffer is empty as

$$P_{.0.}(t)=\exp\left[-\left[\frac{1}{\mu_2}\left(1-e^{-\mu_2 t}\right)\left(\varepsilon+\theta\left(\lambda-\alpha\left(\frac{1}{\mu_2}+\frac{1}{\mu_1}\right)\right)\right)+\frac{\theta}{\mu_2}\alpha t+\frac{\theta}{\mu_2-\mu_1}\left(e^{-\mu_2 t}-e^{-\mu_1 t}\right)\left(\lambda-\frac{\alpha}{\mu_1}\right)\right]\right] \quad (3.5)$$

Taking $s_1 = 1$, $s_2 = 1$ in equation (2.7),we get the probability generating function of the third buffer size as

$$\begin{aligned}P(s_3,t)=\exp\Bigg\{&\frac{(s_3-1)\pi}{\mu_3}\left(1-e^{-\mu_3 t}\right)\left(\varepsilon+\theta\left(\lambda-\alpha\left(\frac{1}{\mu_3}+\frac{1}{\mu_2}+\frac{1}{\mu_1}\right)\right)\right)+\frac{(s_3-1)\pi\theta\alpha t}{\mu_3}+\frac{(s_3-1)\pi\theta}{\mu_3-\mu_2}\left(e^{-\mu_3 t}-e^{-\mu_2 t}\right)\left(\lambda-\alpha\left(\frac{1}{\mu_2}+\frac{1}{\mu_1}\right)\right)\\&+(s_3-1)\mu_2\pi\theta\left(\frac{e^{-\mu_3 t}}{(\mu_2-\mu_3)(\mu_3-\mu_1)}+\frac{e^{-\mu_2 t}}{(\mu_3-\mu_2)(\mu_2-\mu_1)}+\frac{e^{-\mu_1 t}}{(\mu_1-\mu_3)(\mu_2-\mu_1)}\right)\left(\lambda-\frac{\alpha}{\mu_1}\right)\Bigg\}\end{aligned}$$

$$\text{for } \lambda < \min\{\mu_1,\mu_2,\mu_3\} \quad (3.6)$$

Expanding P(s_3,t) and collecting the constant terms, we get the probability that the third buffer is empty as

$$\begin{aligned}P_{..0}(t)=\exp\Bigg\{-\Bigg[&\frac{\pi}{\mu_3}\left(1-e^{-\mu_3 t}\right)\left(\varepsilon+\theta\left(\lambda-\alpha\left(\frac{1}{\mu_3}+\frac{1}{\mu_2}+\frac{1}{\mu_1}\right)\right)\right)+\frac{\pi\theta\alpha t}{\mu_3}+\frac{\pi\theta}{\mu_3-\mu_2}\left(e^{-\mu_3 t}-e^{-\mu_2 t}\right)\left(\lambda-\alpha\left(\frac{1}{\mu_2}+\frac{1}{\mu_1}\right)\right)\\&+\mu_2\pi\theta\left(\frac{e^{-\mu_3 t}}{(\mu_2-\mu_3)(\mu_3-\mu_1)}+\frac{e^{-\mu_2 t}}{(\mu_3-\mu_2)(\mu_2-\mu_1)}+\frac{e^{-\mu_1 t}}{(\mu_1-\mu_3)(\mu_2-\mu_1)}\right)\left(\lambda-\frac{\alpha}{\mu_1}\right)\Bigg]\Bigg\}\end{aligned} \quad (3.7)$$

The mean number of packets in the first buffer is

$$L_1(t)=\frac{\partial P(s_1,t)}{\partial s_1}\Big|_{s_1=1}=\frac{1}{\mu_1}\left(\lambda-\frac{\alpha}{\mu_1}\right)\left(1-e^{-\mu_1 t}\right)+\frac{1}{\mu_1}\alpha t \quad (3.8)$$

The utilization of the first transmitter is

$$U_1(t)=1-P_{0..}(t)=1-\exp\left\{-\left[\frac{1}{\mu_1}\left(\lambda-\frac{\alpha}{\mu_1}\right)\left(1-e^{-\mu_1 t}\right)+\frac{1}{\mu_1}\alpha t\right]\right\} \quad (3.9)$$

The mean number of the packets in second buffer is

$$L_2(t)=\frac{\partial P(s_2,t)}{\partial s_2}\Big|_{s_2=1}=\frac{1}{\mu_2}\left(1-e^{-\mu_2 t}\right)\left(\varepsilon+\theta\left(\lambda-\alpha\left(\frac{1}{\mu_2}+\frac{1}{\mu_1}\right)\right)\right)+\frac{\theta}{\mu_2}\alpha t+\frac{\theta}{\mu_2-\mu_1}\left(e^{-\mu_2 t}-e^{-\mu_1 t}\right)\left(\lambda-\frac{\alpha}{\mu_1}\right) \quad (3.10)$$

The utilization of the second transmitter is

$$U_2(t)=1-P_{.0.}(t)=1-\exp\left\{-\left[\frac{1}{\mu_2}\left(1-e^{-\mu_2 t}\right)\left(\varepsilon+\theta\left(\lambda-\alpha\left(\frac{1}{\mu_2}+\frac{1}{\mu_1}\right)\right)\right)+\frac{\theta}{\mu_2}\alpha t+\frac{\theta}{\mu_2-\mu_1}\left(e^{-\mu_2 t}-e^{-\mu_1 t}\right)\left(\lambda-\frac{\alpha}{\mu_1}\right)\right]\right\} \quad (3.11)$$

The mean number of the packets in third buffer is

$$\begin{aligned}L_3(t)&=\frac{\partial P(s_3,t)}{\partial s_3}\Big|_{s_3=1}\\&=\frac{\pi}{\mu_3}\left(1-e^{-\mu_3 t}\right)\left(\varepsilon+\theta\left(\lambda-\alpha\left(\frac{1}{\mu_3}+\frac{1}{\mu_2}+\frac{1}{\mu_1}\right)\right)\right)+\frac{\pi\theta\alpha t}{\mu_3}+\frac{\pi\theta}{\mu_3-\mu_2}\left(e^{-\mu_3 t}-e^{-\mu_2 t}\right)\left(\lambda-\alpha\left(\frac{1}{\mu_2}+\frac{1}{\mu_1}\right)\right)\\&\quad+\pi\theta\mu_2\left(\frac{e^{-\mu_3 t}}{(\mu_2-\mu_3)(\mu_3-\mu_1)}+\frac{e^{-\mu_2 t}}{(\mu_3-\mu_2)(\mu_2-\mu_1)}+\frac{e^{-\mu_1 t}}{(\mu_1-\mu_3)(\mu_2-\mu_1)}\right)\left(\lambda-\frac{\alpha}{\mu_1}\right)\end{aligned} \quad (3.12)$$

The utilization of the third transmitter is

$$\begin{aligned}U_3(t)=1-P_{..0}(t)=1-\exp\Bigg\{-\Bigg[&\frac{\pi}{\mu_3}\left(1-e^{-\mu_3 t}\right)\left(\varepsilon+\theta\left(\lambda-\alpha\left(\frac{1}{\mu_3}+\frac{1}{\mu_2}+\frac{1}{\mu_1}\right)\right)\right)+\frac{\pi\theta\alpha t}{\mu_3}+\frac{\pi\theta}{\mu_3-\mu_2}\left(e^{-\mu_3 t}-e^{-\mu_2 t}\right)\left(\lambda-\alpha\left(\frac{1}{\mu_2}+\frac{1}{\mu_1}\right)\right)\\&+\pi\theta\mu_2\left(\frac{e^{-\mu_3 t}}{(\mu_2-\mu_3)(\mu_3-\mu_1)}+\frac{e^{-\mu_2 t}}{(\mu_3-\mu_2)(\mu_2-\mu_1)}+\frac{e^{-\mu_1 t}}{(\mu_1-\mu_3)(\mu_2-\mu_1)}\right)\left(\lambda-\frac{\alpha}{\mu_1}\right)\Bigg]\Bigg\}\end{aligned} \quad (3.13)$$

The variance of the number of packets in the first buffer is

$$V_1(t)=\frac{1}{\mu_1}\left(\lambda-\frac{\alpha}{\mu_1}\right)\left(1-e^{-\mu_1 t}\right)+\frac{1}{\mu_1}\alpha t \quad (3.14)$$

The variance of the number of packets in the second buffer is

$$V_2(t)=\frac{1}{\mu_2}\left(1-e^{-\mu_2 t}\right)\left(\varepsilon+\theta\left(\lambda-\alpha\left(\frac{1}{\mu_2}+\frac{1}{\mu_1}\right)\right)\right)+\frac{\theta}{\mu_2}\alpha t+\frac{\theta}{\mu_2-\mu_1}\left(e^{-\mu_2 t}-e^{-\mu_1 t}\right)\left(\lambda-\frac{\alpha}{\mu_1}\right) \quad (3.15)$$

The variance of the number of packets in the third buffer is

$$\begin{aligned}V_3(t)=&\frac{\pi}{\mu_3}\left(1-e^{-\mu_3 t}\right)\left(\varepsilon+\theta\left(\lambda-\alpha\left(\frac{1}{\mu_3}+\frac{1}{\mu_2}+\frac{1}{\mu_1}\right)\right)\right)+\frac{\pi\theta\alpha t}{\mu_3}+\frac{\pi\theta}{\mu_3-\mu_2}\left(e^{-\mu_3 t}-e^{-\mu_2 t}\right)\left(\lambda-\alpha\left(\frac{1}{\mu_2}+\frac{1}{\mu_1}\right)\right)\\&+\pi\theta\mu_2\left[\frac{e^{-\mu_3 t}}{(\mu_2-\mu_3)(\mu_3-\mu_1)}+\frac{e^{-\mu_2 t}}{(\mu_3-\mu_2)(\mu_2-\mu_1)}+\frac{e^{-\mu_1 t}}{(\mu_1-\mu_3)(\mu_2-\mu_1)}\right]\left(\lambda-\frac{\alpha}{\mu_1}\right)\end{aligned} \quad (3.16)$$

The throughput of the first transmitter is

$$\mu_1\left(1-P_{0..}(t)\right)=\mu_1\left[1+\exp\left\{\frac{1}{\mu_1}\left(\lambda-\frac{\alpha}{\mu_1}\right)\left(1-e^{-\mu_1 t}\right)+\frac{1}{\mu_1}\alpha t\right\}\right] \quad (3.17)$$

The mean delay in the first buffer is

$$W_1(t)=\frac{L_1(t)}{\mu_1\left(1-P_{0..}(t)\right)}=\frac{\frac{1}{\mu_1}\left(\lambda-\frac{\alpha}{\mu_1}\right)\left(1-e^{-\mu_1 t}\right)+\frac{1}{\mu_1}\alpha t}{\mu_1\left[1+\exp\left\{\frac{1}{\mu_1}\left(\lambda-\frac{\alpha}{\mu_1}\right)\left(1-e^{-\mu_1 t}\right)+\frac{1}{\mu_1}\alpha t\right\}\right]} \quad (3.18)$$

The throughput of the second transmitter is

$$\mu_2\left(1-P_{.0.}(t)\right)=\mu_2\left[1+\exp\left[\frac{1}{\mu_2}\left(1-e^{-\mu_2 t}\right)\left(\varepsilon+\theta\left(\lambda-\alpha\left(\frac{1}{\mu_2}+\frac{1}{\mu_1}\right)\right)\right)+\frac{\theta}{\mu_2}\alpha t+\frac{\theta}{\mu_2-\mu_1}\left(e^{-\mu_2 t}-e^{-\mu_1 t}\right)\left(\lambda-\frac{\alpha}{\mu_1}\right)\right]\right] \quad (3.19)$$

The mean delay in the second buffer is

$$\begin{aligned}W_2(t)&=\frac{L_2(t)}{\mu_2(1-P_{.0.}(t)}\\&=\frac{\frac{1}{\mu_2}\left(1-e^{-\mu_2 t}\right)\left(\varepsilon+\theta\left(\lambda-\alpha\left(\frac{1}{\mu_2}+\frac{1}{\mu_1}\right)\right)\right)+\frac{\theta}{\mu_2}\alpha t+\frac{\theta}{\mu_2-\mu_1}\left(e^{-\mu_2 t}-e^{-\mu_1 t}\right)\left(\lambda-\frac{\alpha}{\mu_1}\right)}{\mu_2\left[1+\exp\left[\frac{1}{\mu_2}\left(1-e^{-\mu_2 t}\right)\left(\varepsilon+\theta\left(\lambda-\alpha\left(\frac{1}{\mu_2}+\frac{1}{\mu_1}\right)\right)\right)+\frac{\theta}{\mu_2}\alpha t+\frac{\theta}{\mu_2-\mu_1}\left(e^{-\mu_2 t}-e^{-\mu_1 t}\right)\left(\lambda-\frac{\alpha}{\mu_1}\right)\right]\right]}\end{aligned} \quad (3.20)$$

The throughput of the third transmitter is

$$\begin{aligned}\mu_3(1-P_{..0}(t))=\mu_3\Bigg\{1+\exp\Bigg[&\frac{\pi}{\mu_3}\left(1-e^{-\mu_3 t}\right)\left(\varepsilon+\theta\left(\lambda-\alpha\left(\frac{1}{\mu_3}+\frac{1}{\mu_2}+\frac{1}{\mu_1}\right)\right)\right)+\frac{\pi\theta\alpha t}{\mu_3}+\frac{\pi\theta}{\mu_3-\mu_2}\left(e^{-\mu_3 t}-e^{-\mu_2 t}\right)\left(\lambda-\alpha\left(\frac{1}{\mu_2}+\frac{1}{\mu_1}\right)\right)\\&+\pi\theta\mu_2\left(\frac{e^{-\mu_3 t}}{(\mu_2-\mu_3)(\mu_3-\mu_1)}+\frac{e^{-\mu_2 t}}{(\mu_3-\mu_2)(\mu_2-\mu_1)}+\frac{e^{-\mu_1 t}}{(\mu_1-\mu_3)(\mu_2-\mu_1)}\right)\left(\lambda-\frac{\alpha}{\mu_1}\right)\Bigg]\Bigg\}\end{aligned} \quad (3.21)$$

The mean delay in the third buffer is

$$W_3(t)=\frac{L_3(t)}{\mu_3\left(1-P_{..0}(t)\right)}$$

$$
= \frac{\frac{\pi}{\mu_3}\left(1-e^{-\mu_3 t}\right)\left(\varepsilon+\theta\left(\lambda-\alpha\left(\frac{1}{\mu_3}+\frac{1}{\mu_2}+\frac{1}{\mu_1}\right)\right)\right)+\frac{\pi\theta\alpha t}{\mu_3}+\frac{\pi\theta}{\mu_3-\mu_2}\left(e^{-\mu_3 t}-e^{-\mu_2 t}\right)\left(\lambda-\alpha\left(\frac{1}{\mu_2}+\frac{1}{\mu_1}\right)\right) + \pi\theta\mu_2\left(\frac{e^{-\mu_3 t}}{(\mu_2-\mu_3)(\mu_3-\mu_1)}+\frac{e^{-\mu_2 t}}{(\mu_3-\mu_2)(\mu_2-\mu_1)}+\frac{e^{-\mu_1 t}}{(\mu_1-\mu_3)(\mu_2-\mu_1)}\right)\left(\lambda-\frac{\alpha}{\mu_1}\right)}{\mu_3\left\{1+\exp\left[\frac{\pi\theta}{\mu_3}\left(1-e^{-\mu_3 t}\right)\left(\varepsilon+\theta\left(\lambda-\alpha\left(\frac{1}{\mu_3}+\frac{1}{\mu_2}+\frac{1}{\mu_1}\right)\right)\right)+\frac{\pi\theta\alpha t}{\mu_3}+\frac{\pi\theta}{\mu_3-\mu_2}\left(e^{-\mu_3 t}-e^{-\mu_2 t}\right)\left(\lambda-\alpha\left(\frac{1}{\mu_2}+\frac{1}{\mu_1}\right)\right) \pi\theta\mu_2\left(\frac{e^{-\mu_3 t}}{(\mu_2-\mu_3)(\mu_3-\mu_1)}+\frac{e^{-\mu_2 t}}{(\mu_3-\mu_2)(\mu_2-\mu_1)}+\frac{e^{-\mu_1 t}}{(\mu_1-\mu_3)(\mu_2-\mu_1)}\right)\left(\lambda-\frac{\alpha}{\mu_1}\right)\right]\right\}} \tag{3.22}
$$

The mean number of packets in the entire network at time t is

$$
\begin{aligned}
L(t) = {} & \frac{1}{\mu_1}\left(1-e^{-\mu_1 t}\right)\left(\lambda-\frac{\alpha}{\mu_1}\right)+\frac{1}{\mu_1}\alpha t+\frac{1}{\mu_2}\left(1-e^{-\mu_2 t}\right)\left(\varepsilon+\theta\left(\lambda-\alpha\left(\frac{1}{\mu_2}+\frac{1}{\mu_1}\right)\right)\right)+\frac{\theta}{\mu_2}\alpha t \\
& +\frac{\theta}{\mu_2-\mu_1}\left(e^{-\mu_2 t}-e^{-\mu_1 t}\right)\left(\lambda-\frac{\alpha}{\mu_1}\right)+\frac{\pi}{\mu_3}\left(1-e^{-\mu_3 t}\right)\left(\varepsilon+\theta\left(\lambda-\alpha\left(\frac{1}{\mu_3}+\frac{1}{\mu_2}+\frac{1}{\mu_1}\right)\right)\right) \\
& +\frac{\pi\theta\alpha t}{\mu_3}+\frac{\pi\theta}{\mu_3-\mu_2}\left(e^{-\mu_3 t}-e^{-\mu_2 t}\right)\left(\lambda-\alpha\left(\frac{1}{\mu_2}+\frac{1}{\mu_1}\right)\right) \\
& +\pi\theta\mu_2\left(\frac{e^{-\mu_3 t}}{(\mu_2-\mu_3)(\mu_3-\mu_1)}+\frac{e^{-\mu_2 t}}{(\mu_3-\mu_2)(\mu_2-\mu_1)}+\frac{e^{-\mu_1 t}}{(\mu_1-\mu_3)(\mu_2-\mu_1)}\right)\left(\lambda-\frac{\alpha}{\mu_1}\right)
\end{aligned} \tag{3.23}
$$

The variability of the number of packets in the network is

$$
\begin{aligned}
\operatorname{var}(N) = {} & \frac{1}{\mu_1}\left(1-e^{-\mu_1 t}\right)\left(\lambda-\frac{\alpha}{\mu_1}\right)+\frac{1}{\mu_1}\alpha t+\frac{\theta}{\mu_2}\left(1-e^{-\mu_2 t}\right)\left(\varepsilon+\theta\left(\lambda-\alpha\left(\frac{1}{\mu_2}+\frac{1}{\mu_1}\right)\right)\right)+\frac{\theta}{\mu_2}\alpha t \\
& +\frac{\theta}{\mu_2-\mu_1}\left(e^{-\mu_2 t}-e^{-\mu_1 t}\right)\left(\lambda-\frac{\alpha}{\mu_1}\right)+\frac{\pi}{\mu_3}\left(1-e^{-\mu_3 t}\right)\left(\varepsilon+\theta\left(\lambda-\alpha\left(\frac{1}{\mu_3}+\frac{1}{\mu_2}+\frac{1}{\mu_1}\right)\right)\right) \\
& +\frac{\pi\theta\alpha t}{\mu_3}+\frac{\pi\theta}{\mu_3-\mu_2}\left(e^{-\mu_3 t}-e^{-\mu_2 t}\right)\left(\lambda-\alpha\left(\frac{1}{\mu_2}+\frac{1}{\mu_1}\right)\right) \\
& +\pi\theta\mu_2\left(\frac{e^{-\mu_3 t}}{(\mu_2-\mu_3)(\mu_3-\mu_1)}+\frac{e^{-\mu_2 t}}{(\mu_3-\mu_2)(\mu_2-\mu_1)}+\frac{e^{-\mu_1 t}}{(\mu_1-\mu_3)(\mu_2-\mu_1)}\right)\left(\lambda-\frac{\alpha}{\mu_1}\right)
\end{aligned} \tag{3.24}
$$

4 PERFORMANCE EVALUATION OF THE NETWORK

In this section, the performance of the communication network proposed in this section is discussed through numerical illustration. Different values of the parameters are considered for bandwidth allocation and arrival of packets. After interacting with the technical staff at internet providing station, it is observed that the packet parameter λ varies from $2x10^4$ packets/sec to $7x10^4$ packets/sec, α varies from -0.5 to 1.5, θ varies from 0.1 to 0.9, ϵ varies from 0.2 to 0.6 π varies from 0.1 to 0.9, with an average packet size of 53 bytes. After transmitting from node one, the forward transmission rate μ_1 varies from $5x10^4$ packets/sec to $9x10^4$ packets/sec. The rate of transmission from node two μ_2 varies from $15x10^4$ packets/sec to $19x10^4$ packets/sec. The rate of transmission from node three μ_3 varies from $25x10^4$ packets/sec to $29x10^4$ packets/sec. In all the nodes, the dynamic bandwidth allocation strategy is considered i.e., the transmission rate of each packet depends on the number of packets in the buffer connected to it at that instant.

From equations (3.9), (3.11), (3.13), (3.17), (3.19) and (3.21), the utilization of the transmitters and throughput of three nodes are computed for different values of the parameters t, λ, α, θ, ϵ, π, μ_1, μ_2, μ_3 are presented in table 1. The relationship between parameters and utilization of transmitters and throughput of three nodes are shown in Figure 1.

As the time (t) and parameter (λ) increases, the utilization of transmitters are increases for fixed values of the other parameters. It is also observed that as the parameter (α) increases, the utilization of transmitters at three nodes are increases for fixed values of the other parameters. The parameter (θ) increases, the utilization of transmitters at first node is fixed and second and third nodes are increasing for fixed values of the other parameters. The parameter (ϵ) increases, the utilization of first node is constant and second and third nodes are increasing for fixed values of the other parameters. The parameter (π) increases, the utilization of transmitters at first and second nodes are fixed and the third node it is increases for fixed values of the other parameters. As the transmission rate (μ_1) increases, the utilization of the first node decreases and the other two nodes increases when the other parameters remain fixed. As the transmission rate (μ_2) increases, the utilization of first node is constant and the utilization of the second node decreases and third node increases when the other parameters remain fixed. Similarly, as the transmission rate (μ_3) increases the utilization of the first and second nodes are fixed and third node is decreases when other parameters remain fixed.

It is observed as time value (t) increases, the throughput of first, second and third nodes are increasing for fixed values of the other parameters. As the parameter (λ) varies from $3x10^4$ packets/sec to $7x10^4$ packets/sec, the throughput of the first, second and third nodes are increasing when other parameters remain fixed. When the arrival parameter (α) varies from -0.5 to 1.5, the throughput of the first, second and third nodes are increasing when other parameters remain fixed. When the parameter (θ) varies from 0.1 to 0.9, the throughput of the first node is constant, second and third nodes are increasing when other parameters remain fixed. When the parameter (ϵ) varies from 0.2 to 0.6, the throughput of the first node is constant, second and third nodes are increasing when other parameters remain fixed. When the parameter (π) varies from 0.1 to 0.9, the throughput of the first node and second nodes are constant and third node is increases when other parameters remain fixed. When the transmission rate (μ_1) varies from $5x10^4$ packets/sec to $9x10^4$ packets/sec, the throughput of the first, second and the third nodes are increasing when other parameters remain fixed. The transmission rate (μ_2) varies from $15x10^4$ packets/sec to $19x10^4$ packets/sec, the throughput of the first node remains constant, the second and the third nodes are increasing when other parameters remain fixed. Similarly the transmission rate (μ_3) varies from $25x10^4$

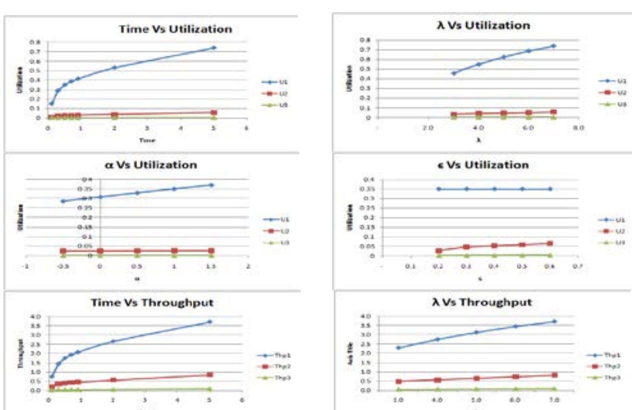
packets/sec to 29x10^4 packets/sec, the throughput of the first and second nodes remain constant and for the third node is increasing when other parameters remain fixed.

Table 1

Values of Emptiness probabilities and Utilization of the communication Network with DBA and NHP arrivals

t^*	$\lambda^\$$	α	θ	ϵ	π	$\mu_1^\$$	$\mu_2^\$$	$\mu_3^\$$	$U_1(t)$	$U_2(t)$	$U_3(t)$	Thp_1	Thp_2	Thp_3
0.1	2	1	0.1	0.2	0.1	5	15	25	0.14926	0.01302	0.00092	0.74630	0.19523	0.02300
0.3	2	1	0.1	0.2	0.1	5	15	25	0.28800	0.02254	0.00143	1.43998	0.33807	0.03564
0.5	2	1	0.1	0.2	0.1	5	15	25	0.34978	0.02639	0.00161	1.74892	0.39580	0.04036
0.7	2	1	0.1	0.2	0.1	5	15	25	0.38684	0.02860	0.00173	1.93420	0.42893	0.04333
0.9	2	1	0.1	0.2	0.1	5	15	25	0.41492	0.03022	0.00183	2.07458	0.45334	0.04568
2.0	2	1	0.1	0.2	0.1	5	15	25	0.53233	0.03750	0.00227	2.66163	0.56250	0.05687
5.0	2	1	0.1	0.2	0.1	5	15	25	0.74334	0.05656	0.00347	3.71670	0.84840	0.08678
0.5	3	1	0.1	0.2	0.1	5	15	25	0.45884	0.03206	0.00201	2.29418	0.48095	0.05033
0.5	4	1	0.1	0.2	0.1	5	15	25	0.54960	0.03771	0.00241	2.74799	0.56560	0.06029
0.5	5	1	0.1	0.2	0.1	5	15	25	0.62514	0.04332	0.00281	3.12569	0.64976	0.07025
0.5	6	1	0.1	0.2	0.1	5	15	25	0.68801	0.04890	0.00321	3.44005	0.73344	0.08021
0.5	7	1	0.1	0.2	0.1	5	15	25	0.74034	0.05444	0.00361	3.70168	0.81662	0.09016
0.5	2	-	0.1	0.2	0.1	5	15	25	0.28504	0.02387	0.00150	1.42518	0.35808	0.03746
0.5	2	0	0.1	0.2	0.1	5	15	25	0.30731	0.02471	0.00154	1.53653	0.37066	0.03842
0.5	2	0.5	0.1	0.2	0.1	5	15	25	0.32888	0.02555	0.00158	1.64440	0.38324	0.03939
0.5	2	1	0.1	0.2	0.1	5	15	25	0.34978	0.02639	0.00161	1.74892	0.39580	0.04036
0.5	2	1.5	0.1	0.2	0.1	5	15	25	0.37004	0.02722	0.00165	1.85018	0.40835	0.04133
0.5	2	1	0.1	0.2	0.1	5	15	25	0.34978	0.02639	0.00161	1.74892	0.39580	0.04036
0.5	2	1	0.3	0.2	0.1	5	15	25	0.34978	0.05216	0.00324	1.74892	0.78242	0.08104
0.5	2	1	0.5	0.2	0.1	5	15	25	0.34978	0.07725	0.00487	1.74892	1.15880	0.12166
0.5	2	1	0.7	0.2	0.1	5	15	25	0.34978	0.10168	0.00649	1.74892	1.52522	0.16221
0.5	2	1	0.9	0.2	0.1	5	15	25	0.34978	0.12546	0.00811	1.74892	1.88194	0.20269
0.5	2	1	0.1	0.2	0.1	5	15	25	0.34978	0.02639	0.00161	1.74892	0.39580	0.04036
0.5	2	1	0.1	0.3	0.1	5	15	25	0.34978	0.03285	0.00201	1.74892	0.49278	0.05034
0.5	2	1	0.1	0.4	0.1	5	15	25	0.34978	0.03927	0.00241	1.74892	0.58912	0.06032
0.5	2	1	0.1	0.5	0.1	5	15	25	0.34978	0.04565	0.00281	1.74892	0.68482	0.07029
0.5	2	1	0.1	0.6	0.1	5	15	25	0.34978	0.05199	0.00321	1.74892	0.77989	0.08026
0.5	2	1	0.1	0.2	0.1	5	15	25	0.34978	0.02639	0.00161	1.74892	0.39580	0.04036
0.5	2	1	0.1	0.2	0.3	5	15	25	0.34978	0.02639	0.00484	1.74892	0.39580	0.12088
0.5	2	1	0.1	0.2	0.5	5	15	25	0.34978	0.02639	0.00805	1.74892	0.39580	0.20114
0.5	2	1	0.1	0.2	0.7	5	15	25	0.34978	0.02639	0.01125	1.74892	0.39580	0.28114
0.5	2	1	0.1	0.2	0.9	5	15	25	0.34978	0.02639	0.01444	1.74892	0.39580	0.36089
0.5	2	1	0.1	0.2	0.1	5	15	25	0.34978	0.02639	0.00161	1.74892	0.39580	0.04036
0.5	2	1	0.1	0.2	0.1	6	15	25	0.31180	0.02705	0.00165	1.87080	0.40582	0.04114
0.5	2	1	0.1	0.2	0.1	7	15	25	0.28016	0.02752	0.00167	1.96110	0.41273	0.04168
0.5	2	1	0.1	0.2	0.1	8	15	25	0.25367	0.02784	0.00168	2.02933	0.41756	0.04208
0.5	2	1	0.1	0.2	0.1	9	15	25	0.23134	0.02807	0.00169	2.08202	0.42100	0.04236
0.5	2	1	0.1	0.2	0.1	5	15	25	0.34978	0.02639	0.00161	1.74892	0.39580	0.04036
0.5	2	1	0.1	0.2	0.1	5	16	25	0.34978	0.02483	0.00162	1.74892	0.39721	0.04045
0.5	2	1	0.1	0.2	0.1	5	17	25	0.34978	0.02344	0.00162	1.74892	0.39842	0.04052
0.5	2	1	0.1	0.2	0.1	5	18	25	0.34978	0.02219	0.00162	1.74892	0.39947	0.04059
0.5	2	1	0.1	0.2	0.1	5	19	25	0.34978	0.02107	0.00163	1.74892	0.40040	0.04065
0.5	2	1	0.1	0.2	0.1	5	15	25	0.34978	0.02639	0.00161	1.74892	0.39580	0.04036
0.5	2	1	0.1	0.2	0.1	5	15	26	0.34978	0.02639	0.00155	1.74892	0.39580	0.04039
0.5	2	1	0.1	0.2	0.1	5	15	27	0.34978	0.02639	0.00150	1.74892	0.39580	0.04042
0.5	2	1	0.1	0.2	0.1	5	15	28	0.34978	0.02639	0.00144	1.74892	0.39580	0.04045
0.5	2	1	0.1	0.2	0.1	5	15	29	0.34978	0.02639	0.00140	1.74892	0.39580	0.04047

*= Seconds, $ = Multiplication of 10,000 Packets/sec



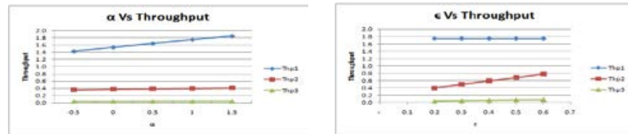


Figure 1 : The relationship between Utilization, Throughput and various other parameters

From equations (3.8), (3.10), (3.12), (3.23) and (3.18), (3.20), (3.22), the mean number of packets in the buffers and in the network, mean delay in transmission of three nodes are computed for different values of t, λ , α, θ, π, $\mu_{1,}$ $\mu_{2,}$ and μ_3 and presented in table 2. The relationship between the parameters and the performance measure are shown in the figure 2.

It is observed that as time (t) varies from 0.1 second to 5 seconds, the mean number of packets in the three buffers and in the network are increasing when other parameters are fixed. When the parameter (λ) varies from 3x10^4 packets/sec to 7x10^4 packets/sec, the mean numbers of packets in the first, second and third buffers and in the network are increasing when other parameters remain fixed. When the parameter (α) varies from -0.5 to 1.5, the mean number of packets in the first, second and third buffers and in the network are increasing when other parameters remain fixed. When the parameter (θ) varies from 0.1 to 0.9, the mean number of packets in the first buffer is constant, second and third buffers and in the network are increasing when other parameters remain fixed. When the parameter (ϵ) varies from 0.2 to 0.6, the mean number of packets in the first buffer is constant, second and third buffers and in the network are increasing when other parameters remain fixed. When the parameter (π) varies from 0.1 to 0.9, the mean number of packets in the first and second buffers are fixed and in the third buffer and in the network are increasing when other parameters remain fixed.

Table 2

Values of mean number of packets and mean delay of the communication network with DBA and NHP Arrivals

t^*	$\lambda^\$$	α	Θ	ϵ	π	$\mu_1^\$$	$\mu_2^\$$	$\mu_3^\$$	$L_1(t)$	$L_2(t)$	$L_3(t)$	$W_1(t)$	$W_2(t)$	$W_3(t)$	$L_n(t)$
0.1	2	1	0.1	0.2	0.1	5	15	25	0.16165	0.01310	0.00092	0.21660	0.06710	0.04002	0.17567
0.3	2	1	0.1	0.2	0.1	5	15	25	0.33967	0.02280	0.00143	0.23589	0.06743	0.04003	0.36390
0.5	2	1	0.1	0.2	0.1	5	15	25	0.43045	0.02674	0.00162	0.24612	0.06756	0.04003	0.45881
0.7	2	1	0.1	0.2	0.1	5	15	25	0.48913	0.02901	0.00173	0.25288	0.06764	0.04003	0.51988
0.9	2	1	0.1	0.2	0.1	5	15	25	0.53600	0.03069	0.00183	0.25837	0.06769	0.04004	0.56852
2.0	2	1	0.1	0.2	0.1	5	15	25	0.75998	0.03822	0.00228	0.28553	0.06795	0.04005	0.80048
5.0	2	1	0.1	0.2	0.1	5	15	25	1.36000	0.05822	0.00348	0.36592	0.06863	0.04007	1.42170
0.5	3	1	0.1	0.2	0.1	5	15	25	0.61403	0.03259	0.00202	0.26765	0.06776	0.04004	0.64864
0.5	4	1	0.1	0.2	0.1	5	15	25	0.79762	0.03844	0.00241	0.29025	0.06796	0.04005	0.83847
0.5	5	1	0.1	0.2	0.1	5	15	25	0.98120	0.04428	0.00281	0.31391	0.06815	0.04006	1.02830
0.5	6	1	0.1	0.2	0.1	5	15	25	1.16478	0.05013	0.00321	0.33859	0.06835	0.04006	1.21813
0.5	7	1	0.1	0.2	0.1	5	15	25	1.34836	0.05598	0.00361	0.36426	0.06855	0.04007	1.40796
0.5	2	-0.5	0.1	0.2	0.1	5	15	25	0.33552	0.02416	0.00150	0.23543	0.06748	0.04003	0.36119
0.5	2	0	0.1	0.2	0.1	5	15	25	0.36717	0.02502	0.00154	0.23896	0.06750	0.04003	0.39373
0.5	2	0.5	0.1	0.2	0.1	5	15	25	0.39881	0.02588	0.00158	0.24252	0.06753	0.04003	0.42627
0.5	2	1	0.1	0.2	0.1	5	15	25	0.43045	0.02674	0.00162	0.24612	0.06756	0.04003	0.45881
0.5	2	1.5	0.1	0.2	0.1	5	15	25	0.46209	0.02760	0.00165	0.24976	0.06759	0.04003	0.49135
0.5	2	1	0.1	0.2	0.1	5	15	25	0.43045	0.02674	0.00162	0.24612	0.06756	0.04003	0.45881
0.5	2	1	0.3	0.2	0.1	5	15	25	0.43045	0.05357	0.00325	0.24612	0.06847	0.04006	0.48727

0.5	2	1	**0.5**	0.2	0.1	5	15	25	0.43045	0.08040	0.00488	0.24612	0.06938	0.04010	0.51573
0.5	2	1	**0.7**	0.2	0.1	5	15	25	0.43045	0.10723	0.00651	0.24612	0.07030	0.04013	0.54419
0.5	2	1	**0.9**	0.2	0.1	5	15	25	0.43045	0.13406	0.00814	0.24612	0.07124	0.04016	0.57265
0.5	2	1	0.1	**0.2**	0.1	5	15	25	0.43045	0.02674	0.00162	0.24612	0.06756	0.04003	0.45881
0.5	2	1	0.1	**0.3**	0.1	5	15	25	0.43045	0.03340	0.00202	0.24612	0.06779	0.04004	0.46587
0.5	2	1	0.1	**0.4**	0.1	5	15	25	0.43045	0.04007	0.00242	0.24612	0.06801	0.04005	0.47293
0.5	2	1	0.1	**0.5**	0.1	5	15	25	0.43045	0.04673	0.00282	0.24612	0.06824	0.04006	0.47999
0.5	2	1	0.1	**0.6**	0.1	5	15	25	0.43045	0.05339	0.00322	0.24612	0.06846	0.04006	0.48706
0.5	2	1	0.1	0.2	**0.1**	5	15	25	0.43045	0.02674	0.00162	0.24612	0.06756	0.04003	0.45881
0.5	2	1	0.1	0.2	**0.3**	5	15	25	0.43045	0.02674	0.00485	0.24612	0.06756	0.04010	0.46204
0.5	2	1	0.1	0.2	**0.5**	5	15	25	0.43045	0.02674	0.00808	0.24612	0.06756	0.04016	0.46527
0.5	2	1	0.1	0.2	**0.7**	5	15	25	0.43045	0.02674	0.01131	0.24612	0.06756	0.04023	0.46850
0.5	2	1	0.1	0.2	**0.9**	5	15	25	0.43045	0.02674	0.01454	0.24612	0.06756	0.04029	0.47173
0.5	2	1	0.1	0.2	0.1	**5**	15	25	0.43045	0.02674	0.00162	0.24612	0.06756	0.04003	0.45881
0.5	2	1	0.1	0.2	0.1	**6**	15	25	0.37368	0.02743	0.00165	0.19974	0.06759	0.04003	0.40275
0.5	2	1	0.1	0.2	0.1	**7**	15	25	0.32872	0.02790	0.00167	0.16762	0.06760	0.04003	0.35829
0.5	2	1	0.1	0.2	0.1	**8**	15	25	0.29258	0.02823	0.00168	0.14418	0.06761	0.04003	0.32250
0.5	2	1	0.1	0.2	0.1	**9**	15	25	0.26310	0.02847	0.00170	0.12637	0.06762	0.04003	0.29326
0.5	2	1	0.1	0.2	0.1	5	**15**	25	0.43045	0.02674	0.00162	0.24612	0.06756	0.04003	0.45881
0.5	2	1	0.1	0.2	0.1	5	**16**	25	0.43045	0.02514	0.00162	0.24612	0.06329	0.04003	0.45721
0.5	2	1	0.1	0.2	0.1	5	**17**	25	0.43045	0.02372	0.00162	0.24612	0.05952	0.04003	0.45579
0.5	2	1	0.1	0.2	0.1	5	**18**	25	0.43045	0.02244	0.00162	0.24612	0.05618	0.04003	0.45452
0.5	2	1	0.1	0.2	0.1	5	**19**	25	0.43045	0.02130	0.00163	0.24612	0.05319	0.04003	0.45338
0.5	2	1	0.1	0.2	0.1	5	15	**25**	0.43045	0.02674	0.00162	0.24612	0.06756	0.04003	0.45881
0.5	2	1	0.1	0.2	0.1	5	15	**26**	0.43045	0.02674	0.00155	0.24612	0.06756	0.03849	0.45874
0.5	2	1	0.1	0.2	0.1	5	15	**27**	0.43045	0.02674	0.00150	0.24612	0.06756	0.03706	0.45869
0.5	2	1	0.1	0.2	0.1	5	15	**28**	0.43045	0.02674	0.00145	0.24612	0.06756	0.03574	0.45864
0.5	2	1	0.1	0.2	0.1	5	15	**29**	0.43045	0.02674	0.00140	0.24612	0.06756	0.03451	0.45859

* = Seconds, $ = Multiplication of 10,000 Packets/sec

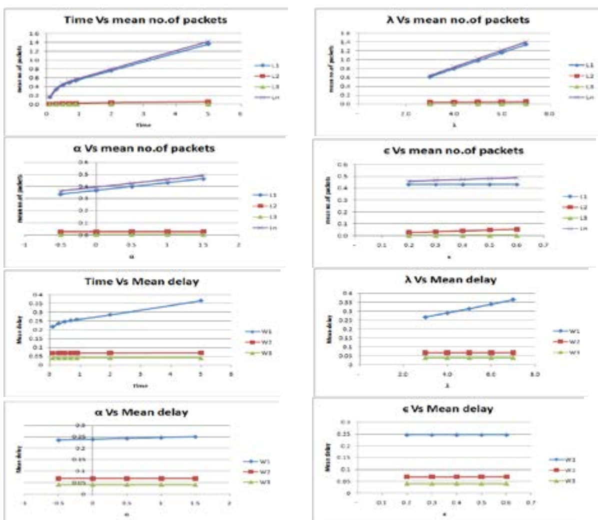


Figure 2 : The relationship between mean no. of packets, mean delay and various parameters

When the transmission rate (μ_1) varies from $5x10^4$ packets/sec to $9x10^4$ packets/sec, the mean number of packets in the first buffer and in the network are decreasing and in the second and third buffers are increasing when other parameters remain fixed. Similarly the transmission rate (μ_2) varies from $15x10^4$ packets/sec to $19x10^4$ packets/sec, the mean number of packets in the first buffer remains fixed and the mean number of packets in the second buffer and in the network are decreasing and the mean number of packets in the third buffer is increases, when other parameters remain fixed. Similarly the transmission rate (μ_3) varies from $25x10^4$ packets/sec to $29x10^4$ packets/sec, the mean number of packets in the first and second buffers remains constant and the mean number of packets in the third buffer and in the network are decreasing when other parameters remain fixed.

It is observed that as time (t) and the parameter (λ) are increasing, the mean delay in buffers are increasing for fixed values of the other parameters. It is also observed that as parameter (α) varies the mean delay in buffers are increasing for fixed values of the other parameters. The parameter (θ) varies the mean delay in first buffer remains constant and second and third buffers are increasing for fixed values of the other parameters. The parameter (ϵ) varies the mean delay in first node is constant and second and third nodes are increasing for fixed values of other parameters. The parameter (π) varies the mean delay in first and second buffers remain fixed and third buffer is increases for fixed values of the other parameters. When the transmission rate (μ_1) increases, the mean delay in the first buffer decreases and in the second and third buffers increasing when the other parameters remain fixed. Similarly, the transmission rate (μ_2) increases the mean delay in the first buffer remains fixed, second buffer decreases, and third buffer increases when other parameter remains fixed. Similarly, the transmission rate (μ_3) increases the mean delay in the first and second buffers are fixed and the third buffer are decreasing, when other parameter remain fixed.

From this analysis it is observed that the dynamic bandwidth allocation strategy has a significant influence on all performance measures of the network. It is further observed that the performance measures are highly sensitive towards smaller values of time. Hence, it is optimal to consider dynamic bandwidth allocation under and non-homogeneous Poisson arrivals and evaluate the performance of the network under transient conditions. It is also to be observed that the congestion in buffers and delays in transmission got reduced to a minimum level by adopting dynamic bandwidth allocation.

5 SENSIVITY ANALYSIS:

Sensitivity analysis of the model is performed to steady the effect of changes in the parameters t, λ, α, θ, ϵ, π, $\mu_{1,}$ $\mu_{2,}$ μ_3 on the mean number of packets, the utilization of transmitters, the mean delay in transmitters and the throughput of the first, second and third nodes.

The following data has been considered for the sensitivity analysis, t = 0.5 sec, $\lambda = 2x10^4$ packets/sec, α=1, θ = 0.1, ϵ =0.2, π = 0.1, μ_1 = 5x 10^4 packets/sec, μ_2 = 15x 10^4 packets/sec, and μ_3 =$25x10^4$ packets/sec.

The mean number of packets, the utilization of nodes, the mean delay, and the throughput of the first, second and third buffers are computed with variation of -15%,-10%, -5%, 0%, +5%, +10% and +15% on the model parameters and presented in table 3.

Table 3: Sensitivity Analysis of the model

Parameter	Performance Measures	% change in Parameter						
		-15	-10	-5	0	5	10	15
	$L_1(t)$	0.40200	0.41206	0.42151	0.43045	0.43892	0.44699	0.45469
	$L_2(t)$	0.01891	0.01932	0.01971	0.02008	0.02042	0.02074	0.02104
	$L_3(t)$	0.00231	0.00236	0.00239	0.00243	0.00247	0.00250	0.00253
	$U_1(t)$	0.33102	0.33771	0.34395	0.34978	0.35527	0.36045	0.36536
	$U_2(t)$	0.01873	0.01914	0.01952	0.01988	0.02021	0.02052	0.02082

t=0.5	$U_3(t)$	0.00231	0.00235	0.00239	0.00243	0.00246	0.00250	0.00253
	$W_1(t)$	0.24289	0.24403	0.24510	0.24612	0.24709	0.24802	0.24890
	$W_2(t)$	0.06730	0.06731	0.06733	0.06734	0.06735	0.06736	0.06737
	$W_3(t)$	0.04005	0.04005	0.04005	0.04005	0.04005	0.04005	0.04005
	Thp_1	1.65511	1.68856	1.71974	1.74892	1.77634	1.80224	1.82678
	Thp_2	0.28092	0.28709	0.29282	0.29817	0.30317	0.30787	0.31230
	Thp_3	0.05780	0.05883	0.05980	0.06071	0.06157	0.06239	0.06317
λ=2	$L_1(t)$	0.37537	0.39373	0.41209	0.43045	0.44881	0.46717	0.48552
	$L_2(t)$	0.01832	0.01891	0.01949	0.02008	0.02066	0.02125	0.02183
	$L_3(t)$	0.00219	0.00227	0.00235	0.00243	0.00251	0.00259	0.00267
	$U_1(t)$	0.31297	0.32547	0.33774	0.34978	0.36161	0.37322	0.38463
	$U_2(t)$	0.01816	0.01873	0.01930	0.01988	0.02045	0.02102	0.02160
	$U_3(t)$	0.00219	0.00227	0.00235	0.00243	0.00251	0.00259	0.00267
	$W_1(t)$	0.23988	0.24195	0.24403	0.24612	0.24823	0.25034	0.25247
	$W_2(t)$	0.06728	0.06730	0.06732	0.06734	0.06736	0.06738	0.06740
	$W_3(t)$	0.04004	0.04005	0.04005	0.04005	0.04005	0.04005	0.04005
	Thp_1	1.56484	1.62733	1.68868	1.74892	1.80806	1.86612	1.92313
	Thp_2	0.27235	0.28096	0.28957	0.29817	0.30676	0.31535	0.32393
	Thp_3	0.05473	0.05672	0.05872	0.06071	0.06270	0.06469	0.06668
α=1	$L_1(t)$	0.42096	0.42412	0.42729	0.43045	0.43361	0.43678	0.43994
	$L_2(t)$	0.01982	0.01991	0.01999	0.02008	0.02016	0.02025	0.02034
	$L_3(t)$	0.00241	0.00242	0.00242	0.00243	0.00244	0.00245	0.00245
	$U_1(t)$	0.34358	0.34566	0.34772	0.34978	0.35184	0.35388	0.35593
	$U_2(t)$	0.01962	0.01971	0.01979	0.01988	0.01996	0.02005	0.02013
	$U_3(t)$	0.00241	0.00241	0.00242	0.00243	0.00244	0.00244	0.00245
	$W_1(t)$	0.24504	0.24540	0.24576	0.24612	0.24649	0.24685	0.24721
	$W_2(t)$	0.06733	0.06733	0.06734	0.06734	0.06734	0.06734	0.06735
	$W_3(t)$	0.04005	0.04005	0.04005	0.04005	0.04005	0.04005	0.04005
	Thp_1	1.71791	1.72828	1.73861	1.74892	1.75919	1.76942	1.77963
	Thp_2	0.29437	0.29564	0.29690	0.29817	0.29943	0.30069	0.30196
	Thp_3	0.06013	0.06032	0.06051	0.06071	0.06090	0.06109	0.06129
θ=0.1	$L_1(t)$	0.43045	0.43045	0.43045	0.43045	0.43045	0.43045	0.43045
	$L_2(t)$	0.01807	0.01874	0.01941	0.02008	0.02075	0.02142	0.02209
	$L_3(t)$	0.00219	0.00227	0.00235	0.00243	0.00251	0.00259	0.00268
	$U_1(t)$	0.34978	0.34978	0.34978	0.34978	0.34978	0.34978	0.34978
	$U_2(t)$	0.01790	0.01856	0.01922	0.01988	0.02053	0.02119	0.02185
	$U_3(t)$	0.00218	0.00227	0.00235	0.00243	0.00251	0.00259	0.00267
	$W_1(t)$	0.24612	0.24612	0.24612	0.24612	0.24612	0.24612	0.24612
	$W_2(t)$	0.06727	0.06729	0.06732	0.06734	0.06736	0.06738	0.06741
	$W_3(t)$	0.04004	0.04005	0.04005	0.04005	0.04005	0.04005	0.04005
	Thp_1	1.74892	1.74892	1.74892	1.74892	1.74892	1.74892	1.74892
	Thp_2	0.26855	0.27843	0.28830	0.29817	0.30802	0.31787	0.32772
	Thp_3	0.05460	0.05664	0.05867	0.06071	0.06274	0.06478	0.06681

The performance measures are highly affected by the changes in the values of time (t) and parameter (λ), (α), (θ), (ϵ) and (π). As t increases from -15% to +15% the average number of packets in the three buffers and in the network are increasing along with an increase in the mean delay in buffers, the utilization of transmitters and throughput of the three nodes. As the parameter (λ) increases from -15% to +15% the average number of packets in the three buffers and in the network are increasing along with an increase in the mean delay, the utilization of transmitters and the throughput of the three nodes. Similarly, for the parameter (α), the utilization of transmitters and the throughput of nodes are increasing in the communication network. Overall analysis of the parameters reflects that the dynamic bandwidth allocation strategy for congestion control will tremendously reduce the delay in communication and improve the voice quality by reducing burstness in buffers.

6 COMPARATIVE STUDY

To study the effect of non homogeneous Poisson arrival assumption on the communication network a comparative study between the performance measures of the network models with non homogeneous Poisson arrivals and Poisson arrivals is performed. The performance measures of both models are computed with fixed values of the parameters (λ , α, θ, ϵ, π, μ_1, μ_2, μ_3) and different values of t = 0.3, 0.5, 2, 5 seconds are presented in table 4.

Table 4
Comparative study of models with Non-Homogeneous and Homogeneous Poisson arrivals

Time (t) Sec	Parameters Measured	Model with NHP α=1	Model With HP α=0	Difference	% Variation
t=0.3	$L_1(t)$	0.3396	0.3107	0.0289	4.45
	$L_2(t)$	0.0227	0.0221	0.0006	1.48
	$L_3(t)$	0.0014	0.0014	0.0000	0.01
	$U_1(t)$	0.2879	0.2671	0.0208	3.76
	$U_2(t)$	0.0225	0.0218	0.0006	1.47
	$U_3(t)$	0.0014	0.0014	0.0001	0.01
	$W_1(t)$	0.2358	0.2326	0.0032	0.68
	$W_2(t)$	0.0674	0.0674	0.0001	0.02
	$W_3(t)$	0.0400	0.0399	0.0001	0.03
	Thp_1	1.4399	1.3355	0.1044	3.76
	Thp_2	0.3380	0.3283	0.0097	1.47
	Thp_3	0.0356	0.0356	0.0001	0.01
t=5	$L_1(t)$	1.3600	0.4000	0.9600	54.55
	$L_2(t)$	0.0516	0.0200	0.0316	44.10
	$L_3(t)$	0.0062	0.0024	0.0038	43.89
	$U_1(t)$	0.7433	0.3297	0.4137	38.55
	$U_2(t)$	0.0502	0.0198	0.0304	43.47
	$U_3(t)$	0.0061	0.0024	0.0037	43.81
	$W_1(t)$	0.3659	0.2427	0.1233	20.25
	$W_2(t)$	0.0684	0.0673	0.0011	0.78
	$W_3(t)$	0.0401	0.0400	0.0001	0.09
	Thp_1	3.7167	1.6484	2.0683	38.55
	Thp_2	0.7537	0.2970	0.4567	43.47
	Thp_3	0.1534	0.0599	0.0935	43.81

As t increases the percentage variation of performance measures between the two models is increasing. For the model with non-homogeneous Poisson arrivals with dynamic bandwidth allocation has more utilization compared to that of the model with Poisson arrivals with dynamic bandwidth allocation. From this analysis it is observed that the assumption of non-homogeneous Poisson arrivals have a significant influence on all the performance measures of the network. This model also includes the two node tandem communication network model when μ_3 is zero.

7. CONCLUSIONS:

In this paper a novel and new communication model which is much useful for analyzing the communication systems more effectively and efficiently is developed and analyzed. The work presented in this paper focus on the improvement of three node tandem communication network with Dynamic Bandwidth Allocation and modified phase type transmission having NHP arrivals for first node and Poisson arrivals for second node. The variable traffic conditions (bursty traffic/ time dependent traffic) is characterized by non homogeneous Poisson process with time dependent arrival rates. It is shown that the dynamic bandwidth allocation can reduce congestion in buffers and delay in transmission by utilizing iid bandwidth. The developed network model much useful for the evaluating the performance of several networks like LAN, WAN, MAN and Computer Communication Systems under variable traffic conditions by predicting performance measures more close to the reality. This mode also includes some of the earlier models as particular cases for specific values of the parameters.

8. REFERENCES:

1] Feldmann A, (2000), Characteristics of TCP connection arrivals, Chapter 15, self similar network traffic and performance evaluation, edited by K.Park, W.willinger, John Wiley & Sons Inc.

2] Gaujal, B. and Hyon, E.(2002), Optimal routing policies in deterministic queues in tandem, Proceedings of Sixth International Workshop on Discrete Event Systems (WODES'02),pp 251-257.

3] IEEE Communications Society(2002), A Brief History of Communications, Report by IEEE Communications Society.

4] Nageswara Rao K., Srinivas Rao K, Srinivasa Rao P (2010) "A tandem communication Network with Dynamic Bandwidth Allocation and modified phase type transmission having Bulk arrivals" in International Journal of Computer science Issues, vol.7, no.3, pp 18-26.

5] Nageswara Rao et al (2011) Performance Evaluation of Two node Tandem Communication Network with Dynamic Bandwidth Allocation having Two Stage Direct Bulk Arrivals, Vol.8, Issue 1, pp.122-130

6] Padmavathi G., K. Srinivas Rao, KVVS Reddy (2009), "Performance Evaluation of parallel and series communication Network with Dynamic Bandwidth Allocation", CIIT International Journal of Networking and communication Engineering Vol.1, No. 7, pp.410-421.

7] Parthasarathy, P.R. and Selvraju, N. (2001), Transient analysis of a Queue where potential customers are Discouraged by Queue length. Mathematical Problems in Engineering, Vol.7, pp.433-454

8] Rakesh Singhai, Shiv Dutt Joshi and Rajendra K P Bhatt (2007), A Novel Discrete distribution and process to model self similar traffic, 9th IEEE international conference on Telecommunication-CONTEL 2007, Zagredb Croatia, pp.167-171

9] Srinivasa Rao, K., Vasanta, M.R., and Vijaya Kumar, C.V.R.S.,(2000), On an interdependent Communication Network, Opsearch Vol.37, No.2, pp 134-143.

10] Sriram, K. (1993), Methodologies for bandwidth allocation, transmission scheduling and congestion avoidance in broadband ATM networks, Computer Network, ISDN System, J.26, pp 43-59.

11] Suresh Varma, P. and Srinivasa Rao, K. (2007), A Communication Network with load dependent Transmission, to appear in International Journal of Mathematical Sciences.Vol.6, No.2, pp.199-210

Mrs.M.V.Rama Sundari is working as Associate Professor, Department of IT, Aditya Engineering College, Surampalem, East Godavari, Andhra Pradesh.

Dr.K.Srinivas Rao is presently working as Professor, Department of Statistics, Andhra University, Visakhapatnam. He is elected as chief editor of Journal of ISPS and elected as Vice-President of Operation Research of India during 2009-10. He guided 28 students for Ph.D. in Statistics, Computer Science, Electronics and Communications and Operations Research. He published 92 research papers in national and international journals with high reputation. His research interests are communication systems, Data Mining and stochastic models

Dr.Peri Srinivasa Rao is presently working as Professor in the Department of Computer Science and Systems Engineering, Andhra University, Visakhapatnam. He got his Ph.D. degree from Indian Institute of Technology, Kharagpur in Computer Science in 1987. He published several research papers and delivered invited lectures at various conferences, seminars and workshops. He guided a large number of students for their M.Tech degrees in Computer Science and Engineering and Information Technology. His current research interests are Communication Networks, Data Mining and Computer Morphology

P.Suresh Varma is presently working as Professor in the Department of Computer Science, Adikavi Nannaya University, Rajahmundry. He published several research papers in national and international journals and presented papers at various conferences, seminars and workshops. He guided a large number of students for their M.Tech, MCA degrees in Computer Science and Engineering and Information Technology. His current research interests are Communication Networks, Data Mining and Cloud Computing.

A Graphical Password Based System for Small Mobile Devices

Wazir Zada Khan[1], Mohammed Y Aalsalem[2] and Yang Xiang[3]

[1] **School of Computer Science, University of Jazan**
Jazan, PoBox # 114, Kingdom of Saudi Arabia

[2] **School of Computer Science, University of Jazan**
Jazan, PoBox # 114, Kingdom of Saudi Arabia

[3] **School of Information Technology, Deakin University, Australia**
221 Burwood Highway, Burwood, VIC 3125, Australia

Abstract

Passwords provide security mechanism for authentication and protection services against unwanted access to resources. A graphical based password is one promising alternatives of textual passwords. According to human psychology, humans are able to remember pictures easily. In this paper, we have proposed a new hybrid graphical password based system, which is a combination of recognition and recall based techniques that offers many advantages over the existing systems and may be more convenient for the user. Our scheme is resistant to shoulder surfing attack and many other attacks on graphical passwords. This scheme is proposed for smart mobile devices (like smart phones i.e. ipod, iphone, PDAs etc) which are more handy and convenient to use than traditional desktop computer systems.

Keywords: ***Smart Phones,*** *Graphical Passwords, Authentication, Network Security.*

1. Introduction

One of the major functions of any security system is the control of people in or out of protected areas, such as physical buildings, information systems, and our national borders. Computer systems and the information they store and process are valuable resources which need to be protected. Computer security systems must also consider the human factors such as ease of a use and accessibility. Current secure systems suffer because they mostly ignore the importance of human factors in security [1]. An ideal security system considers security, reliability, usability, and human factors. All current security systems have flaws which make them specific for well trained and skilled users only. A password is a secret that is shared by the verifier and the customer. "Passwords are simply secrets that are provided by the user upon request by a recipient." They are often stored on a server in an encrypted form so that a penetration of the file system does not reveal password lists [2]. Passwords are the most common means of authentication because they do not require any special hardware. Typically passwords are strings of letters and digits, i.e. they are alphanumeric. Such passwords have the disadvantage of being hard to remember [3]. Weak passwords are vulnerable to dictionary attacks and brute force attacks where as Strong passwords are harder to remember. To overcome the problems associated with password based authentication systems, the researchers have proposed the concept of graphical passwords and developed the alternative authentication mechanisms. Graphical passwords systems are the most promising alternative to conventional password based authentication systems.

Graphical passwords (GP) use pictures instead of textual passwords and are partially motivated by the fact that humans can remember pictures more easily than a string of characters [4]. The idea of graphical passwords was originally described by Greg Blonder in 1996 [62]. An important advantage of GP is that they are easier to remember than textual passwords. Human beings have the ability to remember faces of people, places they visit and things they have seen for a longer duration. Thus, graphical passwords provide a means for making more user-friendly passwords while increasing the level of security. Besides these advantages, the most common problem with graphical passwords is the shoulder surfing problem: an onlooker can steal user's graphical password by watching in the user's vicinity. Many researchers have attempted to solve this problem by providing different techniques [6]. Due to this problem, most graphical passwords schemes recommend small mobile devices (PDAs) as the ideal application environment. Another common problem with graphical passwords is that it takes longer to input graphical passwords than textual passwords [6]. The login process is slow and it may frustrate the impatient users. Graphical passwords serve the same purpose as textual passwords differing in consisting of handwritten designs (drawing), possibly in addition to text. The exploitation of smart phones like ipod and PDA's is increased due to their small size, compact deployment and low cost.

In this paper, considering the problems of text based password systems, we have proposed a new graphical password scheme which has desirable usability for small mobile device. Our proposed system is new graphical passwords based hybrid system which is a combination of recognition and recall based techniques and consists of two phases. During the first phase called Registration phase, the user has to first select his username and a textual password. Then objects are shown to the user to select from them as his graphical password. After selecting the user has to draw those selected objects on a touch sensitive screen using a stylus. During the second phase called Authentication phase, the user has to give his username and textual password and then give his graphical password by drawing it in the same way as done during the registration phase. If they are drawn correctly the user is authenticated and only then he/she can access his/her account. For practical implementation of our system we have chosen i-mate JAMin smart phone which is produced by HTC, the Palm Pilot, Apple Newton, Casio Cassiopeia E-20 and others which allow users to provide graphics input to the device. It has a display size of 240x320 pixels and an important feature of Handwriting recognition. The implementation details are out of the scope of this paper.

The structure of our paper is organized as follows. In Section II, the classification of all existing authentication methods is described. In Section III, all existing graphical password based schemes are classified into three main categories. Section IV reviews existing research and schemes which are strongly related to our work. Section V discusses the problems of all existing graphical password based schemes. In Section VI our proposed system is described in detail. In Section VII we have compared our proposed system with existing schemes by drawing out the flaws in existing schemes. Section VIII provides discussion. Finally Section IX concludes the paper.

2. Classification of Current Authentication Methods

Due to recent events of thefts and terrorism, authentication has become more important for an organization to provide an accurate and reliable means of authentication [14]. Currently the authentication methods can be broadly divided into three main areas. Token based (two factor), Biometric based (three factor), and Knowledge based (single factor) authentication [7], also shown in the Figure 1.

2.1 Token Based Authentication:

It is based on “Something You Possess”. For example Smart Cards, a driver’s license, credit card, a university ID card etc. It allows users to enter their username and password in order to obtain a token which allows them to fetch a specific resource - without using their username and password. Once their token has been obtained, the user can offer the token - which offers access to a specific resource for a time period - to the remote site [15]. Many token based authentication systems also use knowledge based techniques to enhance security [7].

2.2 Biometric Based Authentication:

Biometrics (ancient Greek: bios ="life", metron ="measure") is the study of automated methods for uniquely recognizing humans based upon one or more intrinsic physical or behavioral traits [9]. It is based on “Something You Are” [8]. It uses physiological or behavioral characteristics like fingerprint or facial scans and iris or voice recognition to identify users. A biometric scanning device takes a user's biometric data, such as an iris pattern or fingerprint scan, and converts it into digital information a computer can interpret and verify.

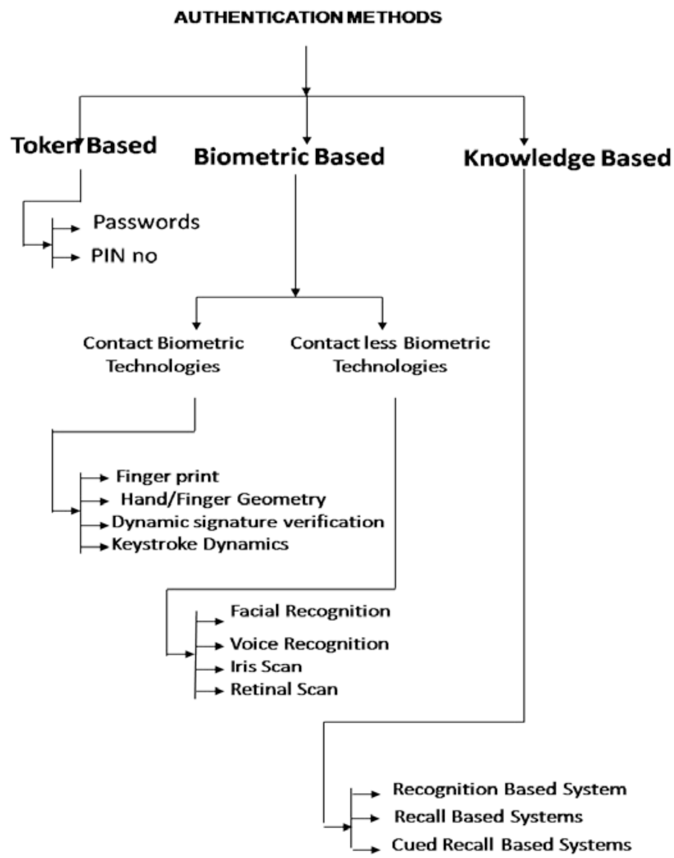


Fig. 1 Classification of Authentication Methods

A biometric-based authentication system may deploy one or more of the biometric technologies: voice recognition, fingerprints, face recognition, iris scan, infrared facial and hand vein thermo grams, retinal scan, hand and finger geometry, signature, gait, and keystroke dynamics [19]. Biometric identification depends on computer algorithms to make a yes/no decision. It enhances user service by providing quick and easy identification [20].

2.3 Knowledge Based Authentication:

Knowledge based techniques are the most extensively used authentication techniques and include both text based and picture based passwords [7]. Knowledge-based authentication (KBA) is based on "Something You Know" to identify you For Example a Personal Identification Number (PIN), password or pass phrase. It is an authentication scheme in which the user is asked to answer at least one "secret" question [17]. KBA is often used as a component in multifactor authentication (MFA) and for self-service password retrieval. Knowledge based authentication (KBA) offers several advantages to traditional (conventional) forms of e-authentication like passwords, PKI and biometrics [16].

3. Classification of Graphical Password Based Systems

Graphical based passwords schemes can be broadly classified into four main categories: First is **Recognition based Systems** which are also known as Cognometric Systems or Searchmetric Systems. Recognition based techniques involve identifying whether one has seen an image before. The user must only be able to recognize previously seen images, not generate them unaided from memory. Second is **Pure Reacll based systems** which are also known as Drwanmetric Systems. In pure recall-based methods the user has to reproduce something that he or she created or selected earlier during the registration stage. Third is **Cued Recall based** systems which are also called Iconmetric Systems. In cued recall-based methods, a user is provided with a hint so that he or she can recall his his/her password. Fourth is **Hybrid systems** which are typically the combination of two or more schemes. Like recognition and recall based or textual with graphical password schemes. Detailed classification of systems, involved in these four categories is shown in Figure 2.

4. Related Work

Haichang Gao et al. [55] have proposed and evaluated a new shoulder-surfing resistant scheme called Come from DAS and Story (CDS) which has a desirable usability for PDAs. This scheme adopts a similar drawing input method in DAS and inherits the association mnemonics in Story for sequence retrieval. It requires users to draw a curve across their password images (pass-images) orderly rather than click directly on them. The drawing method seems to be more compatible with people's writing habit, which may shorten the login time. The drawing input trick along with the complementary measures, such as erasing the drawing trace, displaying degraded images, and starting and ending with randomly designated images provide a good resistance to shoulder surfing. A user study is conducted to explore the usability of CDS in terms of accuracy, efficiency and memorability, and benchmark the usability against that of a Story scheme. The main contribution is that it overcomes a drawback of recall-based systems by erasing the drawing trace and introduces the drawing method to a variant of Story to resist shoulder-surfing.

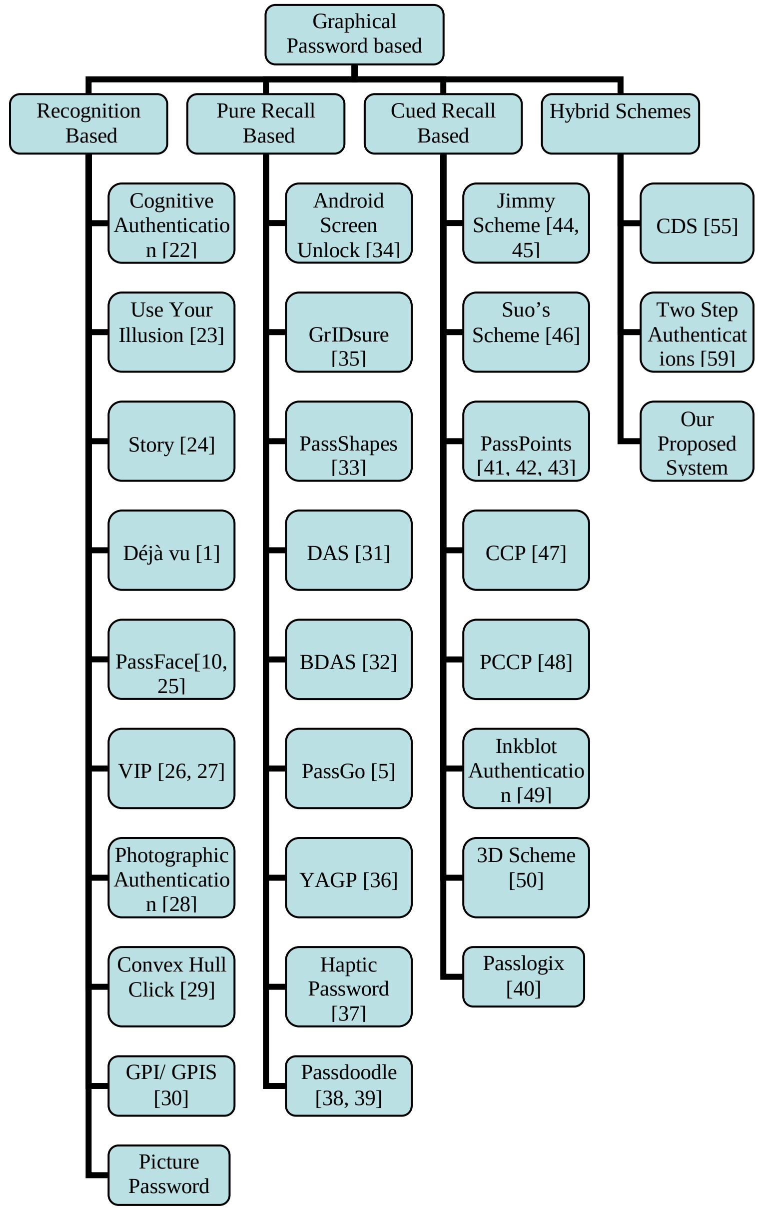


Fig. 2 Classification of Graphical Password Based Systems

P.C.van Oorshot and Tao Wan [59] have proposed a hybrid authentication approach called *Two-Step*. In this scheme users continue to use text passwords as a first step but then must also enter a graphical password. In step one, a user is asked for her user name and text password. After supplying this, and independent of whether or not it is correct, in step two, the user is presented with an image portfolio. The user must correctly select all images (one or more) pre-registered for this account in each round of graphical password verification. Otherwise, account access is denied despite a valid text password. Using text passwords in step one preserves the existing user sign-in experience. If the user's text password or graphical password is correct, the image portfolios presented are those as defined during password creation. Otherwise, the image portfolios (including their layout dimensions) presented in first and a next round are random but respectively a deterministic function of the user name and text password string entered, and the images selected in the previous round.

5. Problem Domain

There are many problems with each of the graphical based authentication methods. These are discussed below:

5.1 Problems of Recognition Based Methods:

Dhamijia and Perrig proposed a graphical password based scheme Déjà Vu, based on Hash Visualization technique [11]. The drawback of this scheme is that the server needs to store a large amount of pictures which may have to be transferred over the network, delaying the authentication process. Another weakness of this system is that the server needs to store the seeds of portfolio images of each user in plaintext. Also, the process of selecting a set of pictures from picture database can be tedious and time consuming for the user [7]. This scheme was not really secure because the passwords need to store in database and that is easy to see.

Sobrado and Birget developed a graphical password technique that deals with the shoulder surfing problem [3]. In their first scheme the system displays a number of pass-objects (pre-selected by user) among many other objects as shown in Fig: 3. To be authenticated, a user needs to recognize pass-objects and click inside convex hull formed by all the pass objects. They developed many schemes to solve the shoulder surfing problem but the main drawback of these schemes is that log in process can be slow.

Another recognition based technique is proposed by Man et al [63]. He proposed a shoulder-surfing resistant algorithm which is similar to that developed by Sobrado and Birget. The difference is that Man et al has introduced several variants for each pass-object and each variant is assigned a unique code. Thus during authentication the user recognize pre-selected objects with an alphanumeric code and a string for each pass-object. Although it is very hard to break this kind of password but this method still requires the user to memorize alphanumeric codes for each pass-object variants.

"Passface" is another recognition based system. It is argued by its developer that it is easy for human beings to remember human faces than any other kind of passwords. But Davis et al [12] have found that most users tend to choose faces of people from the same race. This makes the Passface password somewhat predictable. Furthermore, some faces might not be welcomed by certain users and thus the login process will be unpleasant. Another limitation of this system is that it cannot be used by those people who are face-blind [6].

5.2 Problems of Recall Based Methods:

The problem with the Grid based methods is that during authentication the user must draw his/her password in the same grids and in the same sequence. It is really hard to remember the exact coordinates of the grid. The problem with Passlogix is that the full password space is small. In addition a user chosen password might be easily guessable [6]. DAS scheme has some limitations like it is vulnerable to shoulder surfing attack if a user accesses the system in public environments, there is still a risk for the attackers to gain access to the device if the attackers obtained a copy of the stored secret, and, brute force attacks can be launched by trying all possible combinations of grid coordinates,) Drawing a diagonal line and identifying a starting point from any oval shape figure using the DAS scheme itself can be a challenge for the users, and finally Difficulties might arise when the user chooses a drawing which contains strokes that pass too close to a grid-line, thus, the scheme may not be able to distinguish which cell the user is choosing.

"PassPoints" is the extended version of Blonder's idea by eliminating the predefined boundaries and allowing arbitrary images to be used. Using this scheme it takes time to think to locate the correct click region and determine precisely where to click. Another problem with these schemes is that it is difficult to input a password through a keyboard, the most common input device; if the mouse doesn't function well or a light pen is not available, the system cannot work properly [6]. Overall, with both "PassPoints" and "Passlogix", looking for small spots in a rich picture might be tiresome and unpleasant for users with weak vision.

In Viskey's scheme the main drawback is the input tolerance. Pointing to the exact spots on the picture has proven to be quite hard thus Viskey accepts all input within a certain tolerance area around it. It also allows users to set the size of this area in advance. However, some caution related to the input precision needs to be taken, since it will directly influence the security and the usability of the password. In order to practically set parameters, a four spot VisKey theoretically provides approximately 1 billion possibilities for defining a password. Unfortunately this is not large enough to prevent off-line attacks from a high-speed computer. Therefore no less than seven defined spots are required to overcome the likelihood of brute force attacks.

6. Proposed System

Taking into account all the problems and limitations of graphical based schemes, we have proposed a hybrid system for authentication. This hybrid system is a mixture of both recognition and recall based schemes.

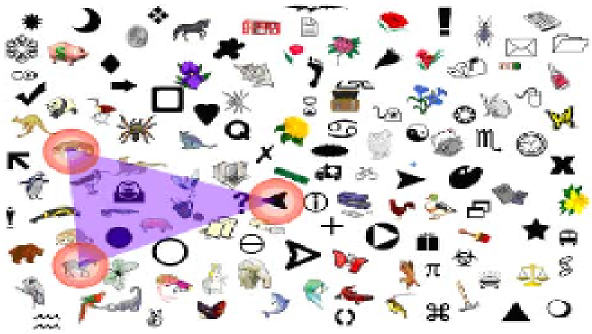


Fig. 3 A shoulder surfing resistant graphical password scheme [3].

Our proposed system is an approach towards more reliable, secure, user-friendly, and robust authentication. We have also reduced the shoulder surfing problem to some extent.

6.1 Working of Proposed System:

Our proposed system comprises of 9 steps out of which steps 1-3 are registration steps and steps 4-9 are the authentication steps.

Step 1

The first step is to type the user name and a textual password which is stored in the database. During authentication the user has to give that specific user name and textual password in order to log in.

Step 2

In this second step objects are displayed to the user and he/she selects minimum of three objects from the set and there is no limit for maximum number of objects. This is done by using one of the recognition based schemes. The selected objects are then drawn by the user, which are stored in the database with the specific username. Objects may be symbols, characters, auto shapes, simple daily seen objects etc. Examples are shown in Figure 4.

Step 3

During authentication, the user draws pre-selected objects as his password on a touch sensitive screen (or according to the environment) with a mouse or a stylus. This will be done using the pure recall based methods.

Step 4

In this step, the system performs pre-processing

Step 5

In the fifth step, the system gets the input from the user and merges the strokes in the user drawn sketch.

Step 6

After stroke merging, the system constructs the hierarchy.

Step 7

Seventh step is the sketch simplification.

Step 8

In the eighth step three types of features are extracted from the sketch drawn by the user.

Step 9

The last step is called hierarchical matching.

Graphical Representation of our proposed system is shown in Figure 5.

Fig. 4 Some examples of objects shown to the user

During registration, the user selects the user name and a textual password in a conventional manner and then chooses the objects as password. The minimum length for textual password is L=6. Textual password can be a mixture of digits, lowercase and uppercase letter. After this the system shows objects on the screen of a PDA to select as a graphical password. After choosing the objects, the user draws those objects on a screen with a stylus or a mouse. Objects drawn by the user are stored in the database with his/her username. In object selection, each object can be selected any number of times. Flow chart of registration phase is shown in Figure 6.

During authentication, the user has to first give his username and textual password and then draw pre-selected objects. These objects are then matched with the templates of all the objects stored in the database. Flow chart of authentication phase is shown in Figure 7. The phases during the authentication like the pre-processing, stroke merging, hierarchy construction, sketch simplification, feature extraction, and hierarchical matching are the steps proposed by Wing Ho Leung and Tsuhan Chen in their paper [13]. They propose a novel method for the retrieval of hand drawn sketches from the database, finally ranking the best matches. In the proposed system, the user will be authenticated only if the drawn sketch is fully matched with the selected object's template stored in the database. Pre-processing of hand drawn sketches is done prior to recognition and normally involves noise reduction and normalization. The noise occur in the image by user is generally due to the limited accuracy of human drawn images. [14]. A number of techniques can be used to reduce noise that includes Smoothing, filtering, wild point correction etc. Here in the proposed system Gaussian smoothing is used which eliminates noise introduced by the tablet or shaky drawing.

$$G(r) = \frac{1}{\sqrt{2\pi\sigma^2}^N} e^{-r^2/(2\sigma^2)}$$

or specifically in two dimensions

$$G(u,v) = \frac{1}{2\pi\sigma^2} e^{-(u^2+v^2)/(2\sigma^2)}$$

Where r is the blur radius (r2 = u2 + v2), and σ is the standard deviation of the Gaussian distribution.

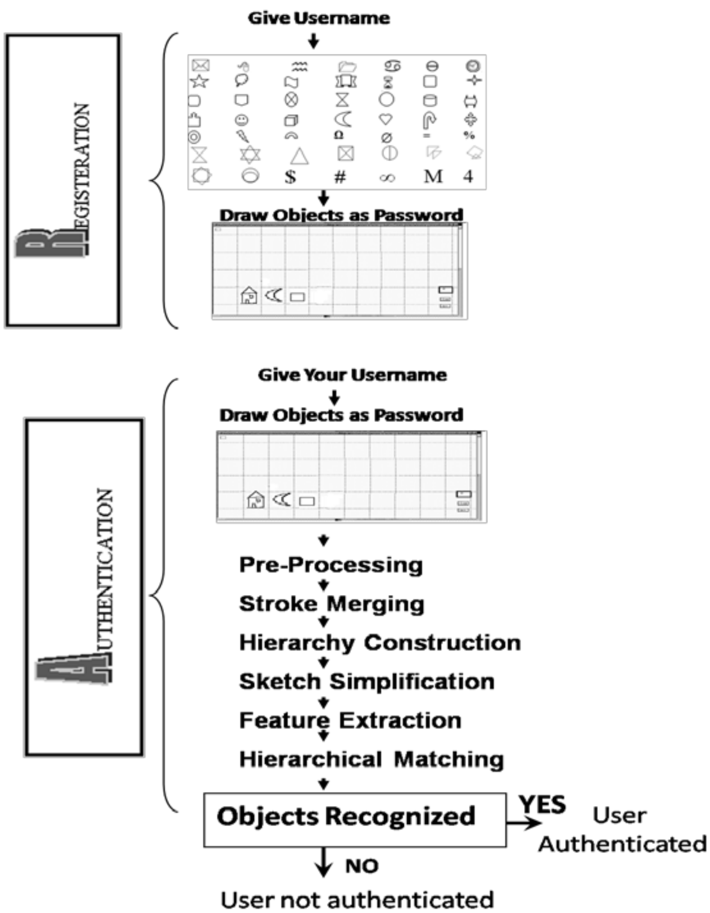


Fig. 5 Graphical Representation of Proposed System

In case, if user draws very large or a very small sketch then the system performs size normalization which adjusts the symbols or sketches to a standard size. The Stroke merging phase is use to merge the strokes which are broken at end points. If the end points are not close, then that stroke is considered as open stroke and it may be merged with another open stroke if the end point of one stroke is close to the end point of the other. The strokes are then represented in a hierarchy to simplify the image and to make it meaningful for further phases [13]. In the next step of sketch simplification, a shaded region is represented by a single hyper-stroke. After sketch simplification three types of features are extracted from the user re-drawn sketch. These features are hyper stroke features, Stroke features, and bi-stroke features.

In the last step of hierarchical matching, the similarity is evaluation the top to bottom hierarchical manner. The user is allowed to draw in an unrestricted manner. The overall process is difficult because free hand sketching is a difficult job. The order in which the user has selected the objects does matter in our proposed system i.e. during the authentication phase, the user can draw his pre-selected objects in the same order as he had selected during the registration phase. So, in this way the total combinations of each password will be $2^n - 1$, 'n' being the number of objects selected by the user as password during the registration phase.

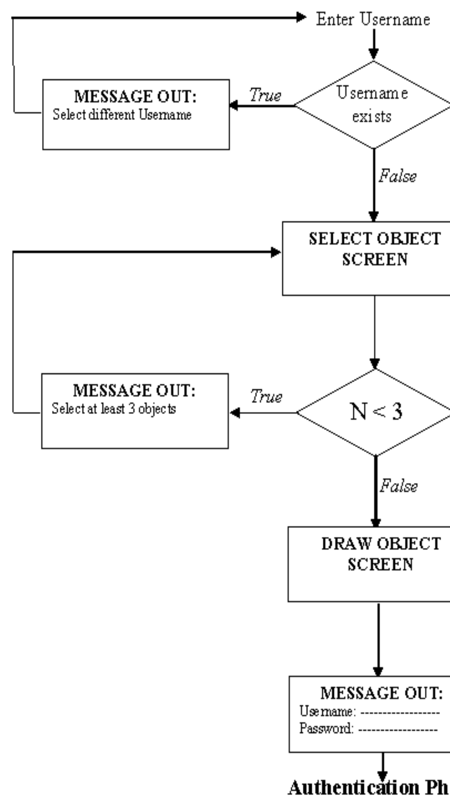


Fig. 6 Flow chart for Registration Phase

7. Comparison of Proposed System with Existing Schemes

Our system offers many advantages over other existing systems as discussed below:

Comparing to the "Passface" system, our system can also be used for those who are face-blind. We have used objects instead of human faces for selecting as password because later on during the authentication phase, the user has to draw his/her password and it is a much more difficult task to draw human faces than simple objects. Also we believe that as compared to human faces, objects are easier to remember which are in daily use. Our system has eliminated the problems with grid based techniques where the user has to remember the exact coordinates which is not easy for the user. Our system just compares the shapes of the objects drawn by the user during authentication.

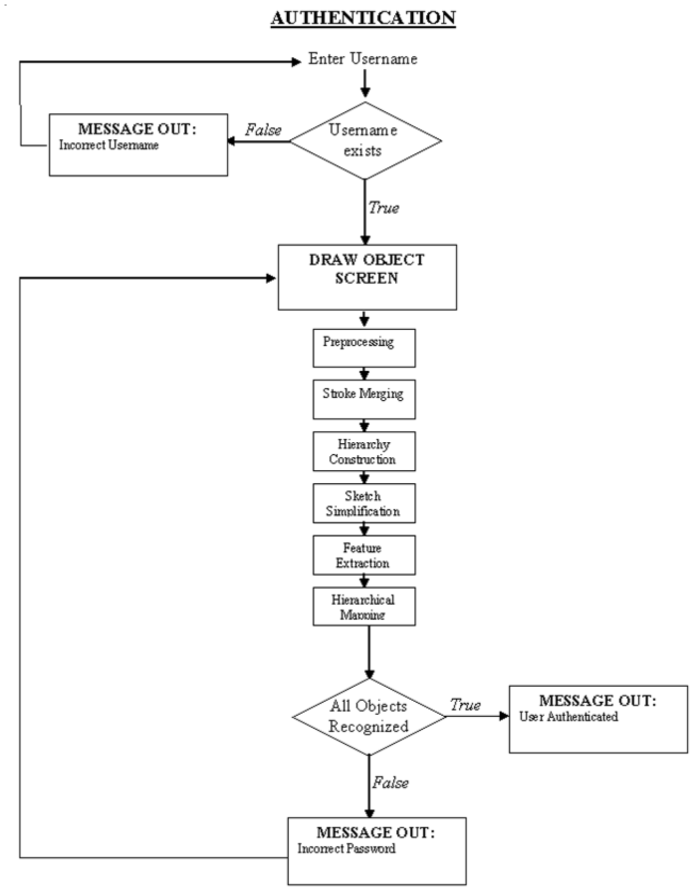


Fig. 7 Flow Chart for Authentication Phase

Our scheme is less vulnerable to Brute force attack as the password space is large. It is also less vulnerable to online and offline dictionary attacks. Since stylus is used, it provides ease to the user for drawing objects and also it will be impractical to carry out dictionary attack. Our scheme is better than Man et al scheme. This is because in his scheme the user has to remember both the objects and string and the code. In our method the user has to remember the objects he selected for password and also the way he has drawn the objects during registration.

Comparing to Van Oorschot's approach, our system is more secure since users not only select graphical password but also draw their password, making it difficult to hack. In our proposed system, even if the textual password is compromised, the graphical password cannot be stolen or compromised since the user is also drawing the graphical password. Our proposed system differs from CDS in that the user has to first select a textual password and then a graphical password, making it more secure. Comparing to Two Step Authentication system, our proposed system works in the same way as Two Step Authentication system i.e the user has to choose a textual password before choosing a graphical password but difference is that in our system during authentication, after giving the username and textual password, the user has to draw his graphical password which is matched with its stored template drawn by the user during the registration phase. This approach protects from hacking the password and prevents them from launching different attacks. Thus our system is more secure and reliable than two step authentication system. As with all graphical based systems our system will also be slow. The normalization and matching will take time. An important issue of our system is that it is somewhat user dependent during authentication. It depends upon the user's drawing ability. Thus, the system may not be able to verify the objects drawn by the user and as a result the actual user may not be authenticated.

The possible attacks on graphical passwords are Brute force attack, Dictionary attacks, Guessing, Spy-ware, Shoulder surfing and social engineering. Graphical based passwords are less vulnerable to all these possible attacks than text based passwords and they believe that it is more difficult to break graphical passwords using these traditional attack methods. Our System is resistant to almost all the possible attacks on graphical passwords. The comparison of our system to existing schemes and systems in resisting attacks on graphical passwords is shown in table 1.

8. Conclusion & Future Work

The core element of computational trust is identity. Currently many authentication methods and techniques are available but each with its own advantages and shortcomings. There is a growing interest in using pictures as passwords rather than text passwords but very little research has been done on graphical based passwords so far. In view of the above, we have proposed authentication system which is based on graphical password schemes. Although our system aims to reduce the problems with existing graphical based password schemes but it has also some limitations and issues like all the other graphical based password

Table 1: Comparison Of Graphical password Schemes resistant to different Attacks

Graphical Password Schemes/ Systems	Type of Scheme	Resistant to Possible Attacks					
		Brute Force Attack	Dictionary Attack	Guessing Attack	Spy-ware or Naïve Key logging	Shoulder Surfing Attack	Phishing Attack or Social Engineering
Blonder's Scheme [62]	Recognition Based	Y	N	Y	N	Y	N
DAS [31]	Pure Recall Based	N	Y	Y	N	Y	N
BDAS [32]	Pure Recall Based	N	-	-	-	-	-
Qualitative DAS [65]	Pure recall Based	N	-	-	-	-	-
Syukri Algorithm [64]	Pure recall Based	N	Y	Y	N	Y	N
PassPoints [41, 42, 43]	Cued Recall Based	Y	N	Y	N	Y	N
PassFace [10, 25]	Recognition Based	Y	Y	Y	N	Y	N
PassGo [5]	Pure Recall Based	Y	-	-	-	-	-
Passlogix [40]	Cued Recall Based	Y	N	Y	N	Y	N
PassMap [66]	Pure Recall Based	Y	N	-	N	Y	N
Passdoodle [38, 39]	Pure Recall Based	N	-	-	-	-	-
Viskey SFR	Pure Recall Based	Y	N	Y	N	Y	N
Perrig and Song [11]	Recognition Based	Y	N	Y	N	Y	N
Sobrado and Birget [3]	Recognition Based	Y	N	Y	N	N	N
Man et al Scheme [63]	Recognition Based	Y	N	N	Y	Y	N
Picture Password Scheme [60]	Recognition Based	Y	N	Y	N	Y	N
CDS [55]	Hybrid	-	-	-	-	Y	-
WIW [57]	Recognition Based	-	-	-	-	Y	-
Association based scheme [58]	Recognition Based	-	-	-	-	Y	-
Déjà Vu [1]	Recognition Based	Y	-	Y	-	-	-
Haptic Password Scheme [37]	Pure Recall Based	-	-	-	-	Y	-
YAGP [36]	Pure Recall Based	Y	-	Y	-	Y	-
Photographic Authentication [28]	Recognition Based	-	Y	-	-	-	-
Two Step Authentication [59]	Hybrid	-	-	-	Y	N	Y
Our Proposed System	Hybrid	Y	Y	Y	Y	Y	Y
Note: Y= Yes resistant to attack N=No not resistant to attack							

techniques. To conclude, we need our authentication systems to be more secure, reliable and robust as there is always a place for improvement. Currently we are working on the System Implementation and Evaluation. In future some other important things regarding the performance of our system will be investigated like User Adoptability and Usability and Security of our system.

Acknowledgment

The authors wish to acknowledge the anonymous reviewers for valuable comments.

References

[1] Rachna Dhamija and Adrian Perrig, "Deja Vu: A User Study. Using Images for Authentication" In Proceedings of the 9th USENIX Security Symposium, August 2000.

[2] Authentication: http://www.objs.com/survey/authent.htm [Last Visited on 15/05/2011].

[3] L.Sobrado and J.C. Birget, "Graphical Passwords", The Rutgers Schloar, An Electronic Bulletin for Undergraduate Research, vol 4, 2002, http://rutgersscholar.rutgers.edu/volume04/sobrbirg/sobrbirg.htm.

[4] Patric Elftmann, Diploma Thesis, "Secure Alternatives to Password-Based Authentication Mechanisms" Aachen, Germany October 2006

[5] Hai tao, ***"Pass-Go, a New Graphical Password Scheme"***, Master Thesis, University of Ottawa Canada, June 2006.

[6] Xiayuan Suo, YingZhu, G. Scott.Owen, ***"Graphical Passwords: A Survey"***, In Proceedings of Annual Computer Security Applications Conference, 2005.

[7] Approaches to Authentication: http://www.e.govt.nz/plone/archive/services/see/see-pki-paper-3/chapter6.html?q=archive/services/see/see-pki-paper-3/chapter6.html [Last Visited on 15/05/2011].

[8] Roman V. Y., "User authentication via behavior based passwords," Systems, Applications and Technology Conference. Farmingdale, NY. 2007.

[9] Biometric Authentication: http://www.cs.bham.ac.uk/ ~mdr/teaching/modules/security/lectures/biometric.html [Last visited on 02/05/11].

[10] Real User, www.realuser.com [Last Visited on 15/05/11].

[11] A.Perrig and D.Song, ***"Hash Visualization: A New Technique to improve Real-World Security"***. In International Workshop on Cryptographic Techniques and E-Commerce, pages 131--138, 1999.

[12] D.Davis, F.Monrose and M.K.Reiter, "*On User Choice in Graphical Password Schemes*". In 13th USENIX Security Symposium, 2004.

[13] Wing Ho Leung and Tsuhan Chen, "*Hierarchical Matching For Retrieval of Hand Drawn Sketches*", In Proceeding of International Conference on Multimedia and Expo - Volume 2 (ICME '03), 2003.

[14] Hafiz Zahid Ullah Khan, ***"Comparative Study of Authentication Techniques"***, International Journal of Video & Image Processing and Network Security IJVIPNS Vol: 10 No: 04

[15] Token Based Authentication: http://www.w3.org/2001/sw /Europe/events/foaf galway/papers/fp/token_based_authentication/ [last visited on 02/05/11].

[16] Knowledge Based Authentication: http://csrc.nist.gov/archive/ kba/index.html [Last Visited on 02/05/11].

[17] Knowledge based Authentication: http://searchsecurity.techtarget.com/definition/knowledge-based-authentication [Last Visited on 02/05/11].

[18] A Survey on Recognition based Graphical User Authentication Algorithms: http://www.scribd.com/doc/23730953/A-Survey-on-Recognition-Based-Graphical-User-Authentication-Algorithms [Last Visited on 02/05/11].

[19] A. Jain, R. Bolle, and S. Pankanti, Eds. ***"Biometrics: personal identification in networked society",*** Boston, MA: Kluwer Academic, 1999.

[20] A.R. Hurson, J. Ploskonka, Y. Jiao, and H. Haridas***, "Security issues and Solutions in Distributed heterogeneous Mobile Database Systems",*** Vol. 61, Advances in Computers, 2004, pp. 107-198.

[21] Robert Biddle, Sonia Chiasson, P.C. van Oorschot, ***"Graphical Passwords: Learning from the First Twelve Years"***, Carleton University - School of Computer Science, Technical Report TR-11-01, January 4, 2011.

[22] D. Weinshall, "***Cognitive authentication schemes safe against spyware"***, (short paper). In IEEE Symposium on Security and Privacy, May 2006

[23] E. Hayashi, N. Christin, R. Dhamija, and A. Perrig., ***"Use Your Illusion: Secure authentication usable anywhere",*** In 4th ACM Symposium on Usable Privacy and Security (SOUPS), Pittsburgh, July 2008.

[24] D. Davis, F. Monrose, and M. Reiter, ***"On user choice in graphical password schemes"***, In 13th USENIX Security Symposium, 2004.

[25] Passfaces Corporation. The science behind Passfaces,White paper, http://www.passfaces.com/enterprise/resources/white_papers.htm, [Last Visited on 05/05/11].

[26] A. De Angeli, L. Coventry, G. Johnson, and K. Renaud, ***"Is a picture really worth a thousand words? Exploring the feasibility of graphical authentication systems"***, International Journal of Human-Computer Studies, 63(1-2):128-152, 2005.

[27] W. Moncur and G. Leplatre, ***"Pictures at the ATM: Exploring the usability of multiple graphical passwords"***, In ACM Conference on Human Factors in Computing Systems (CHI), April 2007.

[28] T. Pering, M. Sundar, J. Light, and R. Want., ***"Photographic authentication through untrusted terminals"***, Pervasive Computing, pages 30-36, January - March 2003.

[29] S. Wiedenbeck, J. Waters, L. Sobrado, and J. Birget, ***"Design and evaluation of a shoulder-surfng resistant graphical password scheme"***, In International Working Conference on Advanced Visual Interfaces (AVI), May 2006

[30] K. Bicakci, N. B. Atalay, M. Yuceel, H. Gurbaslar, and B. Erdeniz, ***"Towards usable solutions to graphical password hotspot problem"***, In 33rd Annual IEEE International Computer Software and Applications Conference, 2009.

[31] I. Jermyn, A. Mayer, F. Monrose, M. Reiter, and A. Rubin, ***"The design and analysis of graphical passwords"***, In 8th USENIX Security Symposium, August 1999.

[32] P. Dunphy and J. Yan, ***"Do background images improve \Draw a Secret" graphical passwords?"***, In 14th ACM Conference on Computer and Communications Security (CCS), October 2007.

[33] R. Weiss and A. De Luca, ***"PassShapes-Utilizing stroke based authentication to increase password memorability"***, In NordiCHI, pages 383{392. ACM, October 2008.

[34] Tafasa. Patternlock, http://www.tafasa.com/patternlock.html [Last Visited on 11/05/11].

[35] GrIDsure. GrIDsure corporate website: http://www.gridsure.com [Last visited on 04/05/11].

[36] H. Gao, X. Guo, X. Chen, L. Wang, and X. Liu, ***"YAGP: Yet another graphical password strategy"***, In Annual Computer Security Applications Conference, 2008.

[37] M. Orozco, B. Malek, M. Eid, and A. El Saddik, ***"Haptic-based sensible graphical password"***, In Proceedings of Virtual Concept, 2006.

[38] J. Goldberg, J. Hagman, and V. Sazawal, ***"Doodling our way to better authentication"***, (student poster), In ACM Conference on Human Factors in Computing Systems (CHI), April 2002.

[39] C. Varenhorst, ***"Passdoodles: A lightweight authentication method'***, MIT Research Science Institute, July 2004.

[40] Passlogix graphical password system, www.passlogix.com [Last Visited on 01/04/11].

[41] S. Wiedenbeck, J. Waters, J. Birget, A. Brodskiy, and N. Memon, ***"Authentication using graphical passwords: Basic results"***, In 11th International Conference on Human-Computer Interaction (HCI International), July 2005.

[42] S. Wiedenbeck, J. Waters, J. Birget, A. Brodskiy, and N. Memon, ***"Authentication using graphical passwords: Effects of tolerance and image choice"***, In 1st Symposium on Usable Privacy and Security (SOUPS), July 2005.

[43] S. Wiedenbeck, J. Waters, J. Birget, A. Brodskiy, and N. Memon, ***"PassPoints: Design and longitudinal evaluation of a graphical password system"***, International Journal of Human-Computer Studies, 63(1-2):102{127, 2005.

[44] K. Renaud and A. D. Angeli, ***"My password is here! An investigation into visio-spatial authentication mechanisms"***, Interacting with Computers, 16(4):1017{1041, 2004.

[45] K. Renaud and E. Smith, ***"Jiminy: Helping user to remember their passwords"***, Technical report, School of Computing, Univ. of South Africa, 2001.

[46] X. Suo, ***"A design and analysis of graphical password"***, Master's thesis, College of Arts and Science, Georgia State University, August 2006.

[47] S. Chiasson, P. C. van Oorschot, and R. Biddle, ***"Graphical password authentication using Cued Click Points"***, In European Symposium On Research In Computer Security (ESORICS), LNCS 4734, pages 359{374, September 2007.

[48] S. Chiasson, A. Forget, R. Biddle, and P. C. van Oorschot, ***"Influencing users towards better passwords: Persuasive Cued Click-Points"***, In Human Computer Interaction (HCI), The British Computer Society, September 2008.

[49] A.Stubblefield and D. Simon, ***"Inkblot Authentication"***, MSR-TR-2004-85, Technical report, Microsoft Research, 2004.

[50] F. Alsulaiman and A. El Saddik, ***"A novel 3D graphical password schema"***, In IEEE Int. Conf. on Virtual Environments Human-Computer Interfaces and Measurement Systems, July 2006.

[51] Arash Habibi Lashkari, Dr. Rosli Saleh, Samaneh Farmand, Dr. Omar Bin Zakria, ***"A wide-range survey on Recall-Based Graphical User Authentications algorithms based on ISO and Attack Patterns"***, (IJCSIS) International Journal of Computer Science and Information Security,Vol. 6, No. 3, 2009.

[52] Farnaz Towhidi, Maslin Masrom, ***"A Survey on Recognition-Based Graphical User Authentication Algorithms"***, (IJCSIS) International Journal of Computer Science and Information Security, Vol. 6, No. 2, 2009.

[53] N. Govindarajulu and S. Madhvanath, ***"Password management using doodles"***, In 9th International Conference on Multimodal Interfaces (ICMI), November 2007.

[54] Michael Kimwele, Waweru Mwangi, Stephen Kimani, ***"Strengths of a Colored Graphical Password Scheme"***, International Journal of Reviews in Computing, 2009-2010 IJRIC& LLS.

[55] Haichang Gao, Zhongjie Ren, Xiuling Chang, Xiyang Liu, Uwe Aickelin, ***"A New Graphical Password Scheme Resistant to Shoulder-Surfing"***, 2010 International Confer-ence on CyberWorlds, Singapore, 20-22 October 2010.

[56] Ahmad Almulhem, ***"A Graphical Password Authentication System", World Congress on Internet Security (WorldCIS-2011), London, UK, February 21-23, 2011.***

[57] Shushuang Man, Dawei Hong, Manton Matthews, ***"A Shoulder Surfing Resistant Graphical Password Scheme-WIW"***, Proceedings of the International Conference on Security and Management, SAM '03, June 23 - 26, 2003, Las Vegas, Nevada, USA, Volume 1.

[58] Zhi Li1, Qibin Sun, Yong Lian, and D. D. Giusto, ***"An Association Based Graphical Password Design Resistant to Shoulder Surfing Attack"***, ICME2005, 2005.

[59] P.C. van Oorschot Tao Wan ***"TwoStep: An Authentication Method Combining Text and Graphical Passwords"*** 4th International Conference, MCETECH 2009, Ottawa, Canada, May 4-6, 2009.

[60]Wayne Jansen Serban Gavrila Vlad Korolev Rick Ayers Ryan Swanstrom, ***"Picture Password: A Visual Login Technique for Mobile Devices",*** NISTIR 7030, July 2003. Wayne Jansen, Serban Gavrila, Vlad Korolev, Rick Ayers, Ryan Swanstrom.

[61] Graphical password authentication, http://www.seminarprojects.com/Thread-graphical-user-based-authentication [Last Visited on 15/05/11].

[62] G. E. Blonder. Graphical password, U.S. Patent 5559961, Lucent Technologies, Inc. (Murray Hill, NJ), August 1995.

[63] S. Man, D. Hong, and M. Mathews, ***"A shoulder surfing resistant graphical password scheme",*** In Proceedings of International conference on security and management. Las Vegas, NV, 2003.

[64] A. F. Syukri, E. Okamoto, and M. Mambo, ***"A User Identification System Using Signature Written with Mouse",*** In *Third Australasian Conference on Information Security and Privacy (ACISP)*: Springer Verlag Lecture Notes in Computer Science (1438), 1998, pp. 403-441.

[65] D. Lin, P. Dunphy, P. Olivier, J. Yan, ***"Graphical Passwords & Qualitative Spatial Relations"***, Proceedings of the 3rd Symposium, On Usable Privacy and Security, Pittsburgh, USA, 2007, pp. 161-162.

Wazir Zada Khan is currently with School of Computer Science, Jazan University, Kingdom of Saudi Arabia. He received his MS in Computer Science from Comsats Institute of Information Technology, Pakistan. His research interests include network and system security, sensor networks, wireless and ad hoc networks. His subjects of interest include Sensor Networks, Wireless Networks, Network Security and Digital Image Processing, Computer Vision.

Dr. Muhammad Y Aalsalem is currently dean of e-learning and assistant professor at School of Computer Science, Jazan University. Kingdom of Saudi Arabia. He received his PhD in Computer Science from Sydney University. His research interests include real time communication, network security, distributed systems, and wireless systems. In particular, he is currently leading in a research group developing flood warning system using real time sensors. He is Program Committee of the International Conference on Computer Applications in Industry and Engineering, CAINE2011. He is regular reviewer for many international journals such as King Saud University Journal (CCIS-KSU Journal).

Dr. Yang Xiang is currently with School of Information Technology, Deakin University. He received his PhD in Computer Science from Deakin University. His research interests include network and system security, distributed systems, and wireless systems. In particular, he is currently leading in a research group developing active defense systems against large-scale distributed network attacks and new Internet security countermeasures. His recent research has been supported by the Australian Research Council (ARC), the University, and industry partners. Dr. Xiang has published more than 100 research papers in international journals and conferences. He has served as Program/General Chair for many international conferences such as ICA3PP 11, IEEE/IFIP EUC 11, TrustCom 11, IEEE HPCC 10/09, IEEE ICPADS 08, NSS 11/10/09/08/07. He has been PC member for many international conferences in distributed systems, networking, and security. He is regular reviewer for many international journals such as IEEE Transactions on Parallel and Distributed Systems, IEEE Transactions on Dependable and Secure Computing, IEEE Transactions on Information Security and Forensics, IEEE Communications Letters, and IEEE Journal on Selected Areas in Communications. He is on the editorial board of Journal of Network and Computer Applications.

POCS Based Super-Resolution Image Reconstruction Using an Adaptive Regularization Parameter

S.S.Panda[1], M.S.R.S Prasad[2] and Dr. G. Jena[3]

[1] **Computer Science & Engg., Pondicherry University, Regency Institute of Technology Yanam, Puducherry 533464, India**

[2] **Computer Science & Engg., Pondicherry University, Regency Institute of Technology Yanam, Puducherry 533464, India**

[3] **Computer Science & Engg., JNTU University, B V C Engineering College Amalapuram, Andhra Pradesh 533210, India**

Abstract

Crucial information barely visible to the human eye is often embedded in a series of low-resolution images taken of the same scene. Super-resolution enables the extraction of this information by reconstructing a single image, at a high resolution than is present in any of the individual images. This is particularly useful in forensic imaging, where the extraction of minute details in an image can help to solve a crime. Super-resolution image restoration has been one of the most important research areas in recent years which goals to obtain a high resolution (HR) image from several low resolutions (LR) blurred, noisy, under sampled and displaced images. Relation of the HR image and LR images can be modeled by a linear system using a transformation matrix and additive noise. However, a unique solution may not be available because of the singularity of transformation matrix. To overcome this problem, POCS method has been used. However, their performance is not good because the effect of noise energy has been ignored. In this paper, we propose an adaptive regularization approach based on the fact that the regularization parameter should be a linear function of noise variance. The performance of the proposed approach has been tested on several images and the obtained results demonstrate the superiority of our approach compared with existing methods.

Keywords: *Super resolution, POCS, regularization*.

1. Introduction

Super-resolution image reconstruction is useful (sometimes essential) to enhance the image resolution in many practical cases such as medical imaging, satellite imaging, and video applications, where several low resolution images of the same content can be obtained. Signal processing techniques are often used to obtain HR images [1, 10]. In these techniques, the fusion of all kinds of information obtained from various LR images is used to enhance the resolution. The most successful methods are stochastic approaches in spatial domain such as ML (Maximum Likelihood), MAP (Maximum A-Posteriori), and POCS (Projection onto Convex Sets) [3]. In these methods, based on a linear model describing the relation of HR and LR images, a cost function is introduced and the HR image is estimated. POCS algorithm has been widely used because it is simple, can be applied to the occasion with any smooth movement, linear variable airspace vague and non-uniform additive noise and can easily join in the prior information. But POCS algorithm is strict to the accuracy of movement estimation. So in order to improve the stability of the algorithm, the relaxation operator will be used to replace ordinary projector operator, at the same time it is not conducive to the resumption of the edge and details of images [4]. However, the linear model used in these methods is an ill-posed problem in the sense that its transformation matrix may be singular and consequently a unique solution cannot be achieved [5,11]. In this paper, the problem of reconstructing an HR image is solved by choosing an adaptive regularization parameter to stabilize the inversion of ill-posed problem and to consider the effect of the noise. The obtained results show the performance of this new approach.

2. POCS Super-Resolution

K low resolution (LR) images can be seen as a high resolution (HR) after geometric distorting, fuzzy linear space, down sample and additional noise. Therefore, the imaging process can be written as [4]

$$G_k = D_k B_k M_k F_k + \xi_k \qquad (1)$$

G_k is the observed image, size is M X M

D_k is down sample matrix, size is M X L

M_k is geometric distortion matrix, size is LX L

B_k is fuzzy linear space matrix, size is L X L

F is the original high-resolution image, size is L X M
ξ_k is Gaussian white noise, size is M ×M

Different observed images are obtained from different geometric distortion, spatial ambiguity, down sampling and additional noise to ideal images. The model above can be expressed as follows:

$$\begin{pmatrix} G_1 \\ . \\ . \\ . \\ G_N \end{pmatrix} = \begin{pmatrix} D_1B_1M_1 \\ . \\ . \\ . \\ D_1B_1M \end{pmatrix} * X + \begin{pmatrix} \xi_1 \\ . \\ . \\ . \\ \xi_N \end{pmatrix} = \begin{pmatrix} H_1 \\ . \\ . \\ . \\ H_N \end{pmatrix} * X + \xi_k \qquad (2)$$

3. POCS Algorithm

The POCS approach to the SR reconstruction problem has been proposed in [7]. Let the motion information be provided. Then, a data consistency constraint set based on the acquisition model (2) can be defined for each pixel with each low-resolution (LR) image. The convex sets in any LR image are given by:

$$C_{n_1,n_2;i,k} = \left\{ x_i(m_1,m_2) : \left| r_k^{x_j}(n_1,n_2) \right| \le \phi_0 \right\} \qquad (3)$$
$$0 \le n_1, n_2 \le N-1, k = 1,2,.....,L$$

Where the value at each pixel is constrained such that it's associated residual

$$r_k^{x_i}(n_1,n_2) = g_k(n_1,n_2) - \sum_{m_1=0}^{M-1}\sum_{m_2=0}^{M-1} x_i(m_1,m_2)\, h_i(m_1,m_2;n_1,n_2) \qquad (4)$$ is

bounded in magnitude by φ_0 for the set. Since φ_0 is determined from the statistics of noise, the ideal image solution is a member of the set satisfying a certain statistical confidence. The projection of an arbitrary $x_i(m_1,m_2)$ onto $c_{n_1,n_2;i,k}$ is defined by

$$P_{n_1,n_2;j,k}[x_i(m_1,m_2)] =$$
$$\begin{cases} x_i(m_1,m_2) + \dfrac{\left(r_k^{x_i}(n_1,n_2) - \phi_0\right) h_i(m_1,m_2;n_1,n_2)}{\sum_{m_1=0}^{M-1}\sum_{m_2=0}^{M-1} h_i^2(m_1,m_2;n_1,n_2)}, & r_k^{x_i}(n_1,n_2) > \phi 0 \\ 0, & -\phi_0 < r_k^{x_i}(n_1,n_2) < \phi_0 \\ x_i(m_1,m_2) + \dfrac{\left(r_k^{x_i}(n_1,n_2) + \phi_0\right) h_i(m_1,m_2;n_1,n_2)}{\sum_{m_1=0}^{M-1}\sum_{m_2=0}^{M-1} h_i^2(m_1,m_2;n_1,n_2)}, & r_k^{x_i}(n_1,n_2) > -\phi 0 \end{cases} \qquad (5)$$

Additional constraints such as bounded energy, positivity, and limited support may be utilized to improve the results. A generally utilized altitude constraint set is

$$P_A[x_i(m_1,m_2)] = \begin{cases} 0, x_i(m_1,m_2) < 0 \\ x_i(m_1,m_2), 0 \le x_i(m_1,m_2) \le 255 \\ 255, x_i(m_1,m_2) > 255 \end{cases}$$

4. Regularization Approach

One of the difficulties in solving Eq. (1) is inverting the forward model without amplifying the effect of noise. In addition, the minimizing process should make the solution insensitive to noise. Because the transformation matrix (H) is ill-conditioned, the obtained result may not be the best one. Thus some forms of regularization must be included in the cost function to solve the problem and to minimize the effect of the noise. The regularization is indeed some sort of constraints imposed on the space of possible solutions and is often independent of measured data [8].

$$J(X) = \|Y - HX\|^2 + \lambda p(X)$$

The function $p(X)$ poses a constraint on the unknown X to direct it to a stable solution, and the coefficient λ represents the strength of this constraint. One of the most common forms for p(X) is Tikhonov regularization; $p(X) = \|TX\|^2$, in which T is a matrix determined according to some aspects of the desired image such as energy or smoothness.

4.1 Proposed Algorithm

In this paper, we take T = I (identity matrix) which is minimal energy regularization and leads to a stable and unique solution [8].

$$J(X) = \|Y - HX\|^2 + \lambda \|X\|^2$$

Another important issue is the proper choice of λ. λ is regularization parameter that controls a trade-off between the fidelity to data (expressed by $\|Y-HX\|^2$) and the energy of the solution (expressed by $\|X\|^2$). Larger values of λ are useful when small number of LR images are available or the fidelity of data is low, which is caused by registration error and noise. On the other hand, smaller values of λ are helpful whenever we have large number of LR images and noise is small [1]. As a result, larger noise variances lead to larger values of λ and vice versa.

Based on above observation, we can conclude that there is a linear relation between λ and noise variance as the form of $\lambda = m\sigma^2 + c$, where σ^2 is the noise variance, m represents the weight of noise energy and c is an offset. This is an adaptive regularization model that more efficiently minimizes Eq. (11) with respect to the effect of the noise.

It is important to note that although like regularization approaches an extra term is included in POCS estimator, because we do not have enough data to accurately estimate correlation matrix of *X*, *C* cannot be a precise estimation. It is the reason that our proposed regularization approach with adaptive regularization parameter can lead to better results.

5. Simulation Result

In our simulations LR images have been obtained by random displacement with uniform distribution over -10 and 10 pixels, blurring using linear motion by 5 pixels with an angle of 5 degrees, and the decimation with L1 =□L2 = □2 . Moreover, LR images have been corrupted by noise (AWGN) with SNR =□20 dB . Fig. 1 shows the original test image of Lena and its LR images.

To quantitatively evaluate the performance of the algorithms, we computed the PSNR using the original and reconstructed images. PSNR in dB is defined as [5]

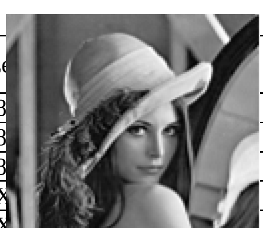
$$PSNR = 10\log\left[\frac{255^2}{\frac{1}{N}\left\|\hat{X}-X\right\|^2}\right]$$

Where N is the total number of pixels, X is the original image, and $\hat{X}$ is the reconstructed image [9].

Firstly, in order to select the best linear model, many values of weight (m) and offset (c) have been tested. Table I shows the results on three well-known test images. The best results were achieved using m =□10^{10} and c = □0.

Secondly, the performance of POCS and the proposed algorithm has been evaluated and compared using different test images. Table II shows PSNR values for different images using POCS, and proposed algorithm with m = □10^{10} and c= □0.

Tested Image	Size	[illegible]	[illegible]	Proposed Algorithm
lena	128[illegible]	[illegible]	[illegible]	69.80
Barbara	128[illegible]	[illegible]	[illegible]	69.95
Peppers	128[illegible]	[illegible]	[illegible]	67.90
Test 1	30X[illegible]	[illegible]	[illegible]	73.04
Test 2	40X[illegible]	[illegible]	[illegible]	80.94
Test 3	50X50	15	48.33	62.70
Test 4	100X100	15	69.73	71.52



(a)

(b)

Figure 1 (a) original image (twice the real size), (b) four frames of LR images used to estimate the high resolution image.

Table I. Comparison Of Psnr Values For Different Images And Different Weights Withc=0 (Psnr In Db)

Tested images	Different values of m				
	10^{-1}	10^{-3}	10^{-5}	10^{-10}	10^{-12}
Lena	8.65	28.22	67.39	72.54	71.54
Peppers	6.58	25.32	61.76	65.26	64.26
Barbara	8.10	29.74	66.37	69.20	68.20

(a) (b)

Figure 2. Column (a) shows the original images and column, Column (b) shows the reconstructed images using proposed algorithm for Lena, Barbara and Peppers .

Table II. Quantitative Performance Comparison Of POCS Algorithms Using PSNR

It can be seen that the proposed algorithm significantly improves the quality of the reconstructed high resolution images for all tested images compared with POCS algorithm.

4. Conclusions

This paper deals with achieving a high resolution image from several low resolution images of the same content. Specifically, we propose an adaptive regularization approach based on our observation that the noise energy should affect the regularization process. We have tested our approach on a variety of different images with different resolutions, and provided superior performance compared with other existing stochastic methods. The quantitative evaluation of the algorithms is based on PSNR which allows a good performance assessment.

References

[1] S.C.Park, M.K.Park, and M.G.Kang, "Super-resolution image reconstruction: a technical review," IEEE Signal Processing Magazine, pp. 21-35, May 2003.

[2] M.B.Chappalli, N.K.Bose, "Simultaneous noise filtering and super-resolution with second-generation wavelets," IEEE Trans. Signal Processing, vol. 12, no. 11, pp. 772-775, November 2005.

[3] H. Shen, P. Li, L.Zhang, and Y. Zhao, "A MAP algorithm to super-resolution image reconstruction," in Proc. IEEE Int. conf.Image and Graphics (ICIG'04), pp. 544-547, 2004

[4] Guo Lei He Zhiming,A Projection on Convex Sets Super-Resolution Algorithm Using Wavelet Transform, Signal Processing, 2008.ICSP 2008. 9th International Conference on
Issue Date: 26-29 Oct. 2008

[5] Sareh Abolahrari Shirazi and Mehran Yazdi A New Approach In Super Resolution Based On An
Adaptive Regularization Parameter, ISCCSP 2008, Malta, 12-14 March 2008

[6] [1] Bin Tian, Jennting T.Hsu, Qiang Liu, "A Wavelet Constrained POCS Super-resolution Algorithm for High Resolution Image Reconstruction from Video Sequence", IEEE Int. Conf. Neural Networks&Signal Processing Nanjing, China, December 14-17, 2003.

[7] D. C Youla, "Generalized image restoration by the method of altemating orthogonal projections", IEEE Ttnm.
Circirirs Xvsr, vol. CAS-25: pp.694-702, 1978

[8] [6] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Advances and challenges in super-resolution, " Wiley Periodicals, Inc, vol. 14, pp. 47-57, 2004.

[9] M. Elad, and A. Feuer, "Restoration of a single superresolution image from several blurred, noisy, and under sampled measured images," IEEE Trans. Image Processing. Israel, vol. 6, no. 12, pp.1646-1658, December 1997.

[10] Marcelo V. W. Zibetti, Fermín S. V. Bazán , Joceli Mayer,Determining the Regularization Parameters for Super-Resolution Problems , Preprint submitted to Elsevier,July 19, 2008

[11] Jing Tiana, , , Li Chena, , Lihong Mab, and Weiyu Yuc,Phase-driven spatially variant regularization for image resolution enhancement ,AEU - International Journal of Electronics and Communications, April 2011.

S. S. Panda got his M.Tech. degree in the year 2010 and doing his research on Digital image processing. He is working as Sr. Assistant Professor in the department of Computer Science and Engineering, Regency Institute of Technology, Yanam.(Puducherry), India. He is a life member of ISTE.

MSRS Prasad got his M.Tech. degree in the year 2010 and doing his research on Digital image processing. He is working as Sr. Assistant Professor in the department of Computer Science and Engineering, Regency Institute of Technology, Yanam.(Puducherry), India. He is a life member of ISTE.

Dr. G. Jena got his Ph.D. degree from F M university and working as Professor and HOD CSE, BVC Engineering College, Amalapuram, India. His area of interest is Image Processing, Signal Processing and Adaptive Channel Equalizers using Time Frequency Domain Transform and Neural Techniques and is a member of the IEEE,FIE,LMISTE, MCSI, India .

Real-Time Blood Donor Management Using Dashboards Based on Data Mining Models

Shyam Sundaram[1] and Santhanam T[2]

**[1] PG & Research, Dept. of Computer Science, DG Vaishnav College
Chennai 600106, Tamil Nadu,India**

**[2] PG & Research, Dept. of Computer Science, DG Vaishnav College
Chennai 600106, Tamil Nadu,India**

Abstract

This study uses data mining modeling techniques to examine the blood donor classification and extending this to facilitate the development of realtime blood donor management using dashboards with blood profile and geo-location data. This enables decision makers the ability to manage and plan the blood donation activities based on key metrics. This capability provides the ability to plan effective targeted blood donation campaigns. The scoring algorithm implemented for the dashboard also helps in the optimized deployment of budget resources and budget allocation determination for blood donation campaigns.

Keywords: **Dashboards, Blood bank, blood transfusion, blood donor, data mining, classification algorithms, Decision support, blood bank information systems**

1. Introduction

It is essential for healthcare systems to have a constant balance of supply and demand for blood products. These play a critical role in the saving and the extension of life. Dashboards are of critical importance to help achieve these objectives. There are two critical aspects of these dashboards that will be covered in this paper. Firstly the ability to have realtime data that identifies the donor profiles based on their patterns of donorship. Secondly rolling up the profile to demographic level both summary and detail level. In specific the linkage to geo-location will be demonstrated. This paper dwells into the effective dashboard creation using data mining techniques coupled with geo-locational linkage. The objective of this research is to effectively translate data mining derived models to enable decision making with the help of dashboards.

This paper is organized as follows. Section two deals with the introduction to blood donorship and relevant peer research in the context of this paper. Section three describes details with reference to the dataset and analysis. The conclusion is given in the final section.

2. Blood Donorship

Major headings are to be column centered in a bold font without underline. A donation is when a donor gives blood for storage at a blood bank for transfusion to an unknown recipient. These can occur at a number of locations including blood donation centers,mobile camps, mobile vans,etc. There a number of types of blood donations such as voluntary blood donation programme. This is the foundation for safe and quality blood transfusion service as the blood collection from voluntary non-remunerated blood donors is considered to be the safest. In order to augment voluntary blood donation in developing countries like India[1] is based on well defined frameworks and operational guide for organizations for this important activity. International healthcare research bodies have extensive frameworks that address context of blood management[2]. In developed countries there are dedicated organizations that have effective blood donor management processes. One such example is the U.S. department of defense (DOD) , which uses an enterprise blood Management software that will manage the blood supply chain including donor management, blood collections, testing, distribution and transfusion. Additionally this also provides a proactive delivery of information and services through a web portal[3].

2.1 Relevant Peer Research

Santhanam et al[4][5] extended the nominal definition based on a standard dataset to derive a CART[6] based decision tree model based on standard donorship. This analysis helped identify the attributes that classify a regular voluntary donor (RVD) in the context of a standard dataset. This provided an extended RVD definition based on the donor definition (along with the application of CART) provides a standard model to determine the donor behavior and provides the capability to build a classification model. This additional nominal class can be easily computed based on the statistical definitions and help assist in decision making.

Chau et al[7] have extensively analyzed the linkages related to the blood donation to the location of the blood donation centers.This research was carried out using donor's past donation profiles to help setup a new blood donation center for the Hong Kong Red Cross.Their findings provide correlations between spatial distance and the incentive for the blood donors which is the uniqueness of this research. This specifically helps in the effective setup of centers with maximal donorship potential.

Saberton et al[8] have extensively analyzed the linkages related to the blood donation to the location of the blood donation centers.Their findings provide correlations between spatial distance and the incentive for the blood donors. This specifically helps in the effective setup of centers with maximal donorship potential.

Bing et al[9] have extensively analyzed the working and implementation of blood bank information systems. Their research provides an extensive background of blood bank information systems. The research also talks about the importance of the decision making capability that is required for effectively running the operations in blood banks. The research also identifies various critical areas that are required for the systems to also have in order to enable decision making.

3. Analysis

3.1 About the Dataset

The blood transfusion dataset (taken from the UCI ML repository)[10] is based on donor database of Blood Transfusion Service Center in Hsin-Chu City in Taiwan. The center passes their blood transfusion service bus to one university in Hsin-Chu City to gather blood donated about every three months. This dataset is derived from I-Cheng Yeh[11].

The data set consists of 748 donors at random from the donor database. These 748 donor data, each one included R (Recency - months since last donation), F (Frequency - total number of donation), M (Monetary - total blood donated in c.c.), T (Time - months since first donation), and a binary variable representing whether he/she donated blood in March 2007 (1 stand for donating blood; 0 stands for not donating blood). There is an imbalance in that the people who have donated blood in 2007 accounts for only 24% in the dataset.

This dataset has been extended to accommodate the following attributes. RVD a boolean attribute that is computed based on the original attributes along with definitions[1]. Additionally a geo-location information was added in the syntax of latitude:longitude. This was randomly assigned to locations in India for analysis. Please note the data used is to be considered only for demonstrative purposes.

3.2 Analysis

The past research of [5] resulted in creating an extended RVD definition based on the donor definition (along with the application of CART) provides a standard model to determine the donor behavior and provides the capability to build a classification model. The ability to easily compute this based on statistical and definition data provided by frameworks[1]. additional nominal class can be easily computed based on the definition. The results of the decision tree help refractor the definitions of the RVD with the following offsets. These have had been defined using suggestive definitions[5]. The dataset is now corrected to these offsets and used for this analysis.

```
IF ( (Frequency > 18.5 ( times)  AND
Recency < 8.5 (months)  )
        RVD = TRUE
ELSE
        RVD = FALSE
```

This results in a finer refinement to the RVD model. The RVD confusion matrix(using the tool weka[12]) post this is provided in the following table.

Table 1: Revised RVD confusion matrix

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
Class 0 (not RVD)	0.92	0	1	0.92	0.96	0.96
Class 1 (RVD)	1	0.08	1	1	1	0.96
Weighted Avg.	1	0.08	1	1	1	0.96

Table 2: Previous RVD confusion matrix

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
Class 0 (not RVD)	1	0.06	0.99	0.92	0.99	0.97
Class 1 (RVD)	1	0.08	1	1	1	0.96
Weighted Average	1	0.08	1	1	1	0.96

The comparison between the RVD before the offset and with it indicate an overall stability to the model with delta change to a better true positive rate for a non RVD and also small increase in the FP rate the non RVD.The inclusion of geospatial location along with the donor data provides critical indicative identification of the RVD. This allows the capability to search by geo-locational attributes which enables targeted blood donation program management including aspects related to logistics and infrastructure linkages. Additional linking to census and demographic information[8] allows the effective determination of blood donor profiles with capabilities to drill-down to the appropriate levels.

Please note the this analysis has been developed using random geo-locational values (dummy values) which have helped to provide a meaningful endpoint of this analysis.

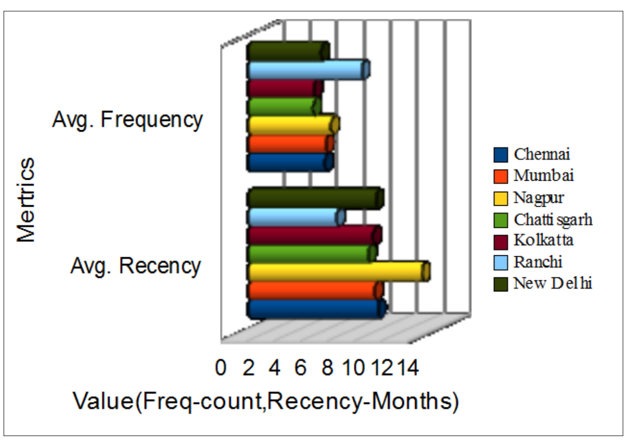


Fig. 1 RVD Geo-locational profiles.

This is further analyzed by a perspective at the overall dashboard across the the indicators and ranking the locations by scores. The algorithm for the dashboard is indicated as follows. Figure 2 provides the implementation of this algorithm.

Geo-location RVD Scoring Algorithm
Step 1: Loop through each unique location L (latitude,longitude) based on geographic division (such as state, district and city).
Step 2: For each location L compute the average frequency,average recency and total RVD count.
Step 3: Calculation of Location level summary scores for the recency,frequency and RVD across the locations.
The RecencyScore (location) is computed as the Rank in descending.
The FrequencyScore (location) is computed as the Rank in ascending.
The RVDScore (location) is computed as the Rank in ascending.
Step 4: Plot the this score in the chart with scores on the X – Axis and locations on the Y-axis.

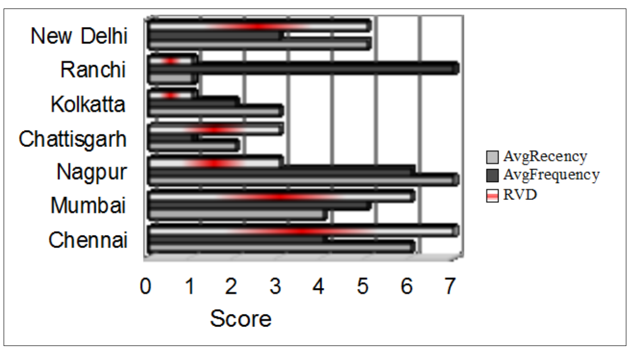


Fig. 2 Location-wise indicator patterns.

The results in comparison with the earlier model [5] reveal an improvement in the true positive rate for RVD class along with a delta increase in the false positive rate.

4. Conclusions

The dashboard indicated in figure 2 provide a quick and relevant score of the the geographic locations based on the RVD profile key indicators. The geo-location RVD scoring algorithm can be modified to rollup to additional attributes as well as handle the requisite geographic division strategy. This capability enables this to be linked effectively to the census tracts as well as health profile systems that enable drill-down to finite levels of information for effective blood donor management. This paper provides a complimentary capability to the recent research[7][8][9]. The application of this across a larger dataset and linkage to both demographic and census tracts will enable the ability to identify meaningful patterns of blood donorship that will assist in the better management.

This provide critical decision makers the ability to make planned decisions. This demonstrates a viable mechanism to manage blood donorship. In specific this helps address the optimized deployment of budget resources related to blood donations drives. This also assists policy makers plan the required budget allocation for overall blood donation related activities in a addressing targeted goals. Such techniques assist in the decision support for healthcare organizations.

Future work will be focused on further enhancing these models to allow integration with blood donor management systems including innovative ways of visualization. The current implementation of the RVD model can also be implemented with other relevant attributes. Similar strategies can also be adopted for other healthcare domains.

References

[1] Ministry of Health and Family Welfare,National AIDS Control Organization Government of India. Name, Voluntary Blood Donation Programme , 2007.

[2] MedlinePlus, U.S.National Library of Medicine, NIH,USA.Blood Transfusion and Donation ,http://www.nlm.nih.gov/medlineplus/bloodtransfusionanddonation.html, 2011.

[3] The Armed Services Blood Program (ASBP, ,http://www.militaryblood.dod.mil/About/default.aspx 2011.

[4] Shyam Sundaram and T. Santhanam, "Classification of Blood Donors using Data Mining", Proceedings of the Semantic E-Business and Enterprise Computing , 2009, pp. 145-147.

[5] Shyam Sundaram and T. Santhanam, "Application of CART Algorithm in Blood Donors Classification", Journal of Computer Science, Vol. 6, No. 5, 2010, pp. 548-552.

[6] L. Breiman, J. Friedman, R. A. Olshen and C. J. Stone. Classification and regression trees, Wadsworth ,1984.

[7] Michael Chau, Eddie Cheng and Chi Wai Chan, "Data Analysis for Healthcare: A Case Study in Blood Donation Center Analysis", Proceedings of Sixteenth Americas Conference on Information Systems(AMICS), 2010.

[8] Saberton, P. J., Paez, A., Newbold, K. B., and Heddle, N. M., "Geographical Variations in the Correlates of Blood Donor Turnout Rates: An Investigation of Canadian Metropolitan Areas", International Journal of Health Geographics ,Vol. 8, No. 56, 2009.

[9] Bing Nan Li, Ming Chui Dong and Sam Chao, "On decision making support in blood bank information systems", Expert Systems with Applications,Vol. 34, No. 2, 2009, pp. 1522-1532.

[10] A. Asuncion and D.J. Newman, UCI repository of machine learning database" www.ics.uci.edu/_mlearn/MLRepository.html,, 2007.

[11] I.C. Yeh , K.J. Yang, Ting and T. Ming, "Knowledge discovery on RFM model using bernoulli sequence", Expert Systems with Applications,Vol. 36, No. 5, 2009, pp. 5866-5871.

[12] Ian H. Witten and Eibe Frank. Data Mining: Practical machine learning tools and techniques, Wadsworth ,Morgan Kaufmann, San Francisco, 2005.

IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 2, September 2011
ISSN (Online): 1694-0814
www.IJCSI.org

Reliable Energy Conservation in Wireless Sensor Networks using Flooding Techniques

L. Nalini Joseph[1], Dr.S.Shanmugavel[2]

[1]Research Scholar, [2]Professor
Department of Electronics &Communication Engineering,
College of Engineering, Anna University,
Chennai, India.

Abstract: A Comparison of Performance in terms of Energy conservation in Wireless Sensor Networks using different Flooding mechanism **has** been analyzed. The following flooding mechanisms analyzed are Classical Flooding, Location Aided Flooding, Adaptive Location Aided Flooding; Diagonal arc based Location Aided Flooding, Diagonal arc based Location Aided Flooding with reliability is considered for the study. The various mechanisms save energy when compared to conventional flooding schemes.

Keywords: *Sensor Networks, Flooding, Virtual Grids, ALAF, DALAF.*

1. Introduction:

Sensor networks are composed of a large number of sensor nodes that are densely deployed either inside the phenomenon or very close to it [1]. The position of sensor nodes need not be engineered or predetermined. This allows random deployment in inaccessible terrains or disaster relief operations. On the other hand, this also means that sensor network protocols and algorithms must possess self-organizing capabilities [2]. Another unique feature of sensor networks is the cooperative effort of sensor nodes [1]. Sensor nodes are fitted with an onboard processor. Instead of sending the raw data to the nodes responsible for the fusion, they use their processing abilities to locally carry out simple computations and transmit only the required and partially processed data.

Conventional protocols use classical flooding [] for disseminating data in a sensor network. Flooding [11] is also used as a preprocessing step in many routing protocols in networks for disseminating route discovery requests. Information dissemination protocols are used in networks for distributing the link state information. Routers in the Internet periodically use flooding to update link state information at other nodes.

Despite many of its uses, flooding suffers from disadvantages such as the broadcast storm problem [12]. There are situations when duplicated messages are sent to the same node and also if two nodes share the same observing region, neighbor nodes receive duplicated messages. The flooding protocol does not take into account of the available energy resources.

Sensor nodes are typically characterized by small form-factor, limited battery power, and a small amount of memory. So there is a need for an energy-efficient flooding mechanism for information dissemination in distributed sensor networks.

2. Adaptive location aided flooding (ALAF)

To overcome the above issues, the concept of integrating non-uniform virtual grids into location aided flooding to form Adaptive Location Aided Flooding (ALAF) is proposed. In ALAF [3], the grids with dense deployment are further sub-divided into smaller grids thereby non-uniform grids of dissimilar grid sizes are present in the sensor network. The node list is frequently stripped off avoiding excessive increase in its size, which in turn conserves appreciable amount of energy on location aided flooding for information transfer to neighbor nodes or to the sink. The description of ALAF structural formation, functional design, its operation and energy conservation principals are presented in this chapter.

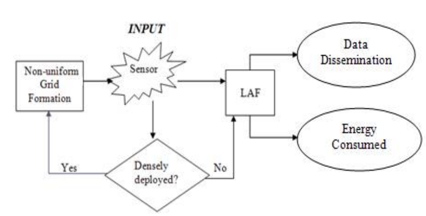


Fig. 1 Block Diagram of ALAF

In ALAF, the LAF[2] mode with densely populated sensor nodes are converted into non-uniform virtual grids for location aided flooding as shown in Fig 1. The virtual grid formation varies as and when the node density increases in the specific location. If the number of nodes in a grid exceeds a certain threshold, then the grid is further split into grids of smaller dimensions. Each node associates itself with a virtual grid and is classified as either gateway node or internal node.

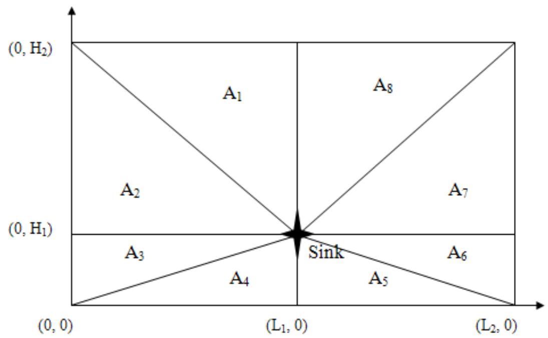


Fig. 2 Non Uniform Grid Formation in ALAF

A source node starting the ALAF broadcasts the packet to all its neighbors. The receiving node does further broadcasts. When a gateway node receives a packet from within its virtual grid, it checks to see if any of its neighbors within the same virtual grid have not yet received the packet. This is done by comparing the node list of the packet with the neighbor list of the node. If such nodes exist, then the gateway node appends the ids of those nodes to the node list of the packet and forwards it to the neighbor nodes that still have not received the message. When a gateway node receives a packet from another gateway node, it strips the packet of its node list and adds its own id and all its neighbors' ids and forwards the packet to all its neighbors. Thus, the packet becomes shorter as it moves across the virtual grids and increases in size as it moves within a virtual grid. When an internal node receives a packet, it modifies the node list of the packet. It includes the ids of all its neighbors in the node list and forwards it to its neighbors if they have not already received a message.

The algorithm for grid formation and node classification is given below:

1. Divide the sensor field into grids based on topography and grid size.
2. For each node,

- Based on its location in one of the grids formed, set the grid id and maintain the node count in each grid.
- If the node count of the grid exceeds threshold value,
- Get the dimensions of the grid and the new grid size.
- Split the grid based on the input parameters.
- Maintain the count of number of grids formed.
- Set new grid id for the node based on its location in one of the new grids formed.
- Identify the neighbors and store in the neighbor list.
- Classify the node
- If all of its neighbors belong to the same grid as that of the node, then set its node type as internal.
- If any of its neighbors belong to a different grid than that of the node, then set its node type as gateway.

3.1. DATA DISSEMINATION:

A source node start broadcasts the packet to all its neighbors. The receiving node does further broadcasts in one of the following ways:

When a gateway node receives a packet from within its virtual grid, it checks to see if any of its neighbors within the same virtual grid have not yet received the packet. This is done by comparing the node list of the packet with the neighbor list of the node. If such nodes exist, then the gateway node appends the ids of those nodes to the node list of the packet and forwards it to the neighbor nodes that still have not received the message.

When a gateway node receives a packet from another gateway node, it strips the packet of its node list and adds its own id and all its neighbors' ids and forwards the packet to all its neighbors. Thus, the packet becomes shorter as it moves across the virtual grids and increases in size as it moves within a virtual grid.

When an internal node receives a packet, it modifies the node list of the packet. It includes the ids of all its neighbors in the node list and forwards it to its neighbors.

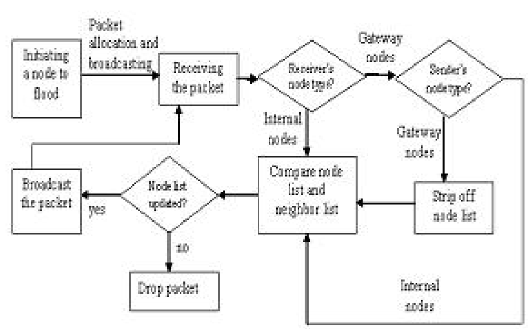


Fig 3. Data Dissemination in ALAF

The proposed algorithm for data dissemination is given below:

1. The source node creates a packet with its id, sequence number, node list, grid id, node type, node list length. The node list contains its id and the ids of all its neighbors.
2. The source node broadcasts the packet to its neighbors.
3. The receiving node checks to see whether the packet's sequence number is already in its received list of sequence numbers.
   a. If so, it drops the packet.
   b. Else, it stores the sequence number of the packet it receives and sets its id, grid id, node type in the packet.
   i. If it is an internal node, it compares its neighbor list with the node list in the packet.
   1. If there are nodes in the neighbor list not present in the node list, it adds those nodes to the node list and broadcasts the packet to its neighbors.
   2. Else, it drops the packet.
   ii. If it is a gateway node, it checks the node type in the received packet.
   1. If it has received from an internal node, it compares its neighbor list with the node list in the packet.
      a. If there are nodes in the neighbor list not present in the node list, it adds those nodes to the node list and broadcasts the packet to its neighbors.
      b. Else, it drops the packet.
   2. If it has received from a gateway node, it checks the grid id in the received packet.
      a. If it has received from a node in the same grid, it compares its neighbor list with the node list in the packet.
      i. If there are nodes in the neighbor list not present in the node list, it adds those nodes to the node list and broadcasts the packet to its neighbors.
      ii. Else, it drops the packet.
      b. If it has received from a node in a different grid, it strips off the node list, adds all its neighbors to the node list and broadcasts the packet.

This method uses the concept of non-uniform virtual grids to partition the sensor nodes into groups of gateway nodes and internal nodes. It exploits the location information available to sensor nodes to prolong the lifetime of sensor network by reducing the redundant transmissions that are inherent in flooding.

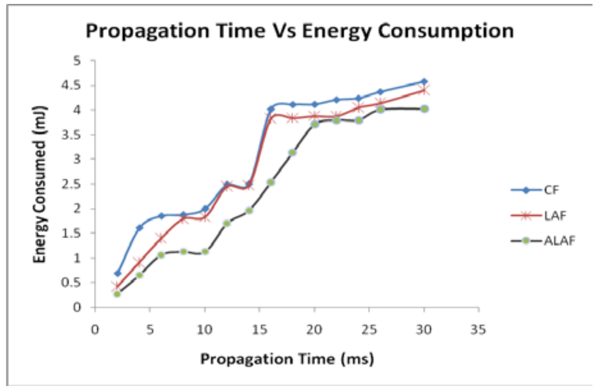


Fig 4. Propagation Time vs energy consumption in ALAF with other flooding techniques

A Graph was plotted for the propagation time versus energy consumed . the above results states the energy consumed by ALAF was minimum than the other flooding schemes. However the non uniform grids some time causes non association of sensor nodes to any of the virtual grids, which in turn cause the lack of information transfer of those sensor nodes. The non uniform grids formed in the network created more virtual sub spaces, whose area cannot be monitored. The unmonitored subspace causes loss in share of information and increased energy consumption of the overall sensor network. These problems are addressed in the work Diagonal-arc based ALAF (DALAF) model, which takes into account of available energy resources. The need for an energy-efficient flooding mechanism for information dissemination in distributed sensor networks is realized with DALAF.

ALAF[] must make very conservative connectivity assumptions because it guesses at connectivity[] (based on a radio model) instead of directly measuring it. Being conservative requires more nodes to stay active than necessary, leading to less energy conservation. Therefore, a need of Flooding mechanism for energy conservation without relying on location information arises This motivates Diagonal based Energy Conservation (DALAF), which, unlike ALAF, does not rely on location information. Further, DALAF itself directly and adaptively measures network connectivity and thus can find network redundancy more accurately so that more energy can be conserved by nullifying subspace and initiate switch off the redundant node radio signal to reduce redundant transmissions, thereby saving energy.

4. Diagonal Arc based Adaptive Location Aided Flooding

This paper has been motivated to develop a Diagonal based Energy Conservation flooding (DALAF) mechanism for energy conservation. DALAF identifies network redundancy and measures network mobility more accurately so that more energy can be conserved by nullifying subspace and initiate switch off the redundant node radio signal. The steps involved in DALAF algorithm is as discussed below.

4.1 Determining Network Redundancy

DALAF organizes nodes into overlapping clusters that are interconnected to each other as shown in Figure 2. A cluster is defined as a subset of nodes that are mutually "reachable" in at most 2 hops. A cluster can be viewed as a circle around the cluster-head with the radius equal to the radio transmission range of the cluster-head. Each cluster is identified by one cluster-head, a node that can reach all nodes in the cluster in 1 hop. A gateway is a node that is a member of more than one cluster. The gateway nodes connect all clusters together to ensure overall network connectivity. A node is ordinary if it is neither a cluster head nor a gateway node and is thus redundant.

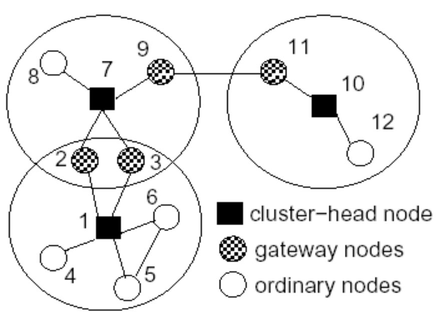


Figure 5: Example of DALAF cluster formation.

4.2 Distributed Cluster Formation

In order to elect cluster-heads and gateway nodes, each node periodically broadcasts a discovery message that contains its node ID, its cluster ID, and its estimated lifetime. The circle around the cluster-head indicates the radio transmission range. A node's estimated lifetime can be conservatively set by assuming the node will constantly consume energy at a maximum rate until it runs out of energy. While forming clusters, DALAF first elects cluster heads, then elects gateways to connect clusters

4.3 Cluster -head Selection

A node selects itself as a cluster-head if it has the longest lifetime of all its neighbor nodes, breaking ties by node ID. Each node can independently make this decision based on exchanged discovery messages. Each node sets its cluster ID to be the node ID of its cluster-head.

4.4 Gateway Node Selection

Among the gateway nodes, those nodes that can hear multiple cluster-heads are primary gateway nodes and those that can hear a combination of cluster heads and primary gateway nodes are secondary gateway nodes. When multiple gateway nodes exist between two adjacent clusters, DALAF suppresses some of them in order to conserve energy since these gateway nodes are redundant. Gateway selection is determined by several rules. First, primary gateway nodes have higher priority than secondary gateway nodes since at least two secondary gateway nodes, instead of just one primary gateway node, are needed to connect adjacent clusters. Second, gateway nodes with more cluster-head neighbors have higher priority, since this will require fewer nodes to be kept awake. Third, gateway nodes with longer lifetimes have higher priority in order to balance node energy. Note that the gateway selection algorithm does not guarantee that only one or one pair of gateway nodes exists between adjacent clusters. In order to support gateway selection, DALAF extends the basic discovery message to include the IDs of the clusters that a gateway node can connect.

Figure 5 shows an example of DALAF cluster formation in which all nodes have the same estimated network operational lifetime. Nodes 1 and 10 can directly decide they are the cluster-heads because they have the lowest ID of all of their neighbors. Node 7 becomes a cluster-head after nodes 2 and 3 choose node 1 as their cluster-head. Nodes 2 and 3 are primary gateway nodes because they are neighbors of two cluster-heads: nodes 1 and 7. Note that one of nodes 2 and 3 is redundant. Nodes 9 and 11 are secondary gateway nodes between clusters 7 and 10.

4.5 Adapting to Network Mobility

With only a subset of the nodes active, it is possible that network mobility could cause a loss of connectivity. If a cluster-head moves then it might no longer be able to serve as a cluster-head. DALAF uses mobility prediction in order to maintain network connectivity.

By estimating how soon a cluster-head will leave its current cluster and inform all nodes in the cluster of that time, the clustered nodes can power themselves on before the cluster-head leaves its cluster. This time is estimated as R/s where s is the cluster-head's current speed and R is its radio transmission range.

Suppose if the R/s estimate is too large, the connectivity between the moving cluster-head and some nodes might be lost before this time. However, if this estimate is too small, DALAF will not be able to conserve any energy. In our DALAF implementation, we set the estimate as R/4s to balance energy conservation and connectivity. We extend the basic discovery message to include the predicted cluster-leaving time. All nodes in a cluster should wake up to reconfigure clusters before the shorter of Ts and the cluster-leaving time of its current cluster-head. The cluster-leaving time estimate is used analogously in the gateway node selection process.

ALAF uses a similar method for dealing with mobility but it anticipates hand-offs by using location information, whereas DALAF uses only local measurements. With such global information, ALAF may have more accurate mobility predications, but DALAF is more practical and localized in nature. DALAF algorithm used to achieve energy conservation has been summarized as follows:

DALAF: ALGORITHM

Step 1: Divide the sensor field into cone shaped grids based on topography and
grid size.
Step 2: For each node,
a. Based on its location in one of the grids formed, set the grid id and maintain the node count in each grid.
b. If the node count of the grid exceeds threshold value,
i. Get the dimensions of the grid and the new grid size.
ii. Split the grid based on the input parameters.
iii. Maintain the count of number of grids formed.
c. Set new grid id for the node based on its location in one of the new grids formed.
d. Identify the neighbors and store in the neighbor list.
e. Classify the node
i. If all of its neighbors belong to the same grid as that of the node, then set its node type as internal.
ii. If any of its neighbors belong to a different grid than that of the node, then set its node type as gateway.

Step 3: The source node creates a packet with its id, sequence number, node list, grid id, node type, node list length. The node list contains its id and the ids of all its neighbors.
Step 4: The source node broadcasts the packet to its neighbors.
Step 5: The receiving node checks to see whether the packet's sequence number is already in its received list of sequence numbers.
c. If so, it drops the packet.
d. Else, it stores the sequence number of the packet it receives and sets its id, grid id, node type in the packet

	CF in (mj)	LAF in (mj)	ALAF in(mj)	DALAF in(mj)	DALAF Single Path	DALAF Multi Path
2	0.068	0.423	0.282	0.221	0.221	0.156
4	1.608	0.915	0.66	0.486	0.486	0.296
6	1.849	1.402	1.07	0.822	0.822	0.524
8	1.874	1.795	1.129	1.093	1.093	0.636
10	1.996	1.834	1.131	1.146	1.146	0.951
12	2.486	2.455	1.706	1.626	1.626	1.263
14	2.498	2.475	1.971	1.79	1.79	1.334
16	4.013	3.826	2.542	2.034	2.034	1.517
18	4.106	3.836	3.144	2.765	2.765	1.639
20	4.11	3.876	3.714	3.602	3.602	1.944
22	4.204	3.878	3.801	3.771	3.771	2.249
24	4.235	4.05	3.802	3.792	3.792	2.553
26	4.241	4.05	4.01	3.986	3.989	2.98
28	4.365	4.14	4.012	4.007	4.007	3.468
30	4.572	4.404	4.028	4.015	4.015	3.773

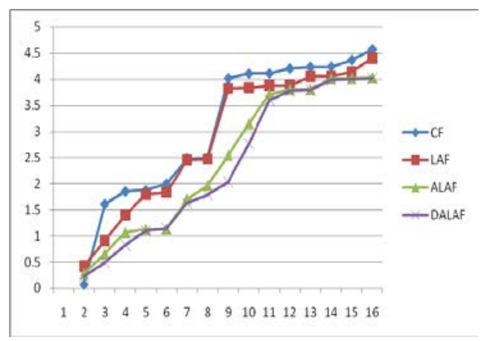


Figure 6. Propogation Time Vs. Energy Consumption

DALAF algorithm which has been reported above has been tested and the evaluation of its performance has been reported in the subsequent section.

5. Results and Findings

We have developed a simulation environment in NS2 to evaluate the performance of DALAF and compared it with ALAF algorithm for energy conservation. It was found that DALAF protocol achieves higher with reliable energy savings when compared with classical flooding, LAF and ALAF. We also found that the nodes with a higher degree (i.e., nodes with more one-hop neighbors) disseminate more data per unit energy in both LAF and modified flooding compared to classical flooding. Thus, dense sensor networks are likely to benefit more from using the DALAF protocol for data dissemination in terms of energy savi

6. Conclusion and Future Directions

Performance analysis in terms of energy consumption is taken and compared for the Classical flooding (CF), Location aided flooding(LAF),Adaptive location aided flooding(ALAF) and Diagonal Arc based Adaptive Location Aided Flooding with reliability,(DALAF) is presented . This mechanism DALAF is capable of measuring reliable network mobility and network redundancy more accurately so that more energy can be conserved by nullifying subspace and initiate switch off the redundant node radio signal.

References:

[1] I.Akyildiz, W.Su, Y.Sankarasubramaniam, and E.Cayirci, "A survey on Sensor Networks," *IEEE Communications Magazine*, vol. 40, Issue:8, pp.102-114, August 2002.

[2] Harshavardhan Sabbineni and Krishnendu Chakrabarty, "Location-Aided Flooding: An Energy-Efficient Data Dissemination Protocol for Wireless Sensor Networks," *IEEE transactions on computers,* vol. 54, no. 1, Jan 2005, pp.36-46.

[3] Nalini.L. and Uma.G.V., "An Energy Efficient Adaptive Location Aided Flooding Protocol for Wireless Sensor Networks" Asian Journal of Information Technology 5(12):1485-1491,2006,Medwell Online ,2006

[4] Nalini.L.,Shanmugavel.S " Conservation of energy in Wireless Sensor Networks using Diagonal arc based Adaptive Location Aided Flooding Mechanism"- European Journal of Scientific Research ISSN 1450-216X vol.51.No.1(2011)PP 44-52

[5] Z. Haas, J. Halpern, and L. Li, "Gossip Based Ad Hoc Routing," IEEE Infocom Conf., pp. 1702-1706, 2002.

[6] J. Jetcheva, Y. Hu, D. Maltz, and D. Johnson, "A Simple Protocol for Multicast and Broadcast in Wireless Ad Hoc Networks," Internet Draft: draft-ietf-manet-simple-mbcast-01.txt, July 2001

[7] J. Kulik, W.R. Heinzelman, and H.Balakrishnan, "Negotiation- Based Protocols for Disseminating Information in Wireless Sensor Networks," Wireless Networks, vol. 8, pp. 169-185, 2002.

[8] K. Whitehouse and D. Culler, "Calibration as Parameter Estimation in Sensor Networks," ACM Int'l Workshop Sensor Networks and Applications, pp. 59-67, 2002

[9] R. Williams, The Geometrical Foundation of Natural Structure: A Source Book of Design. New York**:** Dove Publications, 1979.

[10] J. Wu and F. Dai, "Broadcasting in Ad Hoc Networks Based on Self-Pruning," Proc. Infocom, pp. 2240-2250,2003

[11] Jie Lian, Kshirasagar Naik and Gordon B. Agnew(2006), Modelling and Enhancing the Data Capacity of Wireless Sensor Networks

[12] K.P. Birman, M. Hayden, O. Ozkasap, Z. Xiao, S. Ni, Y. Tseng, Y.Chen, and J. Sheu, "The Broadcast Storm Problem in a Mobile AdHoc Network," Proc. ACM/IEEE Int'l Conf. Mobile Computing andNetworking, pp. 151-162, Aug. 1999.

An Efficient Secure Biometric System with Non-Invertible Gabor Transform

Radha Narayanan[1] and Kathikeyan Subramanian[2]

[1] **Department of Computer Science, Karpagam University Coimbatore,641 021,Tamil Nadu**

[2] **Department of Information Technology,College of Applied Sciences, Sohar, Sultanate of Oman**

Abstract

Biometric scheme are being widely employed because their security merits over the earlier authentication system based on records that can be easily lost, guessed or forged. High scale employments and the related template storage have increased the requirement to guard the biometric data stored in the system. Theft of biometric information is a negotiation of the user's privacy. In Addition, the stolen biometric information can be used to access other biometric systems that have the similar feature provided for the user. Several alternative functions have been identified in literature for creating revocable or non-invertible biometric templates. Although, their security examination either disregards the distribution of biometric features or uses inefficient feature matching. This generally shows the way to unrealistic approximation of security. In this paper a novel approach for the Non-Invertible biometric system is proposed to secure the biometric template. Security of a feature transformation method can be assessed according to the two main factors: i) non-invertibility, and ii) diversity. Non-invertibility represents the complexity in obtaining the original biometric when the secure template is provided and diversity represents the complexity in guessing one secure template when a different secure template created from the identical biometric is provided. The proposed Non-invertible Gabor transform possess both the non invertible and diversity features which enhances the security of the system to a large extent. The proposed approach is very much resistant to minor translation error and rotation distortion. The experimental result shows the better performance of the proposed technique compared to the existing system.

Keywords: *Non-Invertible Gabor Transform, Minutiae, Minucode*

1. Introduction

consistent identity management technique is widely required in order to contest the increased growth in identity theft and to satisfy the increased security needs in a many applications ranging from international border crossings to securing data in databases. Finding the uniqueness of a person is an important process in any identity managing system. Proxy representations of uniqueness like passwords and ID cards are not adequate for trustworthy identity determination since they can be effortlessly misplaced, shared or stolen. Biometric recognition [11, 12] is the science of establishing the identity of an individual with the help of the anatomical and behavioral features. Commonly used biometric features are fingerprint, face, iris, hand geometry, voice, palmprint, handwritten signatures and gait. Biometric features [8] have a number of attractive properties with respect to their use as an authentication token [16], viz., reliability, convenience, universality etc. These features have lead to the wide usage of biometric authentication technique. But there are still some subjects regarding the security of biometric recognition systems that required to be addressed for ensuring the integrity and public acceptance of these systems.

There are five main elements in a general biometric authentication system such as sensor, feature extractor, template database, matcher and decision module. Sensor is the interface between the user and the authentication system and it is used to scan the biometric feature of the individual. Feature extraction phase processes the scanned biometric trait to dig out the significant data (feature set) that is helpful in differentiation of various users. Sometimes the feature extractor is performed after the quality assessment phase that judges whether the scanned biometric trait is of adequate quality for future processing. When the enrollment phase is performed, the extracted feature set is stored in a database as a template (X_T) [6] along with the user's individuality data. As the template database can be physically distributed and enclose millions of records, preserving its security is not a minor work. The matcher module is generally an executable program that accepts two biometric feature sets X_T and X_Q (from template and query, respectively) as inputs and outputs a match score (S) which represents the matching between the two sets. At last, the decision phase suggests the identity conclusion and reply to the query.

Because of the fast growth in sensing and computing techniques, biometric systems have evolved to be reasonable and are effortlessly embedded in a various consumer devices and making this technology weaker for the malicious designs of terrorists and thefts. To avoid any possible security breaks, vulnerabilities of the biometric system must be found and addressed thoroughly. Several studies have analyzed possible security violations in a biometric system and existing techniques to handle those violations. Recognized techniques of vulnerability analysis like attack trees have also been helpful in examining how biometric system security can be compromised.

This paper provides better solution to this problem. In this paper, initially the features are extracted from the fingerprint. Then the secure template is obtained from these features with the help of non-invertible transform called Non-Invertible Gabor Transform [9, 10]. Finally the anonymous matching is performed to identify the authenticated user.

2. Related Work

An Efficient Approach for Non-Invertible Cryptographic Key Generation from Cancelable Fingerprint Biometrics is proposed by Lalithamani *et al.,* [1]. The main drawback of the conventional cryptographic algorithms is the preservation of their key's secrecy. Incorporating the users' biometric features in the generation of strong and repeatable cryptographic keys has gained huge popularity among researchers. The randomness of the user's biometric features, incorporated into the generated cryptographic key, makes the key in order that it cannot be cracked by the attacker lacking noteworthy information of the user's biometrics. However, if a person's biometric is once missed, it will be helpful for the attackers everlastingly as it is naturally belong to the user. To deal with this problem, cancelable biometrics can be used as an effective answer for canceling and re-issuing biometric templates. Here, this paper presents an innovative and efficient technique to create a non-invertible cryptographic key from cancelable fingerprint templates. In the beginning, a one-way transformation is applied on the minutiae points obtained from the fingerprints to accomplish a set of transformed points. Consequently the transformed points are made use of to produce cancelable templates [14, 15]. The cancelable fingerprint templates [7] are then used to create a unique non-invertible key.

Chulhan *et al.,* [2] presents an Alignment-Free Cancelable Fingerprint Templates based on Local Minutiae Information. For every minutia, a rotation and translation invariant value is calculated from the orientation data of neighboring local regions around the minutia [13]. The invariant value is utilized as the input to two changing functions that output two values for the translational and rotational movements of the original minutia correspondingly, in the cancelable template. If a template is compromised, it is modified by a new template generated by various changing functions. The presented technique conserves the original geometric relationships among the enrolled and query templates after their transformation. As a result the transformed templates can be used to identify a person without requiring alignment of the input fingerprint images.

Takahashi *et al.,* [3] put forth a new technique for creating cancelable fingerprint templates with verifiable security based on the well-known chip matching algorithm for fingerprint verification and correlation-invariant random filtering for transforming templates. Ratha *et al.,* [4] demonstrate different techniques to create multiple cancelable identifiers from fingerprint images. A client is given as much biometric identifiers as required by issuing a new transformation key. The identifiers can be cancelled and replaced when cracked by attackers. The performance of various techniques like Cartesian, polar, and surface folding transformations of the minutiae positions are compared.

Huijuan *et al.,* [5] proposed a non-invertible transform approach to perpendicularly project the distances between a pair of minutiae to a circle to create the features. To attain better result, other local features like relative angles between the minutiae pair, and global features like orientation, ridge frequency and total number of minutiae of the randomly sampled blocks around each minutia are also used. The cancelable templates are eventually created by the proposed ldquobin-based quantization (BQ)rdquo. Both feature extraction and cancelable template generation are administrated by a secret key to assure revocability and security.

3. Methodology

This section provides the detailed description of the proposed technique based on the system model shown in Figure1. The proposed Biomapping technique consists of three fundamental procedures: fingerprint feature extraction, noninvertible transform and anonymous matching.

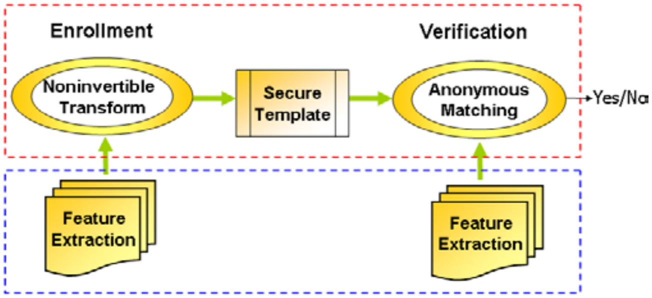


Fig. 1System model for privacy-preserving Biometric Recognition

3.1 Fingerprint feature extraction algorithm

The majority of the existing techniques are based on reference core points of a fingerprint, but, exact evaluation continues to be a complex difficulty and mistakes in the reference core points may led to false reject. To overcome this problem, a minutiae-centered region encoding, MinuCode is used to neglect the reference core point determination and handle with the noise of fingerprint biometrics. The location and orientation attributes of a minutia are used together and these attributes are considered as a 3-tuple (x, y, θ).

Initially, a circular region R is built around every minutia with the similar radius. For every region, the center minutia is taken as the core minutia, and the other minutiae as the neighbor minutiae. Then every neighbor minutia will be transformed into the polar coordinate scheme in accordance with the respective core minutia, and is mentioned as a new 3-tuple (ρ, α, β), where ρ and α represents the radial distance and radial angle respectively, and β indicates the orientation of the neighbor minutia in accordance with the core minutia, α, β∈ [1,360]. An example is provided in Figure 2 (a).

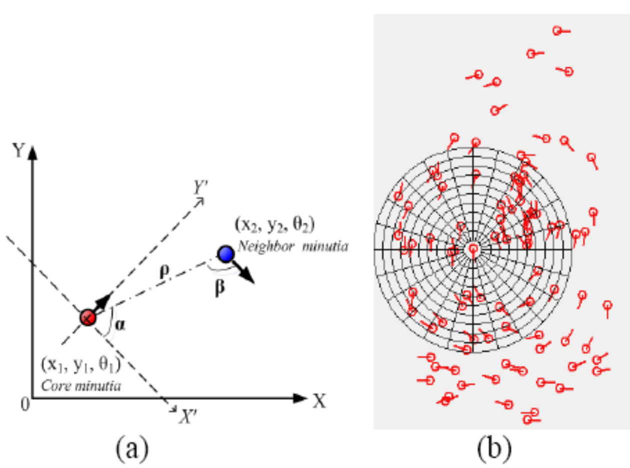


Fig. 2 Illustration for minutiae-centered region encoding, MinuCode

Next, a tessellation quantification is performed on every one of the neighbor minutiae by tessellate the region of interest middle at the core minutia. After that the 3-tuple (ρ, α, β) in the polar coordinate will be quantified into a rougher 3-tuple T (b, a, o).

$$\begin{cases} b = \left\lfloor \dfrac{\rho}{db} \right\rfloor \\ a = \left\lfloor \dfrac{\alpha}{da} \right\rfloor \\ o = \left\lfloor \dfrac{\beta}{do} \right\rfloor \end{cases} \quad (1)$$

Where $\lfloor / \rfloor$ is an operator to compute the quotient.

In the above equation, the parameter db represents the bandwidth of the region tessellation, da represents the distortion tolerable difference of radial angle, and do represents the distortion tolerable difference of the orientation of the neighbor minutia in accordance with the core minutia. Figure 2(b) illustrates an example of tessellation quantification in a fingerprint. The i^{th} neighbor minutia in accordance with the core minutia in the special region is therefore indicated as the 3-tuple $T_i=(b_i, a_i, o_i)$. If there are m neighbor minutiae in a region R, then, R can be indicated as a set of T:

$$M = < T_1, T_2 \ldots . T_m > \quad (2)$$

where the set M is called MinuCode.

Then, if there are N minutiae in a fingerprint, the original fingerprint template can be a gathering of MinuCodes $\{M_1, M_2,\ldots, M_N\}$. This gathering is the outcome of the feature extraction process.

3.2 Non-Invertible Transform Algorithms

This approach integrates the feature extraction and the noninvertible transform. This approach uses the Non-invertible Gabor transform.

3.3 Non-Invertible Construction using Non-Invertible Gabor Transform

The feature extraction points of the fingerprint obtained are used for the Non-Invertible construction of the Gabor transform.

The MinuCodes $\{M_1, M_2\ldots M_N\}$ obtained are supplied to the non invertible Gabor transform in which the translation and modulation operations are performed initially.

The primary building blocks of the Gabor representation are the so-called Gabor systems. To describe a Gabor system, let a > 0 and b > 0 be such that ab = q/p with p and q relatively prime, and let T_{ak} and M_{bl}, for k, l ∈ $\mathbb{Z}$, be the translation and modulation operators represented by

$$T_{ak} f(t) = f(t - ak) \quad (3)$$

$$M_{bl} f(t) = e^{2\pi i b l t} f(t) \quad (4)$$

For $s \in L^2(\mathbb{R})$, the Gabor system $\mathcal{G}(s,a,b)$ is a collection of $\{M_{bl}T_{ak}s(t); (k,l) \in \mathbb{Z}^2$. The composition

$$M_{bl}T_{ak}f(t) = e^{2\pi ibl t}f(t-ak) \quad (5)$$

Which is a unitary operator, is called a time-frequency shift operator. Several technical details in time-frequency analysis are connected to the commutation law of the translation and modulation operators, namely

$$M_{bl}T_{ak} = e^{2\pi iabkl}T_{ak}M_{bl} \quad (6)$$

When p = 1, the time-frequency shift operators commute, i.e. $M_{bl}T_{ak} = T_{ak}M_{bl}$, since $e^{2\pi iabkl} = 1$ for all K, $l \in \mathbb{Z}$. One outcome of the commutation rule that is used in the exposition is the relation

$$< f, M_{bl-bn}T_{ak-am}f > = e^{2\pi iab(l-n)m} < M_{bn}T_{am}f, M_{bl}T_{ak}f > \quad (7)$$

When p = 1 this becomes

$$< f, M_{bl-bn}T_{ak-am}f > = < M_{bn}T_{am}f, M_{bl}T_{ak}f > \quad (8)$$

For $s \in L^2(\mathbb{R})$, the collection $\mathcal{G}(s,a,b)$is a Riesz basis for its closed linear span if there exist bounds A > 0 and B < 1 such that

$$A\|c\|_{\ell^2}^2 \le \left\|\sum_{k,l\in\mathbb{Z}} c_{k,l}M_{bl}T_{ak}s\right\|^2 \le B\|c\|_{\ell^2}^2 \qquad c \in \ell^2 \quad (9)$$

and is a frame when

$$A\|f\|^2 \le \sum_{k,l\in\mathbb{Z}} |< f, M_{bl}T_{ak}s|^2 \le B\|f\|^2 \qquad for\ all \quad f \in \overline{span}\ \{M_{bl}T_{ak}s\} \quad (10)$$

A required condition for $\mathcal{G}(s,a,b)$to comprise a frame for $L^2(\mathbb{R})$ is that ab≤1. In addition if $\mathcal{G}(s,a,b)$is a frame, then it is a Riesz basis for $L^2(\mathbb{R})$ if and only if ab = 1. This paper focus on the regime ab≥1, where $\mathcal{G}(s,a,b)$does not necessarily span $L^2(\mathbb{R})$.

With a Gabor system $\mathcal{G}(s,a,b)$ this paper associate a synthesis operator (or reconstruction operator) $S : \ell^2(\mathbb{Z}^2) \rightarrow L^2(\mathbb{R})$, defined as

$$S_c = \sum_{k,l\in\mathbb{Z}} c_{k,l}M_{bl}T_{ak}s \quad (11)$$

$$S_c = \sum_{k,l\in\mathbb{Z}} c_{k,l}M_{bl}T_{ak}s(t) \qquad for\ every\ c \in l^2(\mathbb{Z}^2) \quad (12)$$

The conjugate $S^*: L^2(\mathbb{R}) \rightarrow l^2(\mathbb{Z}^2)$ of S is known as the analysis operator (or sampling operator), and is given by

$$S^*f = \{\langle f, M_{bl}T_{ak}s\rangle\} \qquad for\ every\ f \in L^2(\mathbb{R}) \quad (13)$$

The Gabor representation of a feature contains the set of coefficients $\{C_{k,l}\}_{k,l\in\mathbb{Z}}$ acquired by inner products with the elements of some Gabor system $\mathcal{G}(s,a,b)$.

$$C_{k,l} = \langle f, M_{bl}T_{ak}S\rangle \quad (14)$$

This procedure can be indicated as an analysis filter-bank. Therefore, s(t) is referred to as the analysis window of the transform. If $\mathcal{G}(s,a,b)$constitutes a frame or Riesz basis for $L^2(\mathbb{R})$, then there exists a function $\upsilon(t) \in L^2(\mathbb{R})$such that any $f(t) \in L^2(\mathbb{R})$ can be reconstructed from the coefficients $\{C_{k,l}\}_{k,l\in\mathbb{Z}}$ using the formula

$$f(t) = \sum_{k,l\in\mathbb{Z}} C_{k,l}M_{bl}T_{ak}\upsilon(t) \quad (15)$$

In this paper, Gabor system is consider that do not essentially span $L^2(\mathbb{R})$ but rather only a (Gabor) subspace. A Gabor space is the set V of all feature that can be represented with some norm-bounded sequence $c_{k,l}$. As perfect recovery cannot be assured for each signal in $L^2(\mathbb{R})$in these cases, the choice of selecting the analysis and synthesis windows based on the implementation constraints. Conversely, for the purpose of analysis and synthesis processes to be stable, it is assured that the systems $\mathcal{G}(s,a,b)$and $\mathcal{G}(\upsilon,a,b)$form frames or Riesz bases for their span.

For tractability, it is assumed that, a and b are positive constants such that ab = q/p, where p and q are relatively prime. In addition, only Gabor spaces are considered whose generators are obtained from the so-called Feichtinger algebra S_0, which is defined by

$$S_0 = \left\{f \in L^2(\mathbb{R}) \middle| \|f\|S_0 := \int |\langle f, M_\omega T_x\psi\rangle| dx\, d\omega < \infty\right\} \quad (16)$$

where $\psi(t)$ represents the Gaussian window. An important possessions of S_0 is that if v(t) and s(t) are elements from S_0 then $\{\langle v, M_{bl}T_{ak}s\rangle\}_{k,l\in\mathbb{Z}}$ is an $\ell^1(\mathbb{Z}^2)$ sequence. Examples of functions in S_0 are the Gaussian and B-splines of strictly positive order. The Feichtinger algebra is an tremendously helpful space of good window functions in the sense of time-frequency localization.

The final secure template of a fingerprint will be a collection of transformed MinuCodes {M'$_1$, M'$_2$,…, M'$_N$}. N is the number of minutiae in the fingerprint.

3.4 Anonymous matching algorithm

The minutia point arrangement process can be divided into two phase. Initially, Biomapping handles the alignment errors in the minutia location and orientation with the help of minutiae-centered quotient computation and tessellation quantification: provided a transformed MinuCode from the secure template and an original MinuCode from query fingerprint, two neighbor minutiae are considered to be matched only if the MinuCode identified after Non-Invertible Gabor Transform are identical.

In the next phase, the threshold mechanisms are used to deal with the replacement errors that carry with the elimination and adding of a few minutiae. Two thresholds tm and tp are used for the anonymous matching: two regions are similar if there are minimum of tm equal neighbor minutiae. At last, two fingerprints are considered to be matched if there are at least tp equal regions.

4. Experimental Result

The experimental result for the proposed biometric system is presented in this section. The experiments are conducted with the help of public domain FVC2002-DB2 database. The database contains 100 fingers, each one has 8 impressions. The first two impressions of every finger are used in this experimentation. As a result, the number of genuine attempts is 100, and the number of the impostor attempts is 9900.

Table 1: FRR and FAR Resulted for the Existing and Proposed Biometric System

Region Threshold (tp)	Existing		Proposed	
	FRR	FAR	FRR	FAR
0	0	98.3	0	80.6
5	4.2	20.5	0	5.2
10	12.1	10.2	4.5	0
15	28.9	4.2	10.2	0
20	40.3	1.1	20.5	0
25	52.1	0	30.9	0
30	60.7	0	38.1	0
35	74.5	0	42.6	0
40	80.3	0	49.9	0
45	88.1	0	55.2	0

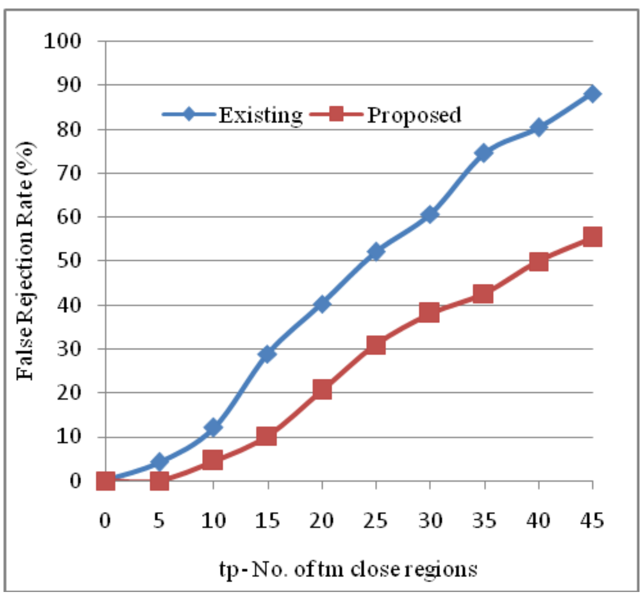


Fig. 3 False Rejection Rate for Different Threshold

Table 1 provides the obtained False Rejection Rate (FRR) and False Acceptance Rate (FAR) for the various region thresholds. It can be clearly observed from the table that initially for the region threshold of 0, the FRR is 0 for both techniques. When the threshold value is increased, the FRR will gradually increase. When the threshold is 15, FRR is 28.9 for the existing method, whereas it is only 10.2 for the proposed technique. When the threshold is set as 45, FRR is 88.1 for the existing method, whereas it is only 55.2 for the proposed technique which is a huge difference. When FAR is considered, the proposed technique will have 0 FAR when the threshold is set above 10, whereas for the existing technique the threshold should be set above 20.

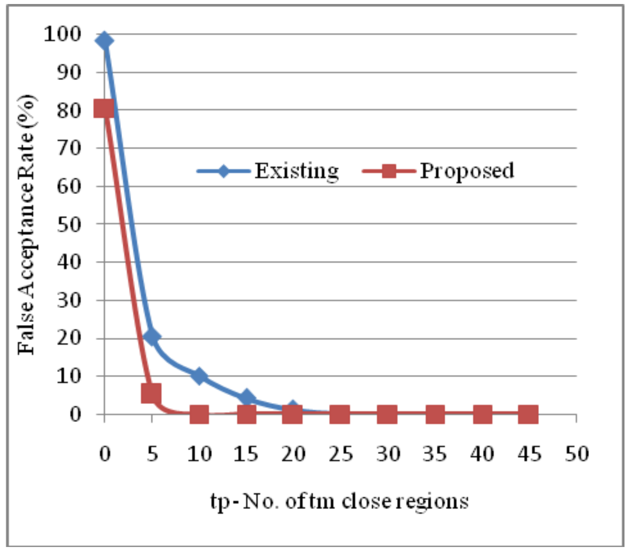


Fig. 4 False Acceptance Rate for Different Threshold

Fig. 3 and Fig. 4 shows the variations in FRR and FAR for various region thresholds respectively. It can be clearly observed that the proposed technique will result in lesser FRR and FAR when compared to the existing biometric system.

5. Conclusion

Biometrics provides easy usability over earlier token and password based authentication techniques, but raises privacy and security concerns. The earlier authentication techniques like credit cards and passwords can be revoked or substituted if it is not handled properly but biometrics is everlastingly linked with a user and cannot be substituted. Cancelable biometrics is helpful to overcome this by generating revocable biometric templates. In this paper, the secure template is obtained from these features with the help of non-invertible transform called Non-Invertible Gabor Transform. The experimental result shows that the proposed biometric system results in better FRR and FAR which in turn indicates the better accuracy and security than the existing biometric system.

This biometric system thus provides a good improvement over the other techniques which are based purely on biometric feature extraction and complex classifier. This approach can be enhanced to higher level in order to further improve the security. This work can be extended in future by using the highly secure non invertible transform such as Baker Non-Invertible Transform. The usage of this Baker Non-Invertible Transform will result in strong mixing of feature points and it will be very hard to break by the attackers and thus it can provide better security.

Moreover, the future enhancement will address the practical issues of the latency for recognition and the size of the secure template.

References

[1] Chulhan Lee, Jeung-Yoon Choi, Kar-Ann Toh and Sangyoun Lee, "Alignment-Free Cancelable Fingerprint Templates Based on Local Minutiae Information", IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, Vol. 37, No. 4, 2007.

[2] Ratha, N.K., Chikkerur, S., Connell, J.H. and Bolle, R.M., "Generating Cancelable Fingerprint Templates", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 29, No. 4, Pp.561-572, 2007.

[3] Jain, A.K., Nandakumar, K. and Nagar, A., "Biometric Template Security", EURASIP Journal on Advances in Signal Processing, Special Issue on ASPPRMB, Vol. 2008, Pp.1-17, 2008.

[4] Andrew Teoh, B.J. and David Ngo, C.L., "Cancellable Biometrics Featuring with Tokenised Random Number", Pattern Recognition Letter, Vol. 26, No. 10, Pp. 1454–1460, 2005.

[5] Daugman, J.G., "Complete Discrete 2D Gabor Transform by Neural Networks for Image Analysis and Compression", IEEE Transaction on Acoustic Speech and Signal Processing, Vol. 36, No. 7, Pp. 1169-1179, 1988.

[6] Dunn, D., Higgins, W. E., "Optimal Gabor Filters for Texture Segmentation," IEEE Transaction on Image Processing, Vol. 4, No. 7, Jul. 1995.

[7] Jain, A.K., Ross, A. and Pankanti S., "Biometrics: A Tool for Information Security," IEEE Transactions on Information Forensics and Security, vol. 1, no. 2, pp. 125–143, 2006.

[8] Ross, A.K., Shah, J. and Jain, A.K., "From Templates to Images: Reconstructing Fingerprints from Minutiae Points," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 29, No. 4, Pp. 544–560, 2007.

[9] Cappelli, R., Lumini, A., Maio, D. and Maltoni, D., "Fingerprint Image Reconstruction from Standard Templates," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 29, No. 9, Pp. 1489–1503, 2007.

[10] Lalithamani, N. and Soman, K.P., "An Efficient Approach for Non-Invertible Cryptographic Key Generation from Cancelable Fingerprint Biometrics", International Conference on Advances in Recent Technologies in Communication and Computing, Pp. 47-52, 2009.

[11] Takahashi, K. and Hitachi, S.H.,"Generating Provably Secure Cancelable Fingerprint Templates based on Correlation-Invariant Random Filtering", IEEE 3rd International Conference on Theory, Applications, and Systems, Pp. 1-6, 2009.

[12] Huijuan Yang, Xudong Jiang and Kot, A.C., "Generating Secure Cancelable Fingerprint Templates Using Local and Global Features", 2nd IEEE International Conference on Computer Science and Information Technology, Pp. 645-649, 2009.

[13] Weicheng, S., Surette, M. and Khanna, R., "Evaluation of Automated Biometrics-Based Identification and Verification Systems", Proceedings of the IEEE, Vol. 85, No.9, Pp. 1464-1478, 1997.

[14] Jain, A.K., Ross, A. and Uludag, U., "Biometric Template Security: Challenges and Solutions," Proceedings of European Signal Processing Conference (EUSIPCO), 2005.

[15]Boult, T.E., Scheirer, W.J. and Woodworth, R., "Fingerprint Revocable Biotokens: Accuracy and Security Analysis," Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Pp. 1–8, 2007.

[16]Ratha, N.K., Connell. J.H. and Bolle, R.M., "An Analysis of Minutiae Matching Strength," Proceedings of Third International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA), 2001, Pp. 223–228, 2001.

First Author

Mrs.N.Radha received M.Sc(Computer Science) in Madurai Kamaraj University and completed M.Phil (Computer Science in bharathiar University,India. Currently she is pursuing her Ph.D. (Computer Science) in Karpagam University, India. She had 14 years of teaching experience. At present she is working in PSGR Krishnammal College for women. She presented papers in Various International and National Journals. Her areas of interests are Information security, Biometric security and oops. She has published more than 10 papers in various National and International Journals.

Second Author:

Dr.Karthikeyan completed M.Sc (Computer Science) in Bharathidasan university in the year 1996.He has done M.Phil(Computer Science) in Bharathiar University in the year 2003 and he completed Ph.D (Computer science) in Alagappa university in the year 2008.He has 15 years of teaching experience hold various positions . Now he is working as an Assistant Professor in College of Applied Sciences, Sultanate of Oman. His areas of interests are Cryptography and Network Security - Security Protocols, Key Management and Encryption Techniques. He is Guiding 16 M.phil scholars and 6 Ph.D scholars. He has published and presented more than 50 papers in various National, International and conference and journals. He served as Reviewer, editor in Chief, Associate editor, Technical editor, Editorial Board member and Guest editor in various National and International Journals. He got senior Membership in IACSIT, Singapore and ACEEE, India. He got membership in CSTA, New York, USA and in International Association of Engineers – Computer Science/Wireless Networks, Canada/Spain.

High Capacity and Resistance to Additive Noise Audio Steganography Algorithm

Haider Shahadi[1], Razali Jidin[2]

[1] Electronic and Communication Department, University Tenaga National, 43000 kajang, Selangor, Malaysia

[2] Electronic and Communication Department, University Tenaga National, 43000 kajang, Selangor, Malaysia

Abstract

Steganography is the art of message hiding in a cover signal without attracting attention. The requirements of the good steganography algorithm are security, capacity, robustness and imperceptibility, all them are contradictory, therefore, satisfying all together is not easy especially in audio cover signal because human auditory system (HAS) has high sensitivity to audio modification. In this paper, we proposed a high capacity audio steganography algorithm with good resistance to additive noise. The proposed algorithm is based on wavelet packet transform and blocks matching. It has capacity above 35% of the input audio file size with acceptable signal to noise ratio. Also, it is resistance to additive Gaussian noise to about 25 db. Furthermore, the reconstruction of actual secret messages does not require the original cover audio signal.

Keywords: *Audio Steganography, Wavelet Packet Transform, Block Matching, Robust Stego-algorithm, High capacity Steganography.*

1. Introduction

The rapid development in information technology, especially in the field of computer networks and wireless communication has enabled hackers to eavesdrop on communication. Eavesdropping can be prevented using techniques of cryptography or/and steganography [1]. Steganography means embedding information in a cover signal to prevent illegal detection [2]. Steganography hides secret messages, while in cryptography, secret message is visible, therefore unauthorized users can access to the secret encrypted data [3].

Steganography is often used in covert communication such as military and government communication and it requires relatively high payload, while a watermarking which is often used in copyright protection however does not require high capacity [4].

The major requirements that must be satisfied for good steganography algorithms are perceptual transparency, payload or capacity and robustness [5]. Steganography, unlike watermarking, needs high capacity and is considered as an important factor, but robustness should be dominant for watermarking. Higher targets for one of the mentioned requirements will negatively affect the other performances for the same stego-algorithm as they are contradictory according to magic triangle [6].

In recently years many techniques appeared to develop information hiding [4, 7, 8], and most techniques used image and video media but lacking of using audio signal as a cover signal especially in high rate data embedding, most likely due to Human Auditory System (HAS) is more sensitive than Human Visual System (HVS) [9]. Although adopting audio signals as cover signals may yield inferior inaudible performance, there are still suitable features such as transitory and unpredictability that makes sound signal as suitable secure cover signal.

In this paper we proposed a new steganography algorithm that has high capacity and resistance to additive noise characteristics, and it is based on both discrete wavelet packet transform (DWPT) and blocks matching. The proposed algorithm is based on matching between the cover and message blocks. Then it uses blocks scaling and replacement in embedded process to keep power of embedded message high in order to resist the additive noise. The output stego-signal has an acceptable quality with high capacity that can reach more than 35% of the input audio size (about 250 *kb/sec*). The message recovery process does not require original audio cover signal, and also the algorithm provides high security due to the arbitrary distribution of the secret message blocks in embedding process instead a known embedded message distribution in most stego-algorithms.

2. Related Work

Generally audio steganography and watermarking can be classified according to embedding domain either time domain or transform domain. The common and simplest hiding technique in time domain with acceptable capacity is the Least Significant Bits (LSB), but it is vulnerable due to any changes in LSB that can destroy the embedded message [6]. In transform domain, there are many transform methods can be employed in information hiding such as Fourier domain [6, 10], discrete cosine domain [6, 11], and wavelet domain [6, 9, 12, 13]. Each domain has its features in signal processing and in information hiding [14], however, the wavelet domain has a major advantage over the others because it divides a signal into different frequency components with different resolution and each component can be used in embedded process according to its power. DWT behaves similar to the time-frequency characteristics of human ears that has high time and low frequency resolution for high frequencies, and high frequency and low time resolutions for the low frequencies components [15].

In general, Discrete Wavelet Transform (DWT) decomposes a signal at a level of decomposition into two components, high and low frequency components. Most power of the input signal is concentrated in low frequency component or called approximation signal, while little power spread over coefficients of high frequency component or called details signal. The decomposition process starts by decompose a signal into two components, high and low frequency components, then will further decompose each the low frequency component into another two components, and the process repeats for further levels of decomposition [15]. The reconstruction of original signal is performed by the Inverse Discrete Wavelet Transform (IDWT), and it bases on approximation and details signals.

The modification in the details signals has little effect on the reconstructed signal, rather depends on the number of levels of decomposition and which selected details level was modified. However modification in the approximation signal or low frequency component may affect significantly on the reconstructed signal. Therefore using details signals as a cover for information embedded process enable high payload and acceptable quality, when it is used in steganography [9]. However, information embedding in details signal has weakness in robustness for watermarking applications such as visible watermarking of TV channels and copyright protection because it is possible to remove a secret message by signal processing attackers for example by resetting details coefficients. Therefore in watermarking applications, it is better to use low frequency component in embedding process because watermarking does not need high capacity. Watermarking however, requires robust algorithm to embed information as a part of the cover signal, render it almost impossible to be removed by attackers [12, 16].

The LSB is the most common method employed in embedding process in DWT domain [9, 13, and 17]. This method has superior stego-signal quality and capacity, but it is sensitive to additive noise especially when data is embedded in high frequencies components which have low power.

In this paper a high capacity and noise resistance audio steganography algorithm will be described. The algorithm uses Discrete Wavelet Packet Transform (DWPT) to decompose audio cover signal. The DWPT is similar to DWT except decomposition is performed for both high frequency and low frequency components. The decomposed signal by DWPT for selected L levels, yields (2^L) components with equal lengths, but only one represents the approximation signal that has the highest power, while the others are details signals with decreasing power, starting from lowest frequency details component to the highest frequencies details component. Subsequently after several steps, the Inverse of Discrete Wavelet Packet Transform (IDWPT) is used to reconstruct the original cover signal.

DWPT decomposition provides two benefits, first, separation the signal components according to their frequency and power, and second, making all the components in the same length and that will facilitate the task of block matching which to be used in our proposed scheme. The proposed scheme however does not use the approximation signal in embedding process to maintain the quality of output of stego-signal. The lowest frequency component of the details signal which has highest power with respect to other details signal is used to embed main key that is generated in matching process. The remaining $(2^L - 2)$ details signals are arranged as blocks and they are used in message embedding process after block matching.

The proposed block matching is based on minimum distance between blocks of secret message after primary scaling. The primary scaling will make the amplitudes of the message signal within the range of the amplitudes of the cover signal. The block matching output is used to generate embedding keys that will be used later in both embedding and extraction processes. The embedding key is encrypted using stream cipher and then it is embedded in lowest frequency details signal as shown in section (3.5).

The proposed algorithm uses block replacement after matching and scaling in the embedding process instead of LSB to satisfy three contradictory things simultaneously, firstly, to maintain message power high to resist the additive noise, second, to keep quality of modified cover signal closes to original input signal, and thirdly, to provide high rate of payload instead of less bits in LSB method. Another advantage for this algorithm over most algorithms is that it cannot estimate the positions of embedded secret message because the blocks are distributed arbitrary according to minimum distance between embedding and replacing blocks and this feature make secret message more protected against detection and tampering. Furthermore the secret message recovery algorithm is does not need the original audio cover signal.

3. Proposed Audio Steganography Scheme

Fig. 1 shows the general block diagram of the proposed hiding algorithm. The main stages of the algorithm are as follow:

3.1 Cover signal decomposition

The input cover audio signal $C = \{c\ (i)\ 0 < i \leq Z,\ Z\text{: number of samples}\}$ is decomposed by using DWPT to L level. Haar filter was chosen in decomposition process because of it is a quadrature mirror filter type and it has finite impulse response. The output from DWPT decomposition stage are 2^L signals, each signal has $Z/2^L$ sample. These signals have different power and different frequencies. One of them which has the greatest power and lowest frequency represents the approximation signal $A = \{a\ (i),\ 0 < i \leq Z/2^L\}$, this signal is kept without any modification to maintain the quality of output stego-signal. The other signals have decreasing powers starting from lowest to highest frequencies details components $D = \{d_j(i),\ j=1,2,..2^L-1,\ 0 < i \leq Z/2^L\}$. Except for the lowest frequency component, other details signal are arranged in the 3D matrix named *Det*.

$$Det = \{\det_{j,k}(i) = 1,2,......., V, k = 1,2,...., W, i = 1,2,......N\} \quad (1)$$

Where $V = 2^L-2$ represents the number of details signals that are being used in the embedding process, $W = Z/(2^L N)$ represents number of blocks in each details signal, and N represents number of samples in each block. This matrix will be fed to block matching process for information embedding process while keeping the lowest frequency details component (D_{2^L-1}) to be used for embedding the key.

3. 2 Secret message pre-processing

The secret message M pre-processing depends on the type of the message. If it is an audio signal has number of bit per sample similar to cover signal, directly it is segmented to Q blocks each block has N samples as shown in Eq. (2).

$$M = \{m(i, j), i = 1,2,...., Q, j = 1,2,......., N\} \quad (2)$$

If the secret message is an image or text, it will have 8 bits per pixel or character. While the audio cover signal for example has 16 bits per sample, therefore it will be combined each two neighbor pixels or character into one sample of 16 bits to reduce number of samples to half. After that arranging the produced message in a one dimension vector, and then segmenting the resultant vector into Q blocks, each block has N samples such as given in Eq. (2).

The produced matrix in Eq. (2) is scaled depends on *Det* matrix which is produced in Eq. (1) as follows:

$$df = mean(M) / mean(Det) \quad (3)$$

$$SM = M / df \quad (4)$$

Where df is a division factor of the primary scale to make message blocks samples within range of details signal matrix. The function *mean* is used to compute average of *M and Det* matrices, and *SM* is a scaled message matrix.

3.3 Block matching Process

Block matching process construct the matching matrix by computing minimum distance between *MS* and *Det* blocks. The matching matrix (*Match)* is a 3D matrix that includes Q layers, where each layer contains $V\times W$ elements. Each element represents a minimum distance between block i from the *MS* matrix and block k from details signal j according to Eq. (5).

$$match(i, j, k) = \frac{1}{N}\sum_{r=1}^{N}\sqrt{(MS_i(r) - Det_{j,k}(r))^2} \quad (5)$$

$$Match = \{match(i, j, k), i = 1,2,......., Q, j = 1,2,...., V, k = 1,2,....., W\} \quad (6)$$

3.4 Key generating and Information Embedding

There are two functions in this stage, first, to generate the key that contains positions of embedding blocks and main scale factors between the message blocks and replacing cover blocks. Second function is embedding of the secret message blocks. The steps of this stage are as follows:

- For each layer i within the *Match* matrix, compute the positions $m_{1,i}$, and $m_{2,i}$ for minimum distance which represent minimum element in a layer. The indices $m_{1,i}$ and $m_{2,i}$ give the position of the block in *Det* matrix that has maximum matching with the block i from the scaled message matrix MS.

- Perform scaling between message block (MS_i) and replacement block ($Det_{m1,i\ ,m2,i}$). This is done as follows:

$$sf_i = mean(MS_i - Det_{m_{1,i},m_{2,i}}) \qquad (7)$$

$$mss_{i,j} = ms_{i,j} - sf_i \qquad (8)$$

$$MSS_i = \{mss_i(j),\ j = 1,2,.....N\} \qquad (9)$$

Where sf_i is the subtraction factor that is obtained by computing the average error between MS_i and $Det_{m_{1,i},m_{2,i}}$ blocks, where $ms_{i,j}$ represents sample j from block MS_i, $mss_{i,j}$ represents sample j from scaled block MSS_i.

- Construct main key vector for each message block i as follows:

$$Key_i = [m_{1,i}, m_{2,i}, sf_i] \qquad (10)$$

- Replace elements of block $Det_{m_{1,i},m_{2,i}}$ by elements of block MSS_i .
- Modify each element that has position $m_{1,i}$, $m_{2,i}$ in *Match* matrix layers $i+1$ to Q equal to large number (for example 10^6) to avoid embedding overwrite if there is any similarity in matching.
- Repeat all the above steps Q times until embedding process of all message blocks completed.

Subsequently, the main key matrix (Key = $\{Key_i(j),\ j = 1, 2, \ldots., Q\}$) and modified details matrix (*Det*) will be generated.

3.5 Main key encryption and embedding

The generated key in previous stage is encrypted and embedded in the lowest frequency details component (D_2^L-1). We chose this component for embedding the main key matrix because it has power higher than other details signals and therefore the embedded key will be more resistance to distortion or lost. In the proposed scheme any type of encryption can be used, either secret or public encryption.

In this work we use a simple stream cipher based pseudo random number generator (*PRN*) to produce a random number vector that has length of $3\times Q$ samples, same length of the main key vector that was produced by converting *Key* matrix to 1D vector as shown in Fig 2. Next, convert both vectors to binary vectors with resolution of 16 bit per sample to obtain random and key binary vectors. The encrypted key (*Enkey*) vector with length of $3\times 16\times Q$ bits is produced by making XOR between two binary vectors.

The encrypted key vector is embedded by using the same embedding technique which was used in [9] with some modification to be resistance to the additive noise. All the lowest frequency details component samples will be converted to integers with scale based on the maximum value. In other word the maximum value will be the highest 16-bit integer. Then for each of these integers d_i, we need to find the biggest P_i which satisfies the following inequality

$$2^{P_i} < d_i < 2^{P_i+1} \qquad (11)$$

For each sample i of the above integers, we have nb_i bits that to be used for embedding encrypted key. We can compute nb_i according to Eq. (12).

$$nb_i = P_i - sb_1 - sb_2 \qquad (12)$$

Where sb_1 is safety bits with direction from the most significant bit to maintain good output quality and sb_2 is the safety bits considered from the least significant bit in order to make embedded bits resistance to the additive noise. Any sample d_i that has zero or negative nb_i will not be used in the key embedding process.

3.6 Stego-Signal Reconstruction

The final stage in the proposed hiding algorithm is the stego-signal reconstruction using Haar IDWPT. The reconstructed process uses modified *Det* matrix and modified D_2^L-1 vector as well as non-modified approximation signal as inputs to IDWPT to obtain on the output stego-signal.

4. Recovery Algorithm of Secret Message

Fig. 3 shows the general block diagram of the secret message recovery algorithm. The algorithm starts by decomposing stego-signal using Haar DWPT, then followed by rearrangement of all the details signal except lowest frequency details signal in 3D matrix *Det* similar to section (3.1) Eq. (1).

The lowest frequency details signal (D_2^L-1) enters the key recovery stage which starts by scaling the signal, then search for the largest power of 2 for each scaled sample with inequality (11) to be satisfied. Depend on the positive values of the nb_is which is computed using Eq. (12), the main key vector *Keyvec* is reconstructed using Eq. (13).

$$Keyvec = \{d_i(j),\ j = 1,..., nb_i,\ i = 1,2,......., \frac{Z}{2^L}\} \qquad (13)$$

Where d_i is a binary vector for *i* sample in vector D_2^{L}*-1* that has positive nb_i.

The key vector is deciphered in decryption stage using same key and procedure that was used in key encryption process in section (3.5) except main key vector *Keyvec* will not be converted to a 1D and binary because it is already a binary vector. The output from this stage is converted to 2D matrix to obtain the main key matrix (*Key*).

The main key and *Det* matrices are fed to message block recovery stage to reconstruct the secret message blocks as Eq. (14):

$$[m_{1,i}, m_{2,i}, sf_i] = Key_i$$
$$MSS_i = Det_{m_{1,i}, m_{2,i}}$$
$$MS_i = \{mss_i(j) + sf_i,\ j = 1,2,\ldots\ldots,N\} \qquad (14)$$

Where mss_i *(j)* is the sample *j* from block *i* in *MSS* matrix. Finally, the secret message is scaled and arranged by inversing procedure in section (3.2) to obtain the required secret message.

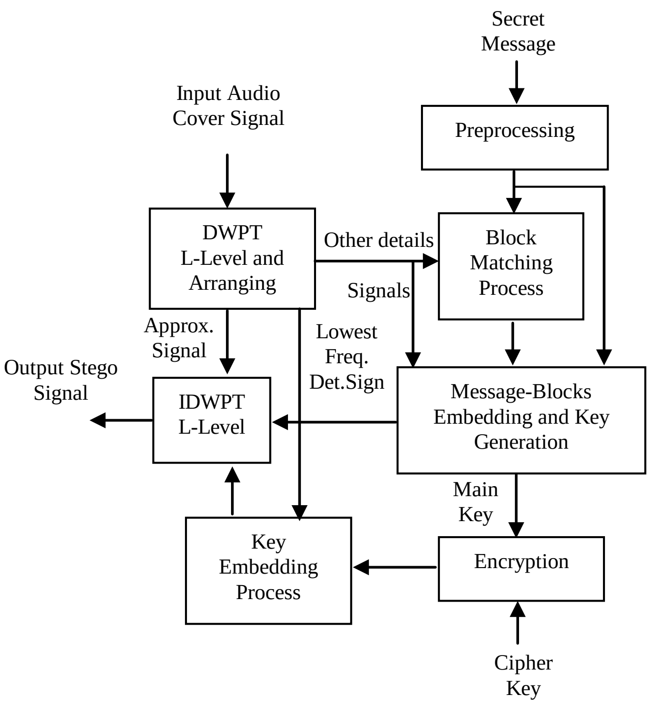


Fig 1 General Block Diagram of Proposed Hiding Scheme

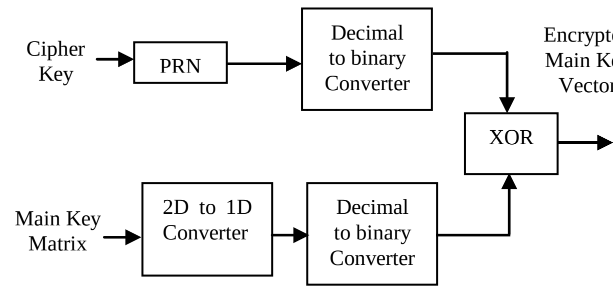


Fig. 2 Main Key Encryption Process

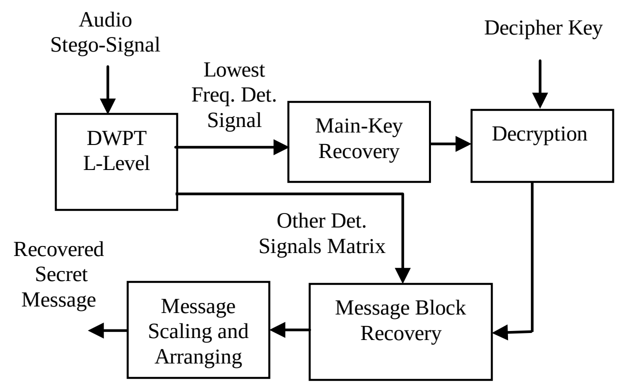


Fig. 3 Recovery Algorithm of Secret Message

5. Results and Discussion

The proposed algorithm was tested using three audio cover signals: male speaker, female speaker and music. Each signal has 16 bit per sample and 44100 samples/sec. Also three data type: audio, image (as given in Fig. 4), and text are used as secret messages in the tests. The quality of output signal in each test was computed using signal to noise ratio (*SNR*) according to Eq. (15).

$$SNR = 10\log_{10}\frac{\sum_{k=1}^{Z} C^2(k)}{\sum_{k=1}^{Z}[C(k) - C'(k)]^2} \qquad (15)$$

Where *C* and C' are input cover signal and output stego-signal respectively, *Z* represents the number of samples in each one of them.

The similarity between extracted secret message and original secret message was computed using normalized correlation *NC* between them according to Eq. (16).

$$NC(M,M') = \frac{\sum_{k=1}^{QN} M(k)M'(k)}{\sqrt{\sum_{k=1}^{QN} M(k)^2}\sqrt{\sum_{k=1}^{QN} M'(k)^2}} \quad (16)$$

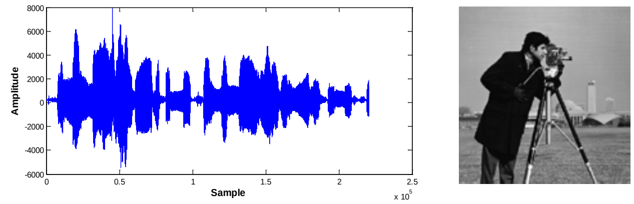
Where M and M' are original and extracted secret message respectively, QN represents number of samples in each one of them.

The proposed scheme was tested for different hiding capacity and the results show that it has an acceptable *SNR* (*35 db* and above according to ref. [18]) with capacity about of 250 *kb/sec*. This capacity represents more than 35% from input audio file size. Fig. 5 show the relationship between *SNR* and embedding capacity for three different message type and three different cover signals. In all experiments of the testing, the following attributes were used: length of block N=100 *sample/sec*, Haar type DWPT, $L=2$, and $sf_1=sf_2=2$ *bits*. From Fig. 5 we can recognize small differences in *SNR* with respect to capacity for different cover signal, and this is due to difference in frequencies. Lower frequency signal has better *SNR* for same capacity than higher frequency signal because it has more percentage power in its non-modified approximation signal.

The extracted secret message is recovered using message recovery algorithm in section (4). The similarity between extracted and original messages was computed using Eq. (16). Without adding noise to the stego-signal, the recovered message is exactly similar to the original message (*NC (M, M') =1*) for different message types and different cover signals. The results show that the proposed algorithm has significant resistance to Additive White Gaussian Noise (AWGN). Fig. 6-8 compares AWGN resistivity of our proposed scheme to the LSB-based DWT scheme. In LSB-based DWT scheme, the secret message will lost nearly about 65 db of AWGN, while in the proposed scheme secret message can be recovered about 25 db of AWGN. This superior performance of the proposed scheme is due to the power of embedding message is much higher compared to that of the LSB-based DWT scheme.

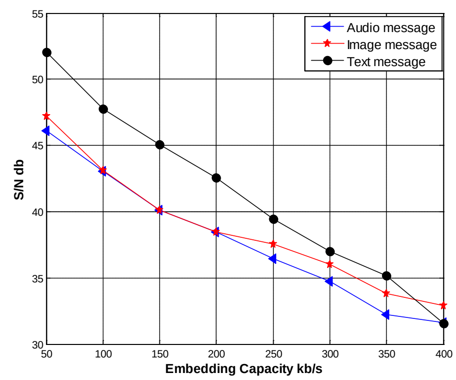
6- Conclusions

We present in this paper a high capacity and AWGN resistance audio steganography algorithm based on DWPT and block matching. Because of the arbitrary block matching that depends on minimum distance between blocks, an arbitrary key is produced that made the proposed algorithm having strong security. The results show that ability of proposed algorithm to embed high capacity of data of different type that can reach to more than 35% from the input audio file size with an acceptable *SNR*. Also it has significant resistance to AWGN. The proposed algorithm can recover a recognized secret message until up to 25 db of AWGN.

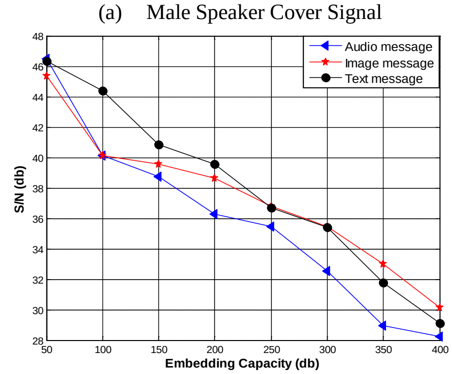


(a) Speech Message Signal (b) Image Message

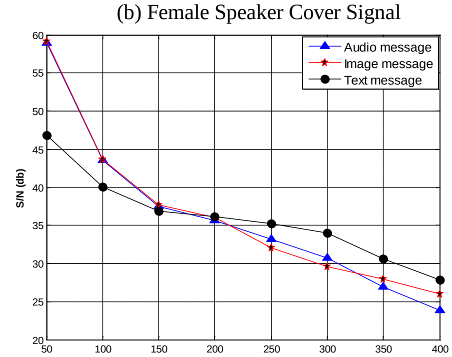
Fig. 4 Input Secret Messages



(a) Male Speaker Cover Signal

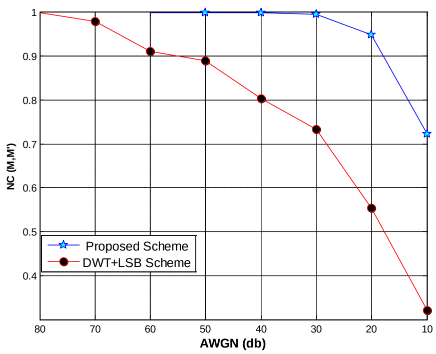


(b) Female Speaker Cover Signal

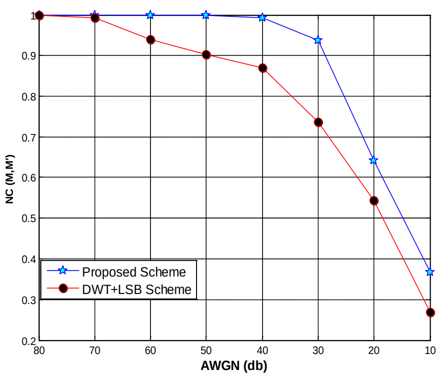


(c) Music Cover Signal

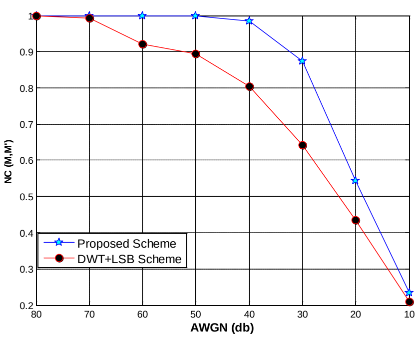
Fig. 5 The Relationship between *SNR* and Embedding Capacity



(a) Audio Message

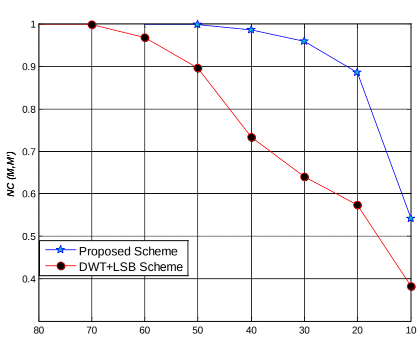


(b) Image Message

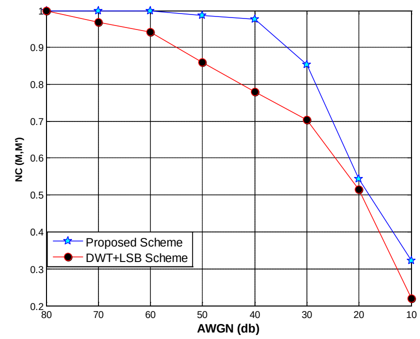


(c) Text Message

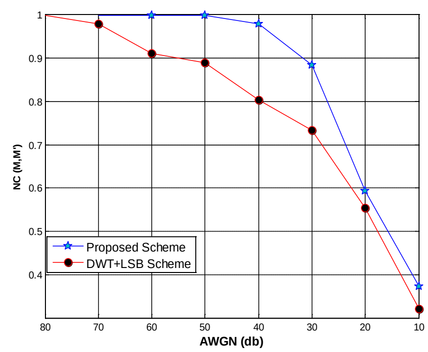
Fig. 6 Relationship between AWGN and *NC* for Male Speaker Cover Signal



(b) Audio Message

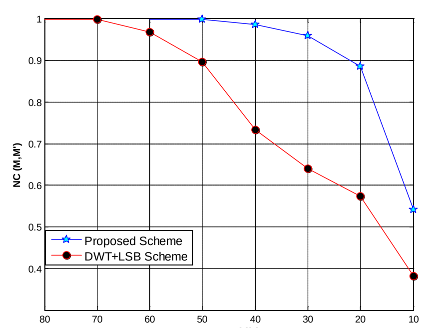


(b) Image Message

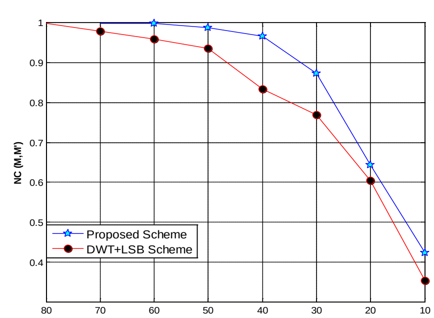


(c) Text Message

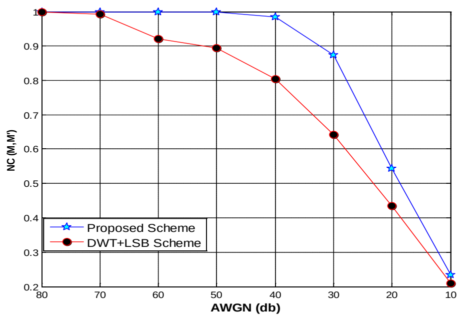
Fig. 7 Relationship between AWGN and *NC* for Female Speaker Cover Signal



(a) Audio Message



(b) Image Message



(c) Text Message

Fig. 8 Relationship between AWGN and *NC* for Music Cover Signal

References

[1] K. Kumar, K. Raja, R. Chhotaray, S. Pattnaik, "Steganography Based on Payload Transformation", IJCSI International Journal of Computer Science Issues, Vol. 8, No. 2, March, 2011, pp. 241-248.

[2] N. Provos and P. Honeyman, "Hide and seek: An introduction to steganography", IEEE Security and Privacy Magazine, Vol. 1, No. 3, June 2003, pp. 32–44.

[3] H. Wang, and S. Wang, "Cyber warfare: Steganography vs. Steganalysis", Communications of the ACM magazine, Vol. 47, No.10, October 2004, pp. 76-82.

[4] A. Cheddad, J. Condell, K. Curran, and P. M. Kevitt, "Digital image steganography: Survey and analysis of current methods", Journal of Signal Processing, Vol. 90, No. 3, Mar. 2010, pp.727–752.

[5] Y. Wang; P.Moulin, "Perfectly Secure Steganography: Capacity,Error Exponents, and Code Constructions", Information Theory IEEE Transactions, Vol. 54, No. 6, Jun 2008, pp. 2706 – 2722.

[6] N. Cvejic,"Algorithms for Audio Watermarking and Steganography", Msc. thesis, Department of Electrical and Information Engineering, Information Processing Laboratory, University of Oulu, Finland Oulu, Finland, 2004.

[7] K. Bailey, and K. Curran, "An Evaluation of image based Steganography methods" Journal of multimedia Tools and Applications, Vol 30, No. 1, July, 2006, pp. 55-88.

[8] N. Meghanathan, and L. Nayak. "Steganalysis algorithms for detecting the Hidden information in image, audio and Video cover media", International Journal of Network Security & Its Application (IJNSA), Vol.2, No.1, January 2010, pp. 43-55.

[9] S. Shahreza, and M. Shalmani, "High capacity error free wavelet Domain Speech Steganography", IEEE International conference on acoustics, speech, and signal processing, March 31 -April 4, 2008, pp. 1729 - 1732.

[10] A. Khashandarag, A. Oskuei, H. Mohammadi and M. Mirnia, "A Hybrid Method for Color Image Steganography in Spatial and Frequency Domain", IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 3, No. 2, May 2011, pp. 113-120.

[11] Z. Zhou, and L. Zhou "A Novel Algorithm for Robust Audio Watermarking Based on Quantification DCT Domain", Third International Conference on International Information Hiding and Multimedia Signal Processing, vol. 1, 26-28 November, 2007, pp. 441-444.

[12] S. Wu, J. Huang, D. Huang and Y. Q. Shi, "A Self-Synchronized Audio Watermarking in DWT Domain", Circuits and Systems, ISCAS 4th. Proceedings of the 2004 International Symposium, Vol. 5, 23-26 May 2004, pp. 712-715.

[13] N. Cvejic, and T. Seppanen,"A wavelet domain LSB insertion algorithm for high capacity audio steganography", Digital Signal Processing Workshop, 2002 and the 2nd Signal Processing Education Workshop. Proceedings of 2002 IEEE 10th, 13-16 October 2002. , pp. 53-55.

[14] M. Hasija1 , A. Jindal2," Contrast of Watermarking Techniques in different domains", IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 3, No. 2, May 2011, pp. 559-563.

[15] G. Tzanetakis, G. Essl, and P. Cook "Audio Analysis using the Discrete Wavelet Transform " In Proceeding WSES International Conference, Acoustics and Music: Theory and Applications (AMTA 2001) , kiathos, Greece, 2001. pp. 318-323.

[16] X. yang,W. P. Niu, M. Lu, , "A Robust Digital Audio Watermarking Scheme using Wavelet Moment Invariance", The Journal of Systems and Software, Vol 84, No. 8 , 2011, pp. 1408-1421

[17] R. Santosa, and P. Bao, "Audio to image wavelet transform based audio steganography", Proceeding of 47th International Symposium, ELMAR, June 2005, pp. 209-212.

[18] T. Wu, H. Lin, W. Hu, and Y. Chen, "Audio Watermarking Scheme with Dynamic Adjustment in Mute Period", Expert System with Applications, Vol. 38, No. 3, pp.6787-6792.

M.Sc. Haider Ismael Shahadi received B.Sc. in Information Engineering in 2001, and M.Sc. in Electronic and Communication Engineering in 2004 from Baghdad university-Iraq. Work as a lecturer in Babylon University since 2005 until 2010. Currently, he is a Ph.D. student in University Tenaga National (UNITEN)-Malaysia. His special fields of interest include Signal and Image Processing, Communication Security, and Artificial Intelligent system.He has 12 publishing in local journal and conference in Iraq.

Assoc. Prof. Dr. Razali bin Jidin received both BSc and MSc in Electrical Engineering in 1986 from Univ. of Bridgeport, Conn, USA. After a career as Computer/Instrument engineer and short teaching, he continued his studies and earned his PhD (in Computer Science and Electrical Engineering) in 2005 from University of Kansas, Lawrence, Kansas, USA. Currently Dr. Razali is a faculty member at Univ. Tenaga in Malaysia. His interest is in Embedded System, Control & Instrumentation in Power Generation, Hardware-Software Co-design. He has published more than twenty five papers at local and international conferences.

Identifying Clusters of Concepts in a Low Cohesive Class for Extract Class Refactoring Using Metrics Supplemented Agglomerative Clustering Technique

A. Ananda Rao[1] and K. Narendar Reddy[2]

[1] **Professor of CSE and Principal, JNTUACE**
Anantapur, Andhra Pradesh, India

[2] **Associate Professor, Dept. of CSE, CVR College of Engineering**
Hyderabad, India

Abstract
Object oriented software with low cohesive classes can increase maintenance cost. Low cohesive classes are likely to be introduced into the software during initial design due to deviation from design principles and during evolution due to software deterioration. Low cohesive class performs operations that should be done by two or more classes. The low cohesive classes need to be identified and refactored using extract class refactoring to improve the cohesion. In this regard, two aspects are involved; the first one is to identify the low cohesive classes and the second one is to identify the clusters of concepts in the low cohesive classes for extract class refactoring. In this paper, we propose metrics supplemented agglomerative clustering technique for covering the above two aspects. The proposed metrics are validated using Weyuker's properties. The approach is applied successfully on two examples and on a case study.

Keywords: *Low Cohesive Class, Metrics, Agglomerative Clustering Technique, Dendrogram, Extract Class Refactoring, Jaccard Similarity Coefficient.*

1. Introduction

The maintainability of object oriented software depends on the quality of software. Software quality is likely to get reduced due to deviation from design principles and due to software deterioration during evolution. The reduction in software quality can be attributed to the presence of bad smells. Low cohesive class is one of the bad smells. Object oriented software with low cohesive classes can increase maintenance cost. Defects (bad smells) in software cause the system to exhibit high complexity, improper behavior, and poor maintainability [1]. It is necessary to detect and correct the defects to make software maintainable. One of the ways to make object oriented software systems maintainable is refactoring. Techniques that reduce object oriented software complexity by incrementally improving the internal software quality without affecting the external behavior come under refactoring [2]. In the context of software under evolution, refactoring is used to improve the software quality. The improvement in the software quality is, in terms of, maintainability, complexity, reusability, efficiency, and extensibility [3].

In the literature active research is being carried out with respect to object oriented software refactoring**.** Considering the importance of low cohesive classes and refactoring, we have proposed an approach for identifying low cohesive classes and clusters of concepts in low cohesive classes for extract class refactoring. The low cohesive class indicates the presence of god class or divergent change bad smell. Divergent change bad smell is present in object oriented software whenever a class needs to be changed for different changes for different reasons in other classes [4]. The changes are propagated to the affected class due to rippling effects. The change propagations depend on strength of coupling between classes. Using strength of coupling between classes, our metric DOCMA(AR) (dependency oriented complexity metric for an artifact affected by ripples) can indicate the class affected by divergent change bad smell [5]. This bad smell indicates that the affected class has low cohesion and is a large class. According to Demeyer [6] the god class is low cohesive and memory consuming class. Object oriented software with low cohesive classes can increase maintenance cost. Low cohesive class performs operations that should be done by two or more classes. One of the design principles in object oriented approach is to design the class with "single minded" function [12]. The violation of this design principle results in large, complex, and low cohesive classes. The low cohesive class cohesion can be improved by splitting the class by extracting cohesive and independent groups of members addressing different functionalities using extract class refactoring [4]. A concept is a cluster (group) of class members addressing a single minded function. In this regard, two aspects are involved; the first one is to identify the low cohesive class and the second one is to identify clusters of concepts in low cohesive class which need to be refactored using extract class refactoring. In this paper, we propose an approach which is based on metrics supplemented hierarchical agglomerative clustering technique for covering the above two aspects. In this paper, "agglomerative clustering technique" means "agglomerative clustering algorithm".

Even though clustering techniques are mainly used in data mining, they are being applied successfully in software engineering. Clustering techniques can identify groups of similar entities [7],[8]. Clustering methods have very good potential to be used in software engineering [9] indicated by its use in software remodularisation[10]. Clustering techniques can identify groups of similar entities where in each group is conceptually different and these groups may address different functionalities.

The contributions of the proposed approach are the following:

- Metrics are proposed to supplement the agglomerative clustering technique (ACT) to handle the situations where some small clusters are formed.
- Identification of low cohesive classes.
- Identification of clusters of concepts in a low cohesive class using the approach which is based on metrics supplemented agglomerative clustering technique.

The organization of the paper is as follows. Section 2 presents the related work. Section 3 presents approach to clustering. The results of experimental cases are presented and discussed in section 4. The conclusions have been placed in section 5.

2. Related Work

The low cohesive classes can be caused by god class or divergent change bad smells. Lot of research has been done in identifying god classes. The main contributions are from [11], [1], [6]. According to Deymer [6] the god class is low cohesive and memory consuming class. Any change to the system may lead to this class. According to Trifu and Marinescu [11] god classes are "large, non-cohesive classes that have access to many foreign data" and they proposed a metrics-based method to identify. Tahvildari and Kontogiannis [1] proposed quality design heuristics and use a diagnosis algorithm based on coupling, complexity, and cohesion to identify design problems (flaws).

Lot of work has been done with respect to remodularising or partitioning or clustering large software modules. Some of them are [13], [14], [15]. In all of the above works remodularisation of software modules in a higher level (like package or file level) is proposed. We need clustering at class level.

Some of the works which focuses on software clustering at class level are: Simon et al. [16] suggested visualization techniques to identify extract class opportunities. Visualizing large classes can be difficult and make it difficult to identify clear clusters. De Lucia et al. [17] propose a methodology that uses structural and semantic metrics for identifying extract class refactoring opportunities. The semantic cohesion metric is based on the names of classes and entities which can be developer dependent hence, may change the results. In a recent work, Joshi and Joshi [18] uses concept lattice for identifying extract class refactoring opportunities. It is identified by the authors that for large systems the lattices can become very complex for the designer to identify problematic cases by inspecting the lattice visually. An algorithm [26] is proposed by the authors to find clusters based on similarity matrix. It is likely to consume more time to identify the clusters and it is threshold value dependant.

In most recent work, Marios Fokaefs et al [19] apply the agglomerative clustering algorithm for several threshold values (ranging from 0.1 to 0.9) and present all possible results to the user. Algorithm also identifies clusters of cohesive entities ranked according to their impact on the design of the whole system and presented to the designer. In contrary to [19], instead of presenting results at different thresholds, we compute clusters at particular threshold and supplement with metrics to merge small clusters with other clusters. In our approach, we supplement the agglomerative clustering technique (ACT) with metrics which can handle situations where ACT alone cannot give acceptable clusters in some situations. Our approach is not much dependent on the threshold value if it is not chosen either near 0 or near to 1. Our approach reduces much human intervention.

3. Proposed Approach

The aim of the proposed approach is to identify the low cohesive classes and clusters of concepts in low cohesive classes.

The proposed approach contains two steps:

Step 1:
Identify low cohesive classes using the metrics LCOM[22], TCC[23], and DOCMA(AR) [5]. LCOM and TCC metrics values can be used to identify the low cohesive class due to god class bad smell. Whereas, DOCMA(AR) metric value can be used to identify the low cohesive class due to divergent change bad smell.

Step 2:
The metrics supplemented agglomerative clustering technique is applied on low cohesive class which is found in step 1. During step 2, the clusters of concepts which need to be refactored are identified. During this step, firstly, similarities between the class members are calculated and then the agglomerative clustering algorithm (technique) is applied to find the clusters at a specified threshold. At the specified threshold there may be some small clusters which need to be merged, this is done with the help of proposed metrics. The agglomerative clustering technique is explained in the following section 3.1.

3.1 Agglomerative Clustering Technique (ACT)

The agglomerative clustering algorithm [8] (which is a hierarchical clustering algorithm) is used in this paper.
The Agglomerative Clustering Algorithm (Technique) is given below:

Step 1: Assign each entity (class member) to a single cluster.

Step 2: Repeat merging while the specified threshold value is not reached.

- Merge two closest clusters according to the considered merging criteria.

Step 3: Display the outcome of the algorithm as a hierarchy of clusters (Dendrogram).

The algorithm requires a threshold value for the similarity metric as a cut-off value. The clusters which are output from the algorithm are at the threshold (cut-off) value. The output hierarchy of the clusters is usually represented by a dendrogram. It has tree like structure. The leaves of the tree represent the individual (single) entities. During the merging process intermediate nodes are formed, they are actual clusters to be output based on cut-off value. The root is the final cluster which contains all the entities. The tree height can be represented using distance metric value or similarity metric value. We used similarity metric value in this paper.

In Hierarchical agglomerative clustering algorithm different linkage methods are available.

1. Single linkage 2. Complete linkage
3. Average linkage 4. Weighted linkage

According to Anquetil and Lethbridge [20] single linkage gives less coupled clusters, complete linkage favors more cohesive clusters, and average linkage gives clusters somewhere in-between the above two. In this paper we have used complete linkage method. The similarity metric we used is the Jaccard similarity metric. Anquetil and Lethbridge [20] indicate in their paper that jaccard distance metric produces good results in software remodularisation. To define the Jaccard similarity metric [24] between two class members we employ the notion of property set of class member. The property set for a method ($PSet\_m_i$) is, the method itself and the methods and fields used (accessed/called) by that method. The property set for a field ($PSet\_f_i$) is, the field itself and the methods using (accessing) it. These property sets are similar to dependency sets used by Simon et al[16].
Similarity Matrix: Similarity matrix (m **x** m) is constructed using the computed values of the above similarity based metrics. Where, m is number of members of the class. Class member means it can be a class method or field (variable).

The **Jaccard index**, also known as the **Jaccard similarity coefficient** (originally coined *coefficient de communauté* by Paul Jaccard [24]), is used for comparing the similarity and diversity of entities (sample sets).

Based on defined property sets we calculate the Jaccard similarity metric between two class members A and B as follows:

$$JSimM(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

Similarity between two entities depends on the properties which are shared [25].

3.2 Proposed Metrics to Supplement Agglomerative Clustering Technique

The clusters identified at cut-off value may give clusters with very few members (may be single member clusters), hence we need to merge them with other clusters. For this purpose we have proposed some metrics. The metrics are given below.

CIM_V - Cluster Identification Metric for Variable = ratio of number of methods within the cluster that reference the variable under consideration to the total number of methods in the cluster

CIM_{VR}_M - Cluster Identification Metric for Method (Based on Variable References) = ratio of number of variables within the cluster referred by the method to the total number of variables in the cluster.

CIM_C_M - Cluster Identification Metric for Method (Based on Method Calls) = ratio of number of methods within the cluster called by method under consideration to the total number of methods in the cluster

CIM_I_M - Cluster Identification Metric for Method (Based on Method Invocations) = ratio of number of methods within the cluster that invoked method under consideration to the total number of methods in the cluster.

Contexts for applying the proposed metrics:

The contexts in which the proposed metrics can be applied are given in Figures 1, 2, 3, 4, and 5. The cluster to be merged with other cluster is indicated by C1.

The contexts are:
Context 1: Only one variable (field) in C1 (Figure 1).
Context 2: Only one method in C1 (Figure 2).
Context 3: Only variables (two or more variables).
Context 4: Only methods (two or more methods)
Context 5: Methods and variables. (At least one method and one variable)

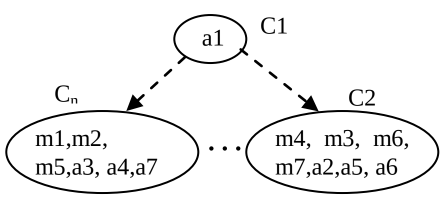


Fig. 1 Context1 for applying proposed metrics

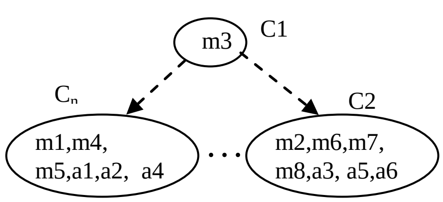


Fig. 2 Context2 for applying proposed metrics

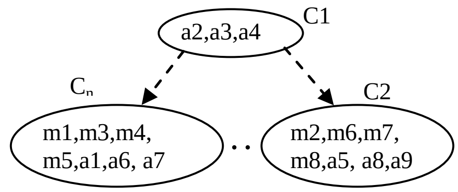


Fig. 3 Context3 for applying proposed metrics

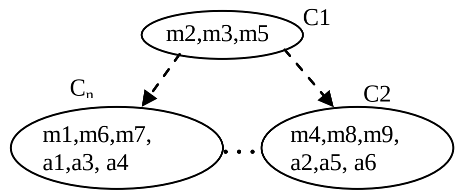


Fig. 4 Context4 for applying proposed metrics

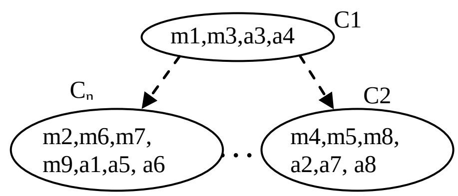


Fig. 5 Context5 for applying proposed metrics

The proposed metrics can be applied for all the five contexts. In all the contexts (Figures 1 to 5) proposed metrics can be used in merging the element(s) from cluster 1 (C1) with cluster 2 (C2) or cluster 3 (C3) or cluster n (C_n).

3.2.1 Validation using Weyuker's Properties

The proposed metrics are validated using the Weyuker's properties [21]. The validation results are given in Table 1.

Table 1: Validation results

	P1	P2	P3	P4	P5	P6	P7	P8	P9
CIM_V	Y	Y	Y	Y	Y	Y	NA	Y	Y
CIM_{VR}_M	Y	Y	Y	Y	Y	Y	NA	Y	Y
CIM_C_M	Y	Y	Y	Y	Y	Y	NA	Y	Y
CIM_I_M	Y	Y	Y	Y	Y	Y	NA	Y	Y

Only eight properties are considered for validation. The seventh property is not considered for validation. Seventh property specifies that the permutation of elements within measured entity can change the metric value. This property is not suitable for OO systems, since the order of methods declaration inside a class does not change the order in which they are executed. Cherniavsky and Smith [27] suggest that this property is not appropriate for OOD metrics.

For contexts 1, 2, and 3 all the proposed metrics satisfied all the eight properties which are applicable to object oriented systems. In case of context 4 the metrics CIM_C_M and CIM_I_M did not satisfy the properties 5 and 9. In case of context 5, all the metrics did not satisfy properties 5 and 9. However, at least, 6 properties are satisfied out of 8 properties in all the contexts. The validation results indicate the suitability of metrics for the purpose for which they are proposed.

4. Experimental Results

Example 1:

The directed graph for the members of the class considered in this example is given in Figure 6. In the graph, the nodes represent methods and variables (fields) and the edges indicate that a dependency exists between two class members.

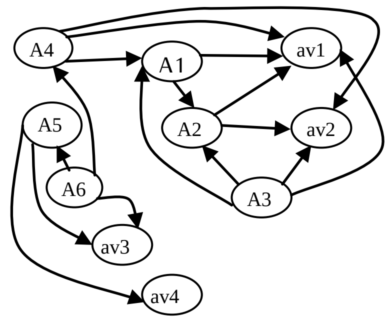


Fig. 6 Directed graph for the members of class in example 1

The class has six methods and four variables (fields).

The methods are: A1, A2, A3, A4, A5, A6
The variables are: av1, av2, av3, av4

The Jaccard similarity metric is used for finding the similarity between the class members. Similarity matrix for the members of the class given in example 1 is shown in Table 2. The hierarchical agglomerative clustering algorithm is applied on the similarity matrix (Table 2). Figure 7 shows the dendrogram produced by the algorithm.

Table 2: Similarity matrix for the class members shown in Figure 6

	A1	A2	A3	A4	A5	A6	av1	av2	av3	av4
A1	1									
A2	0.5	1								
A3	0.6	0.6	1							
A4	0.4	0.4	0.5	1						
A5	0	0	0	0	1					
A6	0	0	0	0.14	0.4	1				
av1	0.5	0.33	0.67	0.5	0	0.12	1			
av2	0.17	0.4	0.5	0.33	0	0.14	0.5	1		
av3	0	0	0	0	0.5	0.75	0	0	1	
av4	0	0	0	0	0.67	0.2	0	0	0.25	1

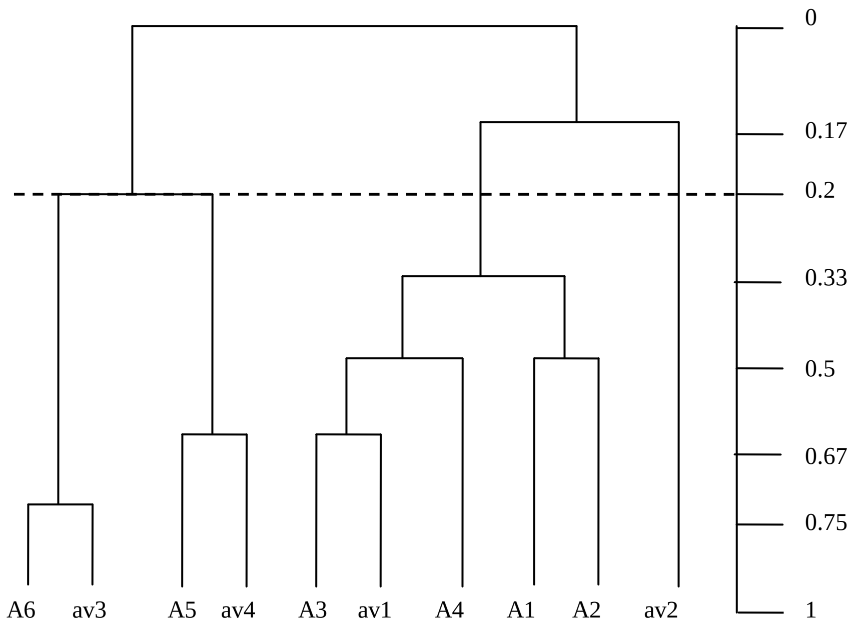


Fig. 7 Dendrogram showing the clusters

Clusters Identification at a given threshold value:

Consider the threshold value of 0.2. At this threshold, three clusters are produced, they are:
G1= {A5, A6, av3, av4}
G2= {av2}
G3= {A1, A2, A3, A4, av1}

Since G2 contain only one member (variable member) it needs to be merged with other cluster. This represents context 1 (Figure 1). For this purpose, **CIM_V** metric is computed.

CIM_V (av2, G1) = 0/2=0,
CIM_V (av2, G3) = ¾ = 0.75
Since **CIM_V** (av2) is more with respect to G3 when compared to G1, hence G2 (av2) is merged with G3.

After merging two clusters are formed. They are:
G1= { A5, A6, av3, av4},
G3= { A1, A2, A3, A4, av1, av2}

Example 2:

The directed graph for the members of the class considered in this example is given in Figure 8.

The class has seven methods and two variables (fields).
The methods are: A1, A2, A3, A4, A5, A6, A7
The variables are: av1, av2

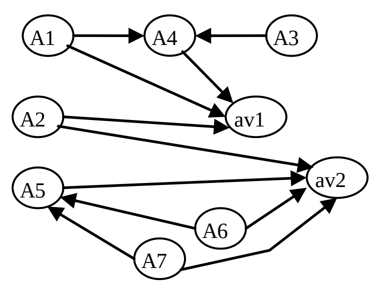


Fig. 8 Directed graph for the members of class in example 2

The similarity matrix is constructed using Jaccard similarity metric values for the class members and the hierarchical agglomerative clustering algorithm is applied on the similarity matrix. The clusters formed at 0.5 threshold value are:

G1= {A1, A4, A3}
G2= {av1}
G3= {A2}
G4= {A5, A6, A7}

G5= {av2}

Since single member clusters are there, they need to be merged. Merging G2 or G5 with other cluster represents context 1. Whereas, merging G3 with other cluster represents context 2 (Figure 2).

For context 1 the metric **CIM_V** is computed**.**

CIM_V (av1, G1) = 2/3=0.66
CIM_V (av1, G4) = 0/3=0

Since **CIM_V** metric value for av1 (G2) with respect to G1 is more when compared to G4, hence G2 is merged with G1. Similarly, av2 (G5) is merged with G4.

Since merging G3 (A2) with other cluster represents context 2, the metrics **CIM_{VR}_M, CIM_C_M,** and **CIM_I_M** are computed**.**

CIM_{VR}_M (A2, G1) = 1/1 = 1
CIM_{VR}_M (A2, G4) = 1/1 = 1
CIM_C_M (A2, G1) = 0/3 = 0
CIM_C_M (A2, G4) = 0/3 = 0
CIM_I_M (A2, G1) = 0/3 =0
CIM_I_M (A2, G4) = 0/3 = 0

Since same metrics values indicate A2 (G3) can be placed either in G1 or G4.
The formed clusters after merging are:
G1= {A1, A4, A3, av1, A2}
G4= {A5, A6, A7, av2}

In situations where a member has equal similarity with two groups, (for example, in the above case) coupling with respect to other classes need to be computed before merging.

Case Study : Bank Application

Bank application developed by students as part of their academic project is given as case study. The novice designer has a tendency to deviate from the design principles and may end up designing low cohesive classes. The bank application is developed in java. It contains the following classes:
CustomerAccount, Bank, ATMCard, Locker, DemandDraft, FixedDeposit, Employee, Manager, Report, Salary, DatabaseProxy. The proposed approach is applied on this case study to find low cohesive classes and clusters of concepts in low cohesive classes. The approach contains two steps.

Table 3: Similarity matrix for the members of class "CustomerAccount"

	M 1	M 2	M 3	M 4	M 5	M 6	M 7	M 8	M 9	M 10	M 11	V 1	V 2	V 3	V 4	V 5	V 6	V 7	V 8	V 9	V 10	V 11
M1	1	.5	.5	.43	.43	.1	0	0	.1	0	0	.17	.2	.33	.25	.22	0	0	0	0	0	0
M2		1	.7	.67	.25	.1	0	0	.1	0	0	.27	.33	0	.25	.38	0	0	0	0	0	0
M3			1	.67	.5	.1	0	0	.1	0	0	.27	.33	0	.25	.38	0	0	0	0	0	0
M4				1	.33	.13	0	0	.13	0	0	.2	.25	0	0	.29	0	0	0	0	0	0
M5					1	.13	0	0	.13	0	0	.33	.43	0	.14	.13	0	0	0	0	0	0
M6						1	.5	.5	.11	0	0	.18	0	0	0	0	.29	.29	.29	0	0	0
M7							1	.6	0	0	0	0	0	0	0	0	33	.33	.33	0	0	0
M8								1	0	0	0	0	0	0	0	0	.33	.4	.33	0	0	0
M9									1	.5	.5	.18	0	0	0	0	0	0	0	.29	.29	.29
M10										1	.6	0	0	0	0	0	0	0	0	.33	.33	.33
M11											1	0	0	0	0	0	0	0	0	.33	.33	.33
V1												1	.56	.11	.33	.44	.1	0	0	.1	0	0
V2													1	.14	.43	.57	0	0	0	0	0	0
V3														1	.2	.17	0	0	0	0	0	0
V4															1	.5	0	0	0	0	0	0
V5																1	0	0	0	0	0	0
V6																	1	.6	.6	0	0	0
V7																		1	.6	0	0	0
V8																			1	0	0	0
V9																				1	.6	.6
V10																					1	.6
V11																						1

M1 – open(), M2 – deposit((), M3 – withdraw(), M4 – display(), M5 – close(), M6 – addCust(), M7 – updateAddr(), M8 – displayAddr(), M9 – apprLoan(), M10- repay(), M11-closeloan(), V1 – CustName, V2 – AcNo, V3 – AcType , V4 – Amount, V5 – Balance, V6 – CustId, V7 – PermAddr , V8 – CommnAddr, V9 – LoanNo, V10- LoanType, V11- LoanAmnt

Step 1:
During this step LCOM and TCC metrics are computed for the different classes. The CustomerAccount class has LCOM metric value of 5 and TCC metric value of 0.49. The high LCOM value and low TCC value indicate that the CustomerAccount is a low cohesive class.

Step 2:
The similarity matrix is constructed using Jaccard similarity metric values for the class members of the class “CustomerAccount” and it is shown in Table 3. The hierarchical agglomerative clustering algorithm is applied on the similarity matrix given in Table 3.

The clusters identified at the threshold value of 0.2 are:

G1 = M1, M2, M3, M4, M5
G2 = V1, V2, V4, V5
G3 = V3
G4 = M6, M7, M8, V6, V7, V8
G5 = M9, M10, M11, V9, V10, V11

Since G3 contains single variable and G2 contains only variables, they need to be merged with other groups. They represent contexts 1 and 3 respectively. Hence the metric **CIM_V** is computed to find the clusters with which G2, G3 can be merged.

Finding the cluster for G3(V3) merging:

CIM_V (V3) with respect to G1 = 1/5 = 0.2
CIM_V (V3) with respect to G4 = 0
CIM_V (V3) with respect to G5 = 0
Since **CIM_V** (V3) is high with respect to G1, hence V3 (G3) is merged with G1.

Finding the cluster for G2 merging:

CIM_V (V1) with respect to G1 = 5/5 = 1
CIM_V (V1) with respect to G4 = 1/3 =0.33
CIM_V (V1) with respect to G5 = 1/ 3 =0.33

CIM_V (V1) with respect to G1 is high when compared to G4 and G5 and **CIM_V** (V2, V4, V5) with respect to G1 is more when compared to G4 and G5, hence merge G2 with G1. After merging three clusters are formed, they are:

G1 = M1, M2, M3, M4, M5, V1, V2, V4, V5, V3
G4 = M6, M7, M8, V6, V7, V8
G5 = M9, M10, M11, V9, V10, V11

4.1 Observations

The LCOM [22] and TCC [23] are computed for the classes considered in the examples and case study before and after refactoring and presented in Table 4. The high LCOM and low TCC values indicate low cohesive classes, whereas low LCOM and high TCC values indicate high cohesive classes. The values in Table 4 indicate the improvement in cohesion due to refactoring in two experimental examples and in one case study.

The high value of LCOM and low value of TCC for the CustomerAccount class of bank application indicate the low cohesive class. The Agglomerative Clustering Technique identified three clusters of concepts for CustomerAccount class. These three clusters are refactored into three classes by using extract class refactoring. The extracted classes are: Account, Customer, and Loan. The 0 (zero) LCOM value and TCC value of 1 for Account (cluster1), Customer (cluster2), and Loan (cluster3) indicate high cohesive classes. Hence our approach could identify the low cohesive class, and the clusters to be refactored. The increase in cohesion after refactoring, indicate the effectiveness of the approach in identifying proper clusters for refactoring.

Table 4. LCOM and TCC values before and after refactoring

	Before Refactoring		**After Refactoring**					
			Cluster1		Cluster2		Cluster3	
	LCOM	TCC	LCOM	TCC	LCOM	TCC	LCOM	TCC
Example 1	7	.47	0	1	0	1	-	-
Example 2	9	.43	0	.5	0	1	-	-
Case Study	5	.49	0	1	0	1	0	1

The proposed approach is useful because of the following reasons:

- It handles the situation where Agglomerative clustering technique (ACT) alone cannot identify proper clusters. Presenting to the user clusters at different thresholds may not solve the problem. In some situations, clusters at any threshold other than maximum (which leads to a single cluster) may contain clusters with single members or small clusters with one member or clusters with only variables. In that situation ACT does not guide us how to merge those small clusters. Hence, proposed metrics will help in those situations.
- In our approach, we need to specify one threshold value somewhere in between 0 and 1, which is neither near to 0 nor near to 1. Merge small clusters with the help of proposed metrics. The rules (in what situations the groups should be merged) can be specified to the tool, which guide the merging process using the proposed metrics. Hence, they reduce the human intervention to study the clusters at different thresholds and decide.
- When a single member is tried to be merged, if metrics values are equal with respect to two or more clusters, it can be merged by considering coupling with respect to other classes.

5. Conclusions

In this paper, we proposed an approach for identifying low cohesive classes and clusters of concepts in low cohesive classes potential for extract class refactoring. Proposed approach consists of two steps. In step 1, low cohesive classes are identified. In step 2, the clusters of concepts in low cohesive classes are identified for extract class refactoring. The proposed approach is based on metrics supplemented agglomerative clustering technique. Agglomerative Clustering Technique is based on the Jaccard similarity metric values between class members. Metrics which are used to supplement agglomerative clustering technique are newly proposed. The metrics are validated using Weyuker's properties. The approach is applied on two examples and on academic software developed by students. In the two examples and case study our approach could find low cohesive classes and clusters of concepts to be refactored. The low cohesive class identified by our approach in the bank application has high LCOM and low TCC value. High LCOM and low TCC values indicate low cohesive class. The clusters identified by our approach are refactored using extract class refactoring. After refactoring the LCOM metric value is decreased whereas TCC metric value is increased. These values indicate increase in cohesion due to refactoring the clusters of concepts into new classes. The increase in cohesion after refactoring, indicate the effectiveness of the approach in identifying proper clusters for refactoring. Hence, our approach could effectively identify low cohesive classes and clusters of concepts to be refactored.

References

[1] L. Tahvildari and K. Kontogiannis. "A Metric-Based Approach to Enhance Design Quality Through Meta-Pattern Transformations" In Proceedings of the 7th European Conference on Software Maintenance and Reengineering, March 26-28, 2003, pp. 183–192.

[2] W .F. Opdyke, "Refactoring : A Program Restructuring Aid in Designing Object-Oriented Application Frameworks", PhD thesis, Univ. of Illinois at Urbana Champaign, 1992.

[3] Tom Mens and Tom Tourwe. "A Survey of Software Refactoring", IEEE Transactions on Software Engineering. Volume 30, Number 2, 2004, pp 126-139.

[4] M. Fowler, K. Beck, J. Brant, W. Opdyke, and D. Roberts. Refactoring Improving the Design of Existing Code, Boston, MA, AddisonWesley, 1999.

[5] K.Narendar Reddy and A.Ananda Rao, "Dependency Oriented Complexity Metrics to Detect Rippling Related Design Defects", ACM SIGSOFT Software Engineering Notes, Volume 34, Number 4, July 2009.

[6] S. Demeyer, S. Ducasse, and O. M. Nierstrasz, Object-Oriented Reengineering Patterns, Morgan Kaufman Publishers, 2002.

[7] P. N. Tan, M. Steinbach, and V. Kumar, Introduction to Data Mining, Addison-Wesley, 2005.

[8] Jiawei Han and M. Kamber, Data Mining Concepts and Techniques, Morgan Kaufmann Publishers, 2005.

[9] V. Tzerpos and R. C. Holt, "Software Botryology: Automatic Clustering of Software Systems", In Proceedings of the International Workshop on Large-Scale Software Composition, 1998.

[10] T. A.Wiggerts, "Using Clustering Algorithms in Legacy Systems Remodularization", In WCRE '97: Proceedings of the 4th Working Conference on Reverse Engineering, 1997.

[11] A. Trifu and R. Marinescu, " Diagnosing Design Problems in Object Oriented Systems", In Proceedings of the 12th Working Conference on Reverse Engineering, 2005.

[12] Roger S. Pressman, Software Engineering A Practitioner's Approach, Sixth Edition, McGraw-Hill Int'l Edition, 2005.

[13] S. Mancoridis, B. S. Mitchell, C. Rorres, Y. Chen, and E. R.Gansner, "Using Automatic Clustering to Produce High-Level System Organizations of Source Code", In Proceedings of the 6th International Workshop on Program Comprehension, 1998 ,pp. 45–52.

[14] D. Doval, S. Mancoridis, and B. S. Mitchell "Automatic Clustering of Software Systems Using a Genetic Algorithm",. In Proceedings of the 5th International Conference on Software Tools and Engineering Practice, 30 August - 2 September 1999.

[15] K. Sartipi and K. Kontogiannis, "Component Clustering Based on Maximal Association", In Proceedings of the IEEE Working Conference on Reverse Engineering, October 2001.

[16] F. Simon, F. Steinbruckner, and C. Lewrentz, "Metrics Based Refactoring", In Proceedings of the 5th European Conference on Software Maintenance and Reengineering, 2011, pp.30–38.

[17] A. D. Lucia, R. Oliveto, and L. Vorraro, "Using Structural and Semantic Metrics to Improve Class Cohesion", In 24th IEEE International Conference on Software Maintenance, 2008.

[18] P. Joshi and R. K. Joshi, "Concept Analysis for Class Cohesion", In European Conference on Software Maintenance and Reengineering, March 24-27 2009, pp. 237–240.

[19] Marios Fokaefs, Nikolaos Tsantalis, and Alexander Chatzigeorgiou,"Decomposing Object-Oriented Class Modules Using an Agglomerative Clustering Technique", In Proc. ICSM, 2009, pp. 93-101.

[20] N. Anquetil and T. Lethbridge, "Experiments with Clustering as a Software Remodularization Method", In WCRE '99: Proceedings of the 6th Working Conference on Reverse Engineering, 1999.

[21] E.J.Weyuker, "Evaluating Software Complexity Measures", IEEE Transactions on Software Engineering, Vol. 14, No. 9, 1988, pp. 1357-1365.

[22] S.R.Chidamber and C.F.Kemerer, "A Metrics Suite for Object Oriented Design", IEEE Transactions on Software Engineering, Vol. 20, No. 6, 1994, pp. 476-493.

[23] J.M.Bieman and B.K.Kang, "Cohesion and Reuse in an Object-Oriented System", In Proc. ACM Symp. Software Reusability (SSR'94), 1995, pp. 259-262.

[24] Paul Jaccard . Étude comparative de la distribution florale dans une portion des Alpes et des Jura. Bulletin de la Société Vaudoise des Sciences Naturelles 37, 1901, pp. 547–579.

[25] Mario Bunge. Treatise on Basic Philosophy, Vol 3: Ontology I, The Furniture of the World, ordrecht_Holland, D.Reidel Publishing Company, 1977.

[26] K.Narendar Reddy and A.Ananda Rao, "Poster paper: Similarity Based Metrics for Performing Extract Class Refactoring for the Class Affected by Ripples", In ISEC(3rd India Software Engg. Conference), Feb 25-27, 2010, Mysore, India.

[27] J.C.Cherniavsky and C.H.Smith, " On Weyuker's Axioms for Software Complexity Measures", IEEE Transactions on Software Engineering, Volume 17, 1991, pp 636-638.

Prof. Ananda Rao Akepogu received B.Sc.(M.P.C) degree from Silver Jubilee Govt. College, SV University, Andhra Pradesh, India. He received B.Tech. degree in Computer Science & Engineering and M.Tech. degree in A.I & Robotics from University of Hyderabad, India. He received Ph.D. from Indian Institute of Technology, Madras, India. He is Professor of Computer Science & Engineering and Principal of JNTUA 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 interests include software engineering and data mining.

Narendar Reddy K 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 Bachelor's degree in Computer Science & Engineering (AMIE(CSE)) from Institution of Engineers, Calcutta, India. Currently he is working as Associate Professor of Computer Science & Engineering at CVR College of Engineering, Hyderabad, India. His main research interests include Object oriented software design, software metrics, refactoring, and software testing. He is a member of IEEE, ACM, AMIE(I), and IAENG.

Amalgamation of Automated Testing and Data Mining : A Novel Approach in Software Testing

Sarita Sharma[1], Anamika Sharma[2]

[1] Computer Science Deptt., DAV Institute of Management
Faridabad, Haryana 121001, India

[1] Computer Science Deptt., DAV Institute of Management
Faridabad, Haryana 121001, India

Abstract

Software engineering comprehends several disciplines devoted to prevent and remedy malfunctions and to warrant adequate behavior. Testing is a widespread validation approach in industry, but it is still largely ad hoc, expensive, and unpredictably effective. In today's industry, the design of software tests is mostly based on the testers' expertise, while test automation tools are limited to execution of pre-planned tests only. Evaluation of test outputs is also associated with a considerable effort by human testers who often have improper knowledge of the requirements specification. This manual approach to software testing results in heavy losses to the world's economy. This paper proposes the potential use of data mining algorithms for automated induction of functional requirements from execution data. The induced data mining models of tested software can be utilized for recovering missing and incomplete specifications, designing a minimal set of regression tests, and evaluating the correctness of software outputs when testing new, potentially inconsistent releases of the system.

Keywords: Automation, bagging, Data Mining

1. Introduction

As we know that creating a reliable and hence a fault free software is one of the major goals of a software developer. This makes software testing one of the most important and critical phases in the software development life cycle.

The process of software testing includes four phases [1] namely modeling the software environment, selection of test cases, running and evaluating test cases and measuring testing progress. As the test case automation is increasing the volume of the test is growing. It makes the selection of test cases very difficult, making the test case reduction highly desirable. After selecting and running the test cases the tester has to evaluate whether the selected test cases expose a fault or not. Traditionally this step was done manually by the human tester, which required a lot of time. As the software systems are growing larger the burden on the human tester is increasing. Using an automated oracle to support the activities of human tester can reduce the cost of the testing process and hence the related maintenance costs [2]. Data Mining is an analytic process designed to explore data (usually large amounts of data - typically business or market related) in search of consistent patterns and/or systematic relationships between variables, and then to validate the findings by applying the detected patterns to new subsets of data. The ultimate goal of data mining is prediction - and predictive data mining is the most common type of data mining and one that has the most direct business applications.

2. Crucial Concepts in Data Mining

The concept of bagging (voting for classification, averaging for regression-type problems with continuous dependent variables of interest) applies to the area of predictive data mining, to combine the predicted classifications (prediction) from multiple models, or from the same type of model for different learning data. It is also used to address the inherent instability of results when applying complex models to relatively small data sets. Suppose the data mining task is to build a model for predictive classification, and the dataset from which to train the model (learning data set, which contains observed classifications) is relatively small. You could repeatedly sub-sample (with replacement) from the dataset, and apply, for example, a tree classifier (e.g., C&RT and CHAID) to the successive samples. In practice, very different trees will often be grown for the different samples, illustrating the instability of models often evident with small data sets. One method of deriving a single prediction (for new observations) is to use all trees found in the different samples, and to apply some simple voting: The final classification is the one most often predicted by the different trees.

3. Automated Software Testing

Today, rigorous application testing is a critical part of virtually all software development projects. As more organizations develop mission – critical systems to support their business activities, the need is greatly increased for testing methods that support business objectives. It is necessary to ensure that these systems are reliable, built according to specification and have the ability to support business processes.

[3]Software testing using an automatic test program will generally avoid the errors that humans make when they get tired after multiple repetitions. The test program won't skip any tests by mistake. The test program can also record the results of the test accurately. The results can be automatically fed into a database that may provide useful statistics on how well the software development process is going. On the other hand, software that is tested manually will be tested with a randomness that helps find bugs in more varied situations. Since a software program usually won't vary each time it is run, it may not find some bugs that manual testing will. Automated software testing is never a complete substitute for manual testing.

Automation allows the testing organization to perform consistent and repeatable test. When applications need to be deployed across different hardware or software platforms, standard or benchmark tests can be created and repeated on target platforms to ensure that new platforms operate consistently.

3.1Automated testing Lifecycle

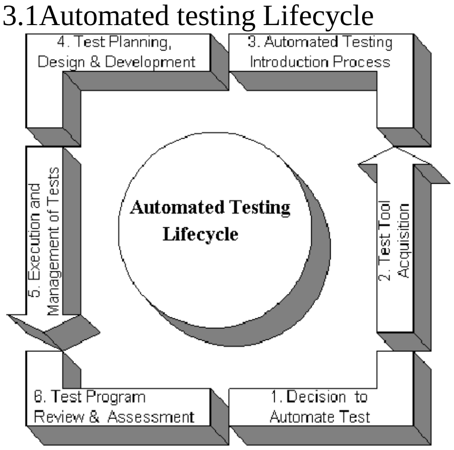


Fig 1. Automated Testing Lifecycle

The Automated Test Lifecycle comprises of six primary components:

- Decision to automate Testing
- Test Tool Acquisition
- Automated Testing introduction process
- Test Planning Design and Development
- Execution and Management of Tests
- Test Program Review and Assessment

3.2 Typical Testing Steps

Most software testing projects can be divided into general steps
a) Test Planning: This step determines like 'which' and 'when'.
b) Test Design: This step determines how the tests should be built the level of quality.
c) Test Environment Preparation: Technical environment is established during this step.
d) Test Construction: At this step, test scripts are generated and test cases are developed.
e) Test Execution: This step is where the test scripts are executed according to the test plans.
f) Test evaluation: After the test is executed, the test results are compared to the expected
results and evaluations can be made about the quality of an application.

3.3 Challenges of Automated Testing

Though automated testing has a lot of benefit, but it also has some associated challenges. Following list includes some of the major challenges of automation [4]:

- Selection of Test Tool
- Customization of Tool
- Selection of Automation Level
- Development and Verification of Script
- Implementation of Test Management System

3.4 New Approach To Automatically Generate Test Cases

Data Mining algorithms can be efficiently used for automated modeling of tested systems. Induced Data Mining models can be utilized for recovering system requirements, identifying equivalence classes in system inputs, designing a minimal set of regression tests, and evaluating the correctness of software outputs.

For this a new approach to automatically generate test cases from SRS and mining of test cases has been discussed.

- A formal transformation of a detailed SRS to a UML state model
- The generation of test cases from the state model
- Mining of Test cases. The introduction of agents can bring enhancement

3.5 Need for effective test automation

The systematic production of high-quality software, which meets its specification, is still a major problem. Although formal specification methods have been around for a long time, only a few safety-critical domains justify the enormous effort of their application. The state of the practice, which relies on testing to force the quality into the product at the end of the development process, is also unsatisfactory. The need for effective test automation adds to this problem, because the creation and maintenance of the test ware is a source of inconsistency itself and is becoming a task of comparable complexity as the construction of the code.

3.6 The Approach

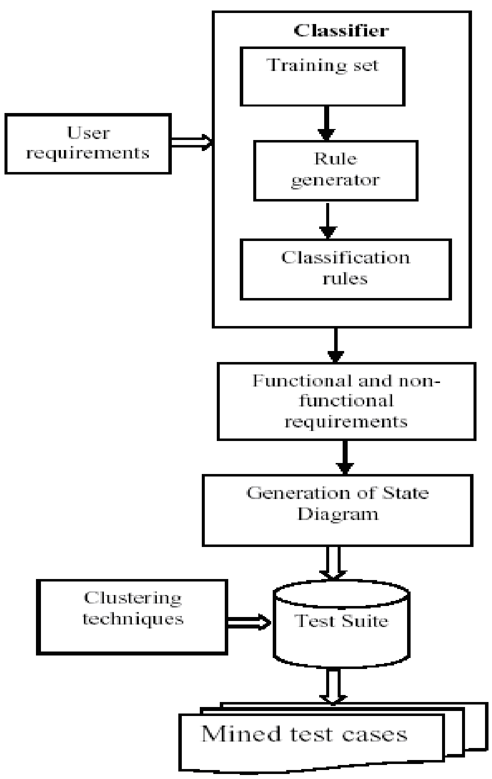


Fig 2. Automatically Generated Test Cases From Software Requirements Specification mining System

(1) Generation of classification rules.
(2) Generate test cases from the UML state machine.
(3) Finally data mining techniques are applied on the generated test cases in order to further reduce the test suite size.

3.7 Mining Techniques for Test Suite Reduction

Data mining is the process of extracting patterns from data. As more data are gathered, with the amount of data doubling every three years, data mining is becoming an increasingly important tool to transform these data into information. It is commonly used in a wide range of profiling practices, such as marketing, surveillance, fraud detection and scientific discovery.

While data mining can be used to uncover patterns in data samples, it is important to be aware that the use of non-representative samples of data may produce results that are not indicative of the domain. Similarly, data mining will not find patterns that may be present in the domain, if those patterns are not present in the sample being "mined". There is a tendency for insufficiently knowledgeable "consumers" of the results to attribute "magical abilities" to data mining, treating the technique as a sort of all-seeing crystal ball. Like any other tool, it only functions in conjunction with the appropriate raw material: in this case, indicative and representative data that the user must first collect. Further, the discovery of a particular pattern in a particular set of data does not necessarily mean that pattern is representative of the whole population from which that data was drawn. Hence, an important part of the process is the verification and validation of patterns on other samples of data. The term data mining has also been used in a related but negative sense, to mean the deliberate searching for apparent but not necessarily representative patterns in large numbers of data. To avoid confusion with the other sense, the terms data dredging and data snooping are often used. Note, however, that dredging and snooping can be (and sometimes are) used as exploratory tools when developing and clarifying hypotheses.

According to [5] , A program fails when it does not do what it is required to do .The purpose of testing a program is to discover faults that cause the system to fail rather than proving the program correctness . A successful test should reveal a problem in software; tests that do not expose any faults are useless, since they hardly provide any indication that the program works properly [6]. In developing a large system, the test of the entire application (system testing) is usually preceded by the stages of unit testing and integration testing . The activities of system testing include function testing, performance testing, acceptance testing, and installation testing Ideally, a minimal test suite can be generated from a complete and up-to-date specification of functional requirements. Unfortunately, frequent changes make the original requirements documentation, even if once complete and accurate, hardly relevant to the new versions of software [7]. To ensure effective design of new regression test cases, one has to recover (reverse engineer) the actual requirements of an existing system. In [8], several ways are proposed to determine input-output relationships in tested software. Thus, a tester can analyze system specifications, perform structural analysis of the system's source code, and observe the results of system execution. While available system specifications may be incomplete or outdated, especially in the case of a "legacy" application, and the code may be poorly structured, execution data seems to be the most reliable source of information on the actual functionality of an evolving system. In paper [9],the idea is extended initially introduced as input-output analysis of execution data can be automated by the IFN (Info-Fuzzy Network) methodology of data mining [11] [10]. In [9] the proposed concept of IFN-based testing has been demonstrated on individual discrete outputs of a small business program. The current study evaluates the effectiveness of the IFN methodology on a complex expert-system application having multiple continuous outputs.

4. RELATED STUDIES

The Paper [12] describes that Data Mining algorithms can be efficiently used for automated modeling of tested systems. Paper [13] proposes a way to process large amounts of data using a machine learning technique from association rule mining. [14] In this paper Case-Based Reasoning and Data mining are used as efficient methods for effort estimation and automated testing .

5. CONCLUSION

With this new proposed approach the system will automatically generate test cases from SRS and mining of test cases. Firstly a formal transformation of a detailed SRS to a UML state model, secondly the generation of test cases from the state model and lastly mining of Test cases. The introduction of agents can bring enhancement to the proposed Approach. Thus it will facilitate the mining and knowledge extraction from test cases and this technique can be utilized with automated software testing. This new approach gives a new direction to automated software testing using data mining and hence a novel approach in software testing.

6. FUTURE SCOPE

More new approaches can be researched to enhance the mining and knowledge extraction from test cases and thus making software testing automation more and more powerful.

REFERENCES

[1] J. Whittaker, "What Is Software Testing? And Why Is It So Hard?", IEEE Software,Vol. 17, No. 1, January/February 2000.
[2] M. Vanmali, M. Last and A. Kandel, "Using a Neural Network in Software Testing Process", International Journal of Intelligent Systems., Vol. 17, No. 1, pp. 45-62, January 2002.
[3] Dave Kelly ,"Software Test Automation and the Product Life Cycle", Symantec Corporation, Volume Number: 13 (1997),Issue Number: 10
[4] Hughes Software Systems Ltd. Test Automation, http://www.hssworld.com/whitepapers/whitepap er_pdf/test_automation.pdf, December 2002.
[5] Pfleeger, S.L. Software Engineering: Theory and Practice.2nd Edition, Prentice-Hall, 2001.
[6] Kaner, C., Falk, J., Nguyen, H.Q. Testing Computer Software. Wiley, 1999.
[7] El-Ramly, M. Stroulia, E. Sorenson, " From Run-time Behavior to Usage Scenarios: An Interaction-pattern Mining Approach.",Proceedings of KDD-2002 (Edmonton, Canada, July 2002), ACM Press, 315 – 327.
[8] Schroeder P. J., and Korel B, "Black-Box Test Reduction Using Input-Output Analysis", Proc. of ISSTA '00, 173-177, 2000.
[9] Last M., and Kandel, A.,"Automated Test Reduction Using an Info-Fuzzy Network. to appear in Annals of Software Engineering", Special Volume on Computational Intelligence in Software Engineering, 2003 .
[10] Last M., and Maimon, O." A Compact and Accurate Model for Classification",IEEE Transactions on Knowledge and Data Engineering, 2003.
[11] Maimon O., and Last, M.,"Knowledge Discovery and Data Mining – The Info-Fuzzy Network (IFN) Methodology",Kluwer Academic Publishers, Massive Computing, Boston, December 2000.
[12] Mark Last,"Data Mining For Software Testing",Ben-Gurion University of the Negev,2005

[13] E.M.H. Lee and K.C.C. Chan (PRC) , "A Data Mining Approach to Discover Temporal Relationship among Performance Metrics from Automated", http://www.actapress.com/Abstract.aspx?paperId=23837

[14] Ilkhani, A.; Abaee, G.; "Extraction test cases by using data mining; reducing the cost of testing ",Dept. of

Comput. Sci., Shahabdanesh Inst. of Higher Educ., Tehran, Iran , 8-10 Oct. 2010

Ms. Sarita Sharma did her Master in Computer Application from IGNOU, New Delhi, India, She did her M.phil (Computer Science) from Ch. Devi Lal University,Sirsa , India and is pusuing Ph.D from Singhania University,Rajasthan, India. She is presently working as Associate Professor in Deptt. Of Computer Science, DAV Institute of Management , India. She has guided more than 90 students in their Projects and has published a number of papers in National/International journals. She is a member of Computer Society of India. Her areas of interest include Software Engineering, Data Mining, Relational Databases, Computer Languages etc. She has about 15 years of teaching experience.

Ms. Anamika Sharma did her Master in computer Science from Gurukul University Haridwar in 1998 and pursuing Ph.D in computer Science from Singhania University Rajasthan. She is having about 13 years of teaching experience of postgraduate courses. She has guided more than 70 students in their project and published number of papers in national/International journals. She is a member of Computer Society of India .Her main Area of research include Computer Graphics, Data mining, Software testing and Quality assurance and object oriented Analysis and Design. At Present she is working in DAV Institute of Management, Faridabad as Associate Professor in Computer Science Dept.

Simulation of Thermal Comfort of a Residential House

Masine Md. Tap[1], Haslinda Mohamed Kamar[2], Abdul Kadir Marsono[3] Nazri Kamsah[4] and Khairul Amry Mohd Salimin[5]

[1] **Department of Manufacturing and Industrial Engineering, Faculty of Mechanical Engineering, Universiti Teknologi Malaysia, Skudai, Johor 81310, Malaysia**

[2] **Department of Thermo-Fluid, Faculty of Mechanical Engineering, Universiti Teknologi Malaysia Skudai, Johor 81310, Malaysia**

[3] **Department of Structural Engineering, Faculty of Civil Engineering, Universiti Teknologi Malaysia Skudai, Johor 81310, Malaysia**

[4] **Department of Thermo-Fluid, Faculty of Mechanical Engineering, Universiti Teknologi Malaysia Skudai, Johor 81310, Malaysia**

[5] **Faculty of Mechanical Engineering, Universiti Teknologi Malaysia Skudai, Johor 81310, Malaysia**

Abstract

In hot and humid climates thermal comfort can become a problem to the occupants of many residential buildings especially when they are not equipped with air-conditioning system. This paper presents outcomes of an ongoing research work to investigate thermal comfort level in a naturally ventilated residential house in Malaysia using computational fluid dynamics (CFD) method. Actual measurements of the temperature distribution, relative humidity and air flow pattern were conducted. CFD simulations on the model of the house allow us to visualize the temperature distribution and air flow pattern and velocity in the house. The thermal comfort in the house was found to be well outside the limits specified by ASHRAE standards. CFD simulation was used to investigate the effects of using a ceiling fan installed in the middle of the hall section and rotating at 150 RPM. It was found that the fan produced swirling flow pattern in the hall section resulting in a more uniform temperature distribution inside the house. However, there is no significant improvement in the thermal comfort level in the house. Results of CFD simulations also show that the use of small extractor fans installed on the front and back walls has no significant effects on the thermal comfort level in the house. Although the mechanical ventilation devices did not help improve the thermal comfort in the house being studied, the CFD simulation results can be used by building designers and engineers to further improved the level of thermal comfort in residential houses in hot and humid climates that are naturally ventilated.

Keywords: *Thermal Comfort in a Residential Building, Computational Fluid Dynamics, Naturally Ventilated Residential House.*

1. Introduction

In hot and humid climates, thermal comfort in both residential and commercial buildings is essential. Most residential buildings are not equipped with air-conditioning systems. Instead they rely mostly on natural ventilation, passive cooling system and mechanical ventilation devices such as ceiling and extractor fans to achieve certain level of thermal comfort. Special attention must be given to the design and installation of these devices so as to optimise their effects to thermal comfort. Moreover, the quality and energy efficiency of these devices varies widely.

It is therefore critical to assess several important building characteristics at the design stage. These include the ability to improve energy efficiency, effects of solar radiation, effects of air flow due to wind and the most importantly the occupant's comfort. To assess the thermal comfort, it is necessary to analyse the air velocity, temperature distribution, and relative humidity of the air in the interior space of the building. Standards that may be used to evaluate thermal comfort are widely available and numerical methods such as computational fluid dynamics (CFD) may be utilized to assist in the analysis.

A CFD software such as Fluent is a useful tool that can be used to create a virtual model of the building interior and simulate air flow, temperature profile and humidity which are directly related to thermal comfort, before the actual construction can be done. Modifications to an existing building can also be simulated using the CFD method prior to any physical renovations.

2. Human Thermal Comfort

Human thermal comfort is defined as the state of mind that expresses satisfaction with the surrounding environment [1]. Maintaining the thermal comfort for occupants of buildings or other enclosures is one of the important goals of the Heat, Ventilation and Air-conditioning (HVAC) design engineers. Thermal comfort in buildings is affected by the transfer of heat energy by conduction, convection, radiation, and evaporative heat loss. Thermal comfort will be maintained when the heat generated by human metabolism is allowed to dissipate, thus a person will maintain thermal equilibrium with his surroundings. Any heat gain or loss beyond this level will generate a sensation of discomfort. The sensation of feeling hot or cold is not just dependent on air temperature alone. Other factors that can affect human thermal comfort are the relative humidity of the ambient air, the air movement pattern and its velocity, radiant heat exchange, metabolic rate of a person and the person's clothing [2].

The thermal comfort condition in a ventilated building may be assessed through the air flow pattern and its velocity, the temperature distribution and the indoor air quality [3]. A good indoor climate will not only make its occupants comfortable but also promote the energy saving and its sustainability [4]. Natural ventilation implies that air is supplied and removed from the indoor space of a building by natural means. The effectiveness of natural ventilation therefore depends very much on the design features of the house [5]. Natural ventilation is usually coupled with the use of mechanical ventilation system such as extractor and ceiling fans to provide better ventilation and thus thermal comfort condition.

Computational fluid dynamics (CFD) is a branch of fluid mechanics that uses numerical methods and algorithms to solve and analyze problems that involve fluid flows. Computers are used to perform the calculations required to simulate the interaction of liquids and gases with surfaces defined by boundary conditions. With high-speed supercomputers, better solutions can be achieved. Many researchers have used CFD to analyze thermal comfort in building spaces and investigating the effects of natural and stratified ventilations on the thermal comfort. Some used this technique to improve the efficiency of energy usage for the building. These are very difficult to be carried out by using other methods [6]. Bastide *et.al.* [7] used CFD to predict building energy efficiency. Serra & Semiao [8] used CFD to evaluate and compare between two different ventilation strategies and Stravrakakis *et.al.* [9] used CFD to evaluate the effect of window sizes to thermal comfort and hygiene in buildings.

This paper presents a study on thermal comfort in a single-storey terrace residential house in Malaysia, which is not furnished with an air-conditioning system. The goal of this study is to assess the level of thermal comfort in the house when it is naturally ventilated and when a ceiling fan is used together with the natural ventilation. Fluent CFD software was used to construct a virtual model of the house. CFD simulations were performed to predict the air flow and temperature distributions in the house. The simulation procedure was validated by comparing results of the CFD analysis with data from actual measurement. The effects of a ceiling fan, installed in the hall section of the house and extractor fans installed at the front and rear walls of the house, on the air flow and temperature distributions were also investigated.

3. Thermal Comfort Study of a Residential House

Figure 1 shows a virtual model of the interior regions of a residential single-storey terrace house we used as a case study developed using Fluent CFD software. The house comprises of several regions namely the hall, stack and kitchen. The house is not furnished with air-conditioning system and is naturally ventilated, i.e. both the front and rear doors opened.

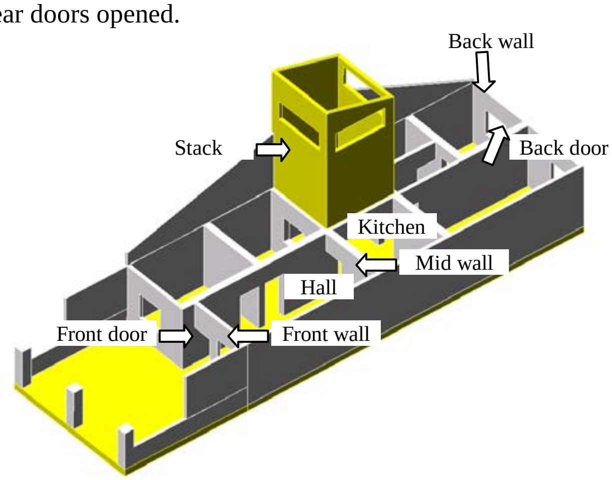


Fig. 1 CFD model of a residential house considered in this study.

To assess the level of thermal comfort in the house, we carried out measurements to acquire the average temperature of the ambient air, the relative humidity of the air and the average velocity of the air inside the house. The data were recorded during a time span from 12AM to 9PM. The data was compared with the upper and lower limits of thermal comfort specified by the ASHRAE standard [1] in order to assess the level of thermal comfort in the house.

3.1 Comfort Conditions with Natural Ventilation

Figure 2 shows variation of the average temperature of the ambient air inside the various sections of the house when the house is naturally ventilated. The air temperature in all sections decreases from about 29°C at 12AM to about 27°C at 6AM. Thereafter, the air temperature increases and reaching the highest value of about 31°C at 6PM. Also shown in Figure 2 are limits of air temperature for acceptable level of thermal comfort, as specified by the ASHRAE standard [1]. It can be seen that at any given time, the average temperature of the ambient air inside the house is well outside the limits of acceptable thermal comfort.

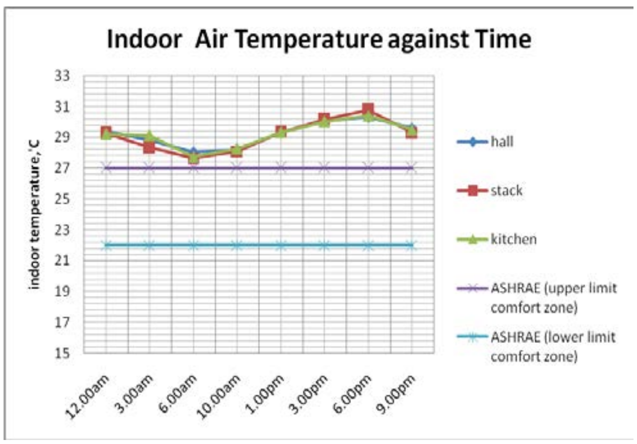


Fig. 2 Variation of average temperature of the air inside the house.

Figure 3 shows variation of the relative humidity of the ambient air inside the house. Also shown in this figure are the limits of relative humidity for acceptable level of thermal comfort, as specified by the ASHRAE standard [1]. The figure shows that the relative humidity of the air fluctuates within the range from about 71% at 10AM to 81% at 3AM. At any given time, the ambient air inside the house is too humid and this condition would cause thermal discomfort to the occupants.

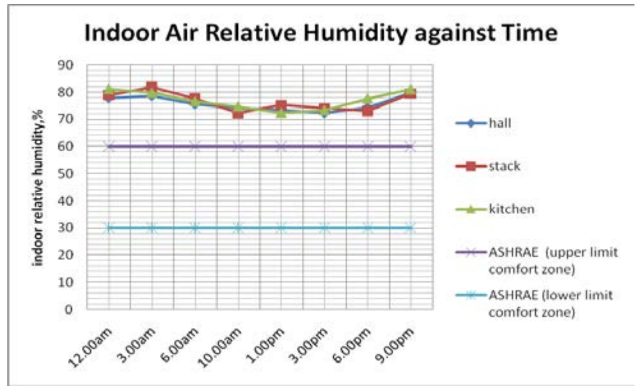


Fig. 3 Variation of relative humidity of the air inside the house.

The variation of average air velocity in various sections of the house is shown in Figure 4. Also shown in this figure are the limits of air velocity for acceptable level of thermal comfort, as specified by ASHRAE standard [1]. It can be seen that the air velocity in the hall and stack sections is fairly constant at about 0.1 m/s from 1PM to 6PM. The air velocity in the kitchen area appears to fluctuate from 0.1 m/s to 0.2 m/s. At any given time, the average velocity of the air inside the house is below the acceptable level of thermal comfort specified by ASHRAE standard [1].

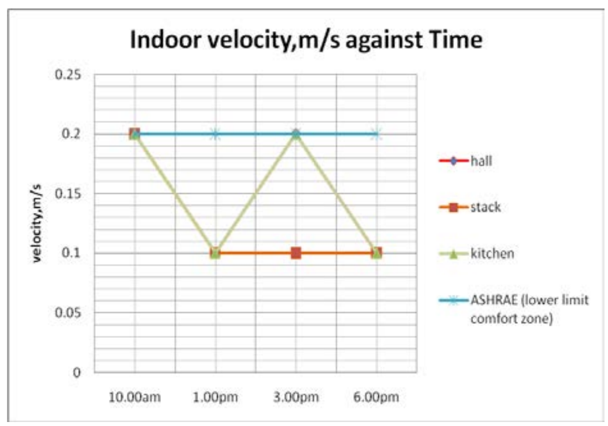


Fig. 4 Variation of average velocity of the air inside the house.

3.2 Comfort Conditions with Combined Natural Ventilation and a Ceiling Fan

The above findings clearly show the thermal comfort in the house is well below a satisfactory level when the house is ventilated naturally. Clearly, some means of increasing the comfort level is required. The simplest way is by the use of a ceiling fan. We then turned on the ceiling fan which was installed in the hall section of the house, during the entire data collection period. After a steady-state condition was established, we repeated the measurements for the average temperature of the ambient air, relative humidity of the air and the average velocity of the air in the various sections of the house. The goal is to investigate the effects of the use of the ceiling fan on these thermal comfort parameters.

Figure 5 shows the variation of average temperature of the air when the ceiling fan was turned on. It can be seen that the air temperature in all regions of the house decreases from about 30°C at 12AM to about 27°C at about 10AM. Thereafter, the temperature increases as more sensible heat is gained by the air, reaching the highest value of about 31°C at 6PM. With the ceiling fan turned on, the temperature of the air inside the house is still outside the thermal comfort limits specified by ASHRAE [1].

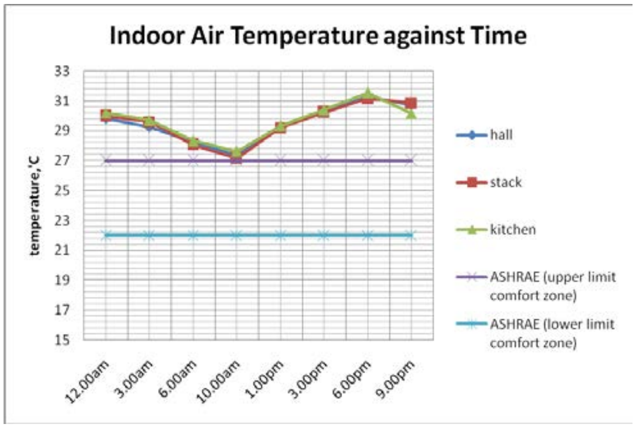


Fig. 5 Variation of average temperature of the air inside the house when the ceiling fan was turned on.

The variation of relative humidity of the air inside the house when the ceiling fan was turned-on is shown in Figure 6. It is seen that the relative humidity of the air in all sections of the house fluctuates during the data collection period, reaching the highest value of 80% at 1PM and the lowest value of 65% at 6PM. The relative humidity of the air is also outside the thermal comfort limits specified by ASHRAE standard [1].

The variation of air velocity inside the house with the ceiling fan turned on is shown in Figure 7. It is seen here that the air velocity in the hall section is still above the limit of thermal comfort level specified by ASHRAE standard [1]. However, the air velocity in the stack section appears to be below the comfort level limit. The air velocity in the kitchen section is in the acceptable level of thermal comfort.

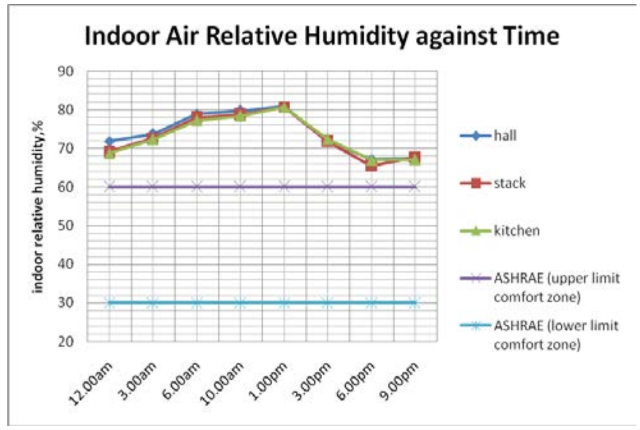


Fig. 6 Variation of relative humidity of the air inside the house when the ceiling fan was turned on.

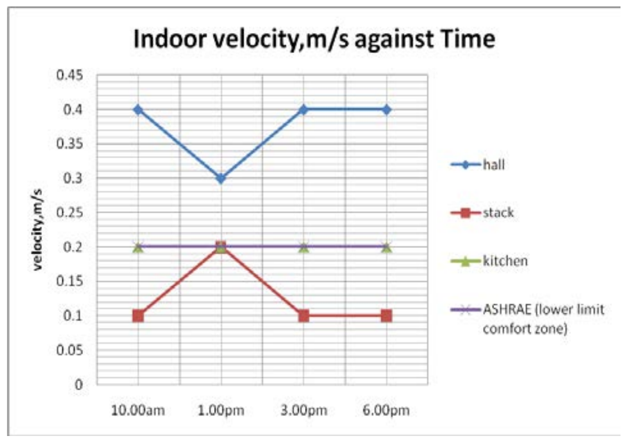


Fig. 7 Variation of air velocity inside the house when the ceiling fan was turned on.

The above results show that although a ceiling fan was turned on during the entire data collection period, the thermal comfort conditions in the house were still outside the recommended acceptable level specified by the ASHRAE standard [1]. Clearly, some other means of mechanical ventilation need to be implemented in order to further improve the thermal comfort conditions inside the naturally ventilated house.

In the next section, we describe the use of computational fluid dynamics (CFD) method to model and simulate the thermal comfort conditions of the house. Our goal is to predict the temperature distribution and velocity profile of the ambient air inside the house for two conditions. First is when the house was naturally ventilated and secondly is when the ceiling fan was turned on. We used the thermal comfort data we collected to verify the CFD modeling and simulation procedure. Once verified, we then used the model to investigate if the thermal comfort conditions in

IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 2, September 2011
ISSN (Online): 1694-0814
www.IJCSI.org

the house could be improved by the use of other mechanical ventilation method.

4. CFD Analysis of the Residential House

Figure 1 shows the virtual model of the residential house we considered in this study. The flowchart shown in Figure 8 summarizes the procedure we employed in performing the CFD modeling and simulating the thermal comfort conditions inside the house.

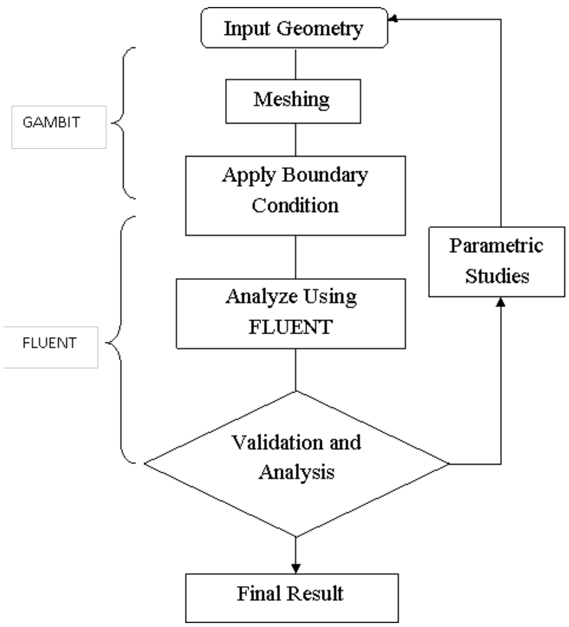


Fig. 8 CFD modeling and simulation procedure.

Three CFD models of the house were developed in this study. The first model represents the house with natural ventilation. The second model represents the house with natural ventilation and furnished with a ceiling fan, installed in the middle of the hall section. The third model represents the house furnished with the ceiling fan and two extractor fans. One extractor fan was installed on the front wall near the door while the other on the real wall, also near the door. The CFD simulations were performed on these models to investigate the distribution of air temperature and air velocity inside the house, under the given ventilation conditions.

4.1 CFD Simulation on Naturally Ventilated House

We first performed CFD simulation on the model which represents the house that is naturally ventilated. After meshing the model, we imposed the boundary conditions (BCs) on the model that represent the natural ventilation conditions, namely with the front door, middle door and back door opened. In addition, there is an opening near the top of the stack, on the front wall to allow outside air to flow into the house. These boundary conditions are shown in Table 1. The values of all parameters were taken from the actual measurement from the house.

Figure 9 shows the distribution of air velocity obtained from the CFD simulation. With natural ventilation, the air is seen to flow in a streamline condition from the front door towards the door on the middle partition wall, with decreasing velocity. It appears that air circulation also occurs in the middle of the hall but there is almost no air flow near wall far from the door. The air flows from the middle door towards the back door also in a streamline condition and with increasing velocity. Uneven flow condition is seen in the stack section and in the kitchen, in the area close to the rear wall.

Table 1 Boundary conditions for the house with natural ventilation.

Types of BCs	Zone	Parameters
Inlet Air Velocity	Front door	Velocity = 0.4 m/s Temperature = 29.6ºC
	Stack	Velocity = 0.4 m/s Temperature = 29.8ºC
Outlet Air Pressure	Back Door	Pressure = 101 kPa Temperature = 29.3ºC
Wall Thermal Conditions	Stack	Heat gain = 10 W/m^2 Temperature = 30ºC
	Front wall	Heat gain = 29.8 W/m^2 Temperature = 29.3ºC
	Back wall	Heat gain = 19.8 W/m^2 Temperature = 29.4ºC

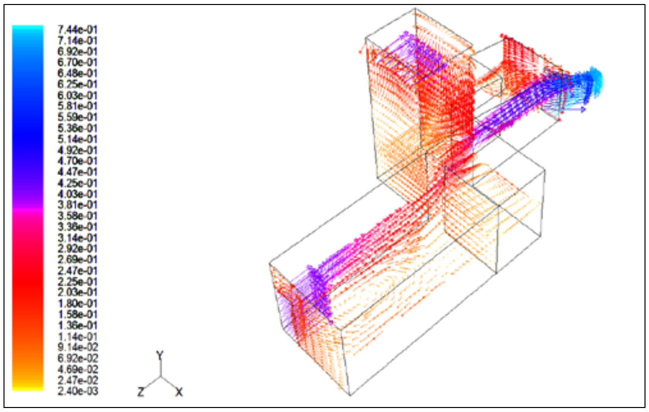


Fig. 9 Distribution of air velocity (in m/s) when the house is naturally ventilated.

Figure 10 shows the contour of temperature distribution on the walls and the air inside the house. The temperature distribution for the air is taken at a height of 1.5 m above the floor level. It can be seen that the air temperature in the hall is almost uniform at about 302K (29°C). However, the air temperature is slightly higher close to the front wall. The air temperature in the kitchen is also quite uniform at about 303K (30°C). There is a significant temperature variation on the front wall surface with the highest temperature is about 305K (32°C). This is due to a high heat gain through this wall. The front wall of the stack and back wall of the house are at a fairly uniform temperature of about 303K (30°C).

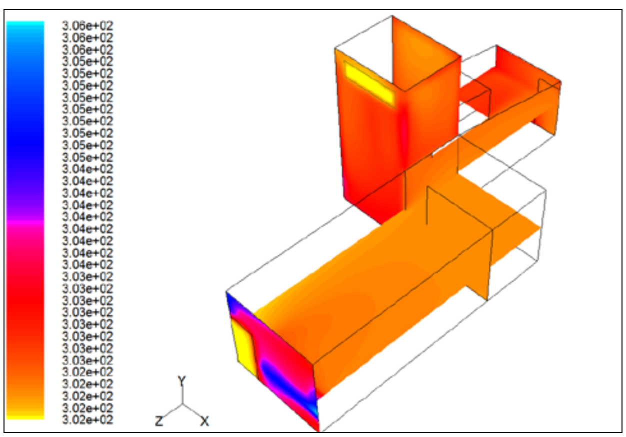


Fig. 10 Temperature (in Kelvin) distribution in the house when it is naturally ventilated.

4.2 CFD Simulation on Naturally Ventilated House with Ceiling Fan

Next, we performed similar CFD simulation on the second model of the house, which represents the naturally ventilated house with the ceiling fan turned on. The goal is to investigate the effects of the ceiling fan on the air temperature and velocity distributions. The boundary conditions prescribed on this model are similar to those given in Table 1. An additional boundary condition was prescribed on the model to represent the ceiling fan that was installed in the middle of the hall and rotating at a speed of 150 RPM.

Figure 11 shows the distribution of air velocity in the house. It can be seen that the ceiling fan had caused swirling air flow in the hall section. The swirling flow intensity appears to be highest in the area closed to the front wall. The air flow velocity increases compared to the natural ventilation condition, where the highest air velocity is about 1.3 m/s near the front wall. Swirling air flow condition can also be seen in the area closed to the middle partition wall. The air is seen to flow in a streamline condition from the door on the partition wall towards the back door, with a uniform velocity. Lower air velocity with uniform pattern occurs in the stack section. Vortex flow condition can also be seen in the kitchen, in the area closed to the wall, near the door.

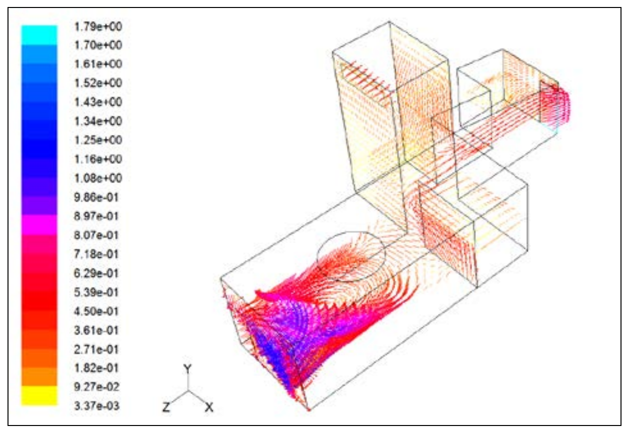


Fig. 11 Distribution of air velocity (in m/s) when the ceiling fan was turned on.

Figure 12 shows the temperature distribution in the house when the ceiling fan was turned on. It is seen that the temperature distribution in the hall appears to be uniform at about 302K (29°C). The temperature in the kitchen also appears to be uniform at about 303K (30°C). There are no more high temperature spots on the front wall when the fan is turned on. The temperature distribution on the back wall and front side of the stack is also uniform at about 303K (30°C).

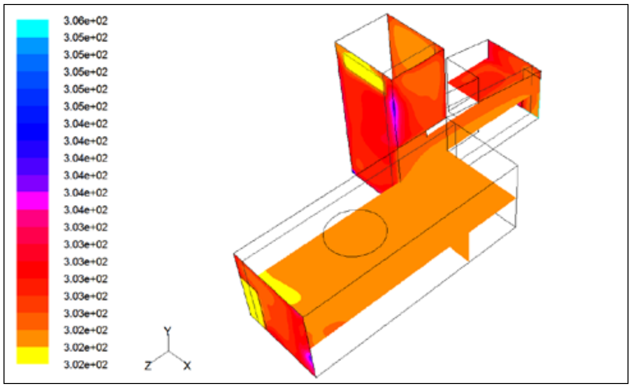


Fig. 12 Temperature (in Kelvin) distribution in the house when the ceiling fan was turned on.

The above observation shows that when the ceiling fan was turned on, vortex and swirling air flow conditions were created in the hall area. The fan also increases the air flow velocity everywhere else inside the house. But the air velocity is much higher than the level specified by ASHRAE standard [1] which could cause discomfort to the occupants. However, the ceiling fan does help to make the temperature distribution in the house to become more uniform, particularly in the hall section.

4.3 CFD Simulation on Naturally Ventilated House with Ceiling and Extractor Fans

Finally we perform CFD simulation on the third model of the house. This model represents the naturally ventilated house furnished with a ceiling fan and two extractor fans which are installed on the front and back wall, near the top of the doors. We anticipated that the extractor fans will improve the air flow and circulation conditions inside the house, thus improving the temperature distribution. We prescribed the same boundary conditions as those for the second model. In addition, we introduced extractor fans by prescribing little openings on the front and back walls, through which air at a pressure of 101 kPa and temperature of 29°C flows with a velocity of 1.4 m/s.

Figure 13 shows the distribution of air velocity inside the house when the extractor fans were turned on along with the ceiling fan. It can be observed that the extractor fan installed on the front wall does not seem to have significant effects on the flow condition and velocity of the air inside the hall section. The amount of air and velocity of the air blown by this fan are probably too small to have any significant effects on the swirling flow condition of the air created by the ceiling fan. Nevertheless the extractor fan installed on the back wall appears to affect the air flow condition in the stack and kitchen sections.

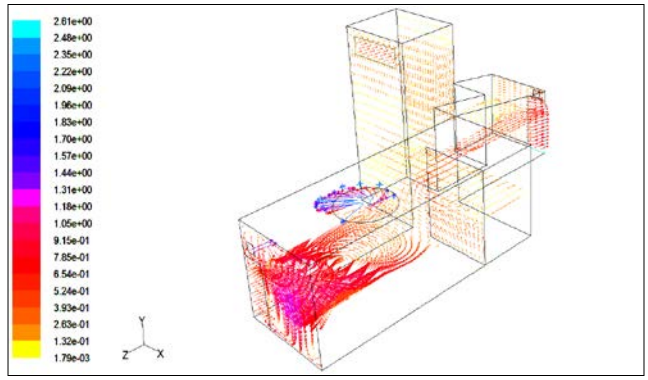


Fig.13 Distribution of air velocity inside the house when the ceiling fan is used along with the extractor fans.

Figure 14 shows the distribution of air temperature inside the house, on a plane 1.5 m above the floor. It can be seen that the temperature distribution inside the hall section remains uniform at about 302K (29°C) and appears to be unaffected by the air blown by the extractor fan on the front door. The temperature distribution on the front and back wall appears to the more uniform with a couple of high temperature spots occur near the corners of the house. High temperature spot can also be seen at the bottom corner of the stack. The front wall, back wall and stack wall show higher temperature than the air inside the house due to the heat gain from the surrounding.

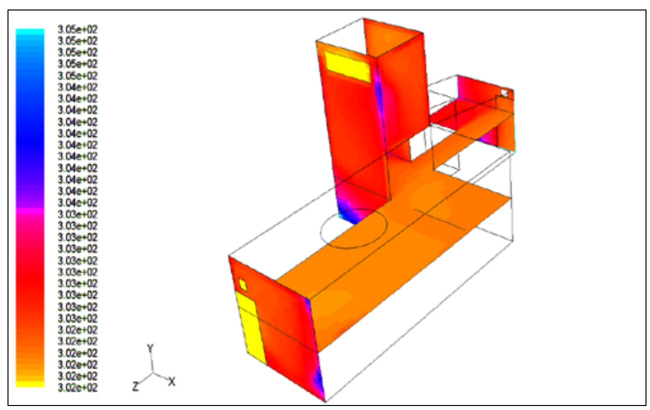


Figure 14 Distribution of air temperature inside the house when the ceiling fan is used along with the extractor fans.

The above finding suggests that the use of small extractor fans installed on the front and back walls closed to the door has almost negligible effects on the temperature distribution in the house. The flow condition and velocity of the air in the hall section are also unaffected by the extractor fans. Both parameters are still outside the range of thermal comfort level specified by ASHRAE [1].

5. Conclusion

A thermal comfort study on a naturally ventilated single-storey terrace residential house was presented. Actual measurements on the air temperature, relative humidity and air flow velocity show that the thermal comfort in the house is well outside the recommended limits specified by the ASHRAE standard. CFD simulations were conducted on the representative models of the house. Using this method, we were able to observe the temperature distribution and air flow conditions inside the house when it was naturally ventilated. The use of a ceiling fan, installed in the middle of the hall section and rotating at 150 RPM, produces a swirling flow of the air in the hall section. The distribution of air temperature becomes more uniform almost everywhere inside the house. However, the level of thermal comfort in the house still remains outside the recommended level. Results of CFD analysis also show that the use of two extractor fans, installed on the front and back wall respectively, has insignificant effects on the temperature distribution as well as air flow pattern inside the house. The research will continue to find the effective methods of improving the thermal comfort in the house.

Acknowledgments

The authors would like to acknowledge Universiti Teknologi Malaysia for supporting this research through the Research University Grant scheme.

References

[1] American Society of Heating, Refrigerating and Air-Condition Engineers, PIA ANSI/ASHRAE 55-2004, Atlanta, GA: ASHRAE, 2004.

[2] S. Mohamed, and K. Srinavin, “Thermal environment effects on construction workers’ productivity”, Work Study, Vol. 51, No. 6, 2002, pp. 297-302.

[3] A. Stamou, and I Katsiris, “Verification of a CFD model for indoor airflow and heat transfer”, Building and Environment, Vol 41, 2006, pp. 1171-1181.

[4] J.F. Nicol, and M.A. Humphreys, “Adaptive thermal comfort and sustainable thermal standards for buildings”, Energy and Buildings, Vol 34, 2002, pp. 5563-572.

[5] A. T. Balasbaneh, “Hot climate air flow study and effect of stack ventilation in residential building”, MSc Thesis, Faculty of Civil Engineering, Universiti Teknologi Malaysia, Johor Bahru, Malaysia, 2010.

[6] N. Djongyang, R. Tchinda, and D Njomo, “Thermal Comfort: A review paper”, Renewable and Sustainable Energy Reviews, Vol. 14, 2010, pp. 2626-2640.

[7] A. Bastide, P. Lauret, F. Garde, and H. Boyer, “Building energy efficiency and thermal comfort in tropical climates: Presentation of a numerical approach for predicting the percentage of well-ventilated living spaces in buildings using natural ventilation”, Energy and Buildings, Vol. 38, 2006, pp.1093-1103.

[8] N. Serra, and V. Semiao, “Comparing displacement ventilation and mixing ventilation as HVAC strategies through CFD”, Engineering Computation: International Journal for Computer-Aided Engineering and Software, Vol. 26, No. 8, 2009, pp. 950-971.

[9] G.M. Stavrakakis, D.P. Karadimou, P.L. Zervas, H. Sarimveis, N.C. Markatos, “Selection of window sizes for optimizing occupational comfort and hygiene based on computational fluid dynamics and neural networks”, Building and Environment, Vol. 46, 2011, pp. 298-314.

Masine Md. Tap, received her Bachelor’s Degree in Mechanical Engineering from Universiti Teknologi Malaysia in 1986, MPhil in Computer Aided Engineering from Herriot-Watt University, United Kingdom in 1989 and PhD. from Dundee University, United Kingdom in 1999. She is now an associate professor in the Faculty of Mechanical Engineering, Universiti Teknologi Malaysia. Her areas of interest are industrial engineering, work design and operations research. She is a co-inventor for Industrial Building System (IBS) product at national and international level.

Haslinda Mohamed Kamar, received her Bachelor’s Degree in Mechanical Engineering from University of Glasgow, Scotland in 1993, Master and PhD. in Mechanical Engineering from University Teknologi Malaysia in 1997 and 2009, respectively. She is now a senior lecturer in the Faculty of Mechanical Engineering, Universiti Teknologi Malaysia. Her areas of interest are automotive air-conditioning system, thermal comfort & energy efficiency in hot climates, and natural ventilation as passive cooling strategy in buildings.

Abdul Kadir Marsono, received his Bachelor of Engineering (civil) from Universiti Technologi Malaysia in 1985, Master of Philosophy for Heriot University in 1989 in the decipline of structural engineering and PhD in Structural Engineering from Dundee University in the year of 2000. Currently he is an associate professor in structural engineering, taught Information technology and tall building system analysis and design. He carry out research in non-linear analysis of reinforced concrete of tall buildings and industrialized building system (IBS) as well as sustainable product design for civil engineering. He is a principal patent inventor for Industrial Building System

(IBS) product at national and international level.

Nazri Kamsah received his Bachelor's Degree in Mechanical Engineering from University of Sunderland, United Kingdom in 1983, Masters of Engineering from Universiti Teknologi Malaysia in 1988, and PhD in Mechanical Engineering from University of New Hampshire, USA in 2001. He is currently a senior lecturer in the Faculty of Mechanical Engineering, Universiti Teknologi Malaysia. His areas of interest include computational solid mechanics modeling and simulation, thermal management in microelectronics, thermal comfort & energy efficiency in hot climates, natural ventilation as passive cooling strategy in buildings and computational fluid dynamics (CFD) modeling and simulation.

Khairul Amry Mohd Salimin, received his Bachelor's Degree in Mechanical Engineering from Universiti Teknologi Malaysia in 2011.

A Novel Data Encryption Technique by Genetic Crossover of Robust Finger Print Based Key and Handwritten Signature Key

Tanmay Bhattacharya[1], Sirshendu Hore [2] and S. R. Bhadra Chaudhuri [3]

[1] **Assistant Prof, Dept. of IT, JIS College Engineering, Kalyani, West Bengal, India.**

[2] **Assistant Prof, Dept. of CSE, Hooghly Engineering & Technology College, Pipulpati, Hooghly, West Bengal, India.**

[3] **Professor, Dept. E&TC, Bengal Engineering & Science University, Shibpur, Howrah, West Bengal, India.**

Abstract

Unimodal biometric systems have to contend with a variety of problems such as noisy data, intra-class variations, restricted degrees of freedom, non-universality; phishing attacks spoof attacks, and high false acceptance rates. In order for the biometrics to be ultra-secure and to provide more-than-average accuracy, more then one form of biometric identification is required. Hence some of these limitations can be addressed by a combination of different biometric recognition technologies that integrate the evidence presented by multiple sources of information. In the proposed work Fingerprint Key and Signature Keys are generated from the Fingerprint and Handwritten Signature of the legitimate user. The system is quite robust because it is trained by Artificial Neural Network and Machine Intelligence. Two Combined Keys are generated by Genetic crossover of those two keys. Finally by interleaving the combined keys Encryption Key is generated. In this approach there is a significant improvement over the traditional Unimodal biometric authentication techniques.

Keywords: *ANN; Minutiae; Center of Gravity;Aspect Ratio; Training; SHA-512, Crossover*

1. Introduction

Biometrics, described as the science of recognizing an individual based on her physiological or behavioral traits, is beginning to gain acceptance as a legitimate method for determining an individual's identity. Different biometrics such as fingerprints, hand geometry, iris, retina, face, hand vein, facial thermogram,signature, voice, etc. to either validate or determine an identity [1].Most biometric systems deployed in real-world applications are unimodal, i.e., they rely on the evidence of a single source of information for authentication (e.g., single fingerprint *or* face). Some of the limitations imposed by Unimodal biometric systems can be overcome by including multiple sources of information for establishing identity [2]. Such systems, known as multimodal biometric systems, are expected to be more reliable due to the presence of multiple, (fairly) independent pieces of evidence [3]. Minutiae are local discontinuities in the fingerprint pattern. All popular AFIS are minutiae-based [4,5,6]. Usually each minutiae is described by the position in the coordinate, the direction it flows and the type, whether it is ridge ending or bifurcation. [7,8, 9,10]. Signature is a socially accepted authentication method and is widely used as proof of identity in our daily life. Signature verification can be classified into two categories: signature verification [11, 12] and offline signature verification [13].All the Signature verification system of today are verification system and the features that are mostly used are Baseline Slant Angle, Aspect Ratio, Center of Gravity.Figure 1 illustrates the structure of Minutiae, center of gravity and Baseline slant angle of handwritten signature.

Figure 1: Minutiae (a) Ridge ending (b) Bifurcation Signature (c) Center of Gravity (d) Slant angel

Artificial Neural Network: Artificial neural networks are constituted of artificial neurons. An ANN is a system consisting of processing elements (PE) with links between them. A certain arrangement of the PEs and links produce a certain ANN model, suitable for certain tasks [14]. A Multi-Layer-Perceptron (MLP) is a kind of feed-forward ANN model consisting of three adjacent layers; the input, hidden and output layers. Each layer has several PEs. Figure 2 illustrates the structure of a MLP

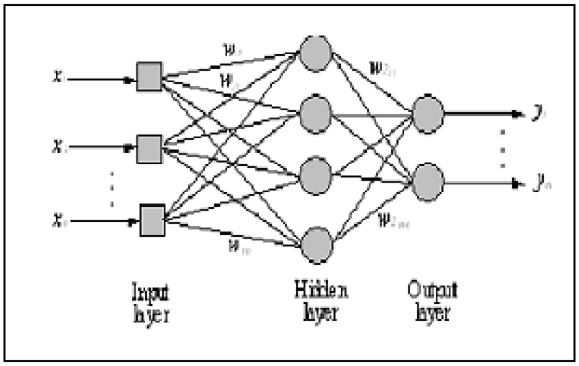


Figure 2: A schematic diagram of a MLP neural network

2. Proposed Algorithm

Following are the main steps of proposed algorithm.

a. Biometric Fingerprint Key Generation
   - Step 1 Image Acquisition
   - Step 2 Enhancement of the Image
   - Step 3 Feature extraction
   - Step 4 Training with different sample images using ANN
   - Step 5 Template Finger print is obtained
   - Step 6 Biometric Key of 512 bit is generated from the given template.

b. Handwritten Signature Key generation

c. Generation of two intermediate keys by genetic crossing over of Biometric key and Signature key.

d. Final Encryption Key generation

e. Encryption of data using Final Encryption Key once fingerprint is match with the template.

f. Decryption of data is done using the Final Encryption key after fingerprint is match with the template.

The sequence of steps for complete authentication process is given in the schematic diagram. Figure 3 illustrates the scheme.

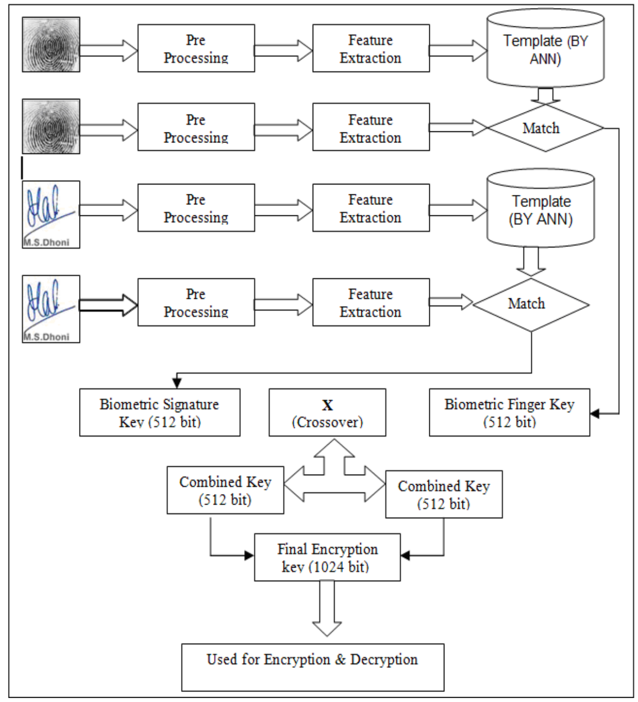


Figure 3: Schematic diagram: The sequence of multimodal authentication process.

3. Explanation of Algorithm

Step 1: In the initial phase the Fingerprint and Signature image are obtained using Bio-sensor scanner which is a flatbed scanner with 600 DPI.

Step 2: Then the image (Fingerprint) is preprocessed to remove the noise using various preprocessing techniques like segmentation, Normalization, Orientation, Ridge frequency estimation and Gabor filtering. Figure 4 illustrate the Image enhancement process.

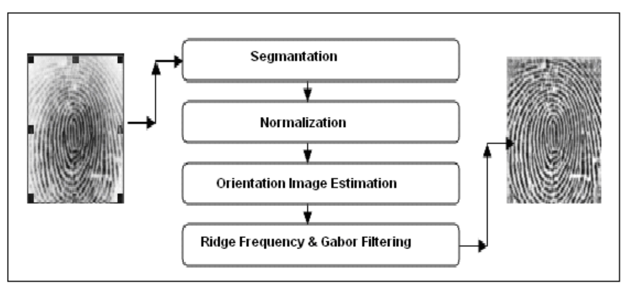


Figure 4: A schematic diagram of a fingerprint Enhancement

Step 3: Enhanced image is then binaries and thinning operation is performed to make the ridges single pixel width. Try to find the location of "1" in the thinned image. The number "1" basically represent the ridges. Taking a '3*3' window mask with 1 as a starting point finding the absolute difference between the center pixel and neighborhoods pixel. If the value is 1(ridge ending) or 3 (bifurcation) then find the angel at which the ridge is moving. Store the coordinates, angles and the calculated values.

Step 4: Before the ANN training the data was divided into three datasets; the training, validation and test. Here data are Minutiae points (ridge ending and bifurcation) which are extracted from a set of fingerprint images. The training set was used to train the MLP, the validation set was used for early-stopping of the training process and the test set was used to evaluate the MLP performance after completion of the training process. The training data set consist of different sample images.

Steps involved:

Forward propagation: *The output of each node in the successive layers is calculated*

O (output of a node) =1/ (1+exp (-ΣW$_{ij}$ x$_i$)) *(a)*

The Error E (Im) of an image pattern Im is calculated with respect to Target (T)

E (Im) =1/2(Σ T (Im)-O (Im)) 2 *(b)*

Reverse Propagation: The error δ for the nodes in the output layer is calculated

δ(output layer) = o(T) - o(Im) *(c)*

The new weights between output layer and hidden layer are updated

W (n+1) =W (n) + ή δ(output layer) *(d)*

The training of the network is stopped when the desired mean square error (MSE) is achieved

E (MSE) = ΣE (Im) *(e)*

Step 5: A Finger Template is created using the training sets .The implementation and simulation were carried out with the aid of neural networks built in function using Matlab (7.5.0) and Java

Step 6: A finger biometric key of length 512 bit is generated using SHA512 hash algorithm. With SHA512 a variable-length message is converted into a fixed-length output of 512 bits. The input message is broken up into chunks of 1024 -bit blocks (sixteen 64-bit little endian integers); the message is padded so that its length is divisible by 1024. The padding works as follows: first a single bit, 1, is appended to the end of the message. This is followed by as many zeros as are required to bring the length of the message up to 128 bits less than a multiple of 1024. The remaining bits are filled up with a 128-bit integer representing the length of the original message; after initialization of SHA512 buffer with a Eight-word buffer (A,B,C,D,E,F,G,H) compute the message digest and finally process message in 16-word blocks to get the output. Figure 5 illustrates the Biometric key generation process. Figure 5 illustrates the Key Generation process with SHA512

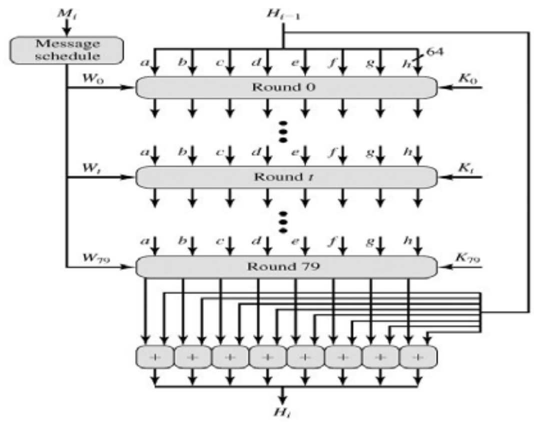


Figure 5: A schematic diagram of key Generation process with SHA512

Step 7: After obtaining the signature image various preprocessing operation are performed to remove the noise caused by the scanner. The image is then cropped, to the bounding rectangle of the signature. Finally transform the signature image from color to grayscale, and to black and white. Figure 6 illustrate the Image enhancement process.

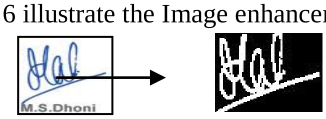


Figure 6: A schematic diagram of a Signature Enhancement

Step 8: From the enhanced signature image we calculate the center of gravity, slant angel also the aspect ratio of the image .stores those value. Training, signature template creation as well as generating the signature biometric key from the signature template is same as step 4, step 5 and step 6.

Step 9: Two Combined keys of 512 bit each are generated by Genetic Crossover of Finger Key and Signature Key. Figure 7 illustrates the Genetic finger biometric key and genetic signature key generation process

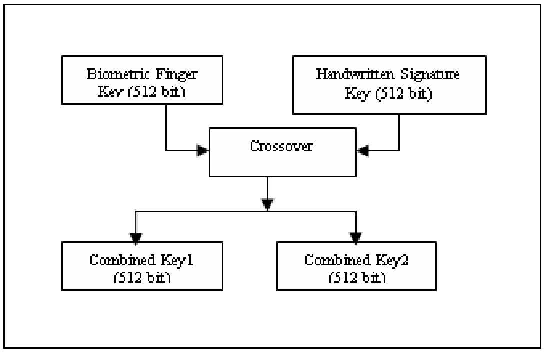


Figure 7: Two combine keys are generated by Crossover

Step 10: A Final key of length 1024 bit is generated by interleaving two Combined Keys obtained from the previous step. Figure 8 illustrates the Final Key generation process.

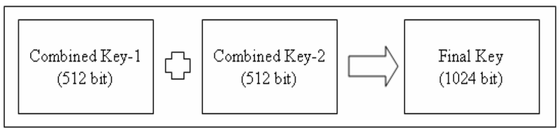


Figure 8: A Final key of length 1024 bit is generated

Step 11: Taking the File from the user encrypts the data using the Final Key. After taking the file from the user store it in a temporary array after converting the character of the file into their corresponding binary format. Stored the binary values in the 8x8 matrix which is filled up row wise where each row corresponds to a single character. Perform columnar transposition on the matrix. Finally perform the bitwise AND operation on the data using Final key .Given file is now encrypted.

Step 12: Taking the Sample fingerprint and Handwritten Signature from the user extract the feature and compare it with the template to find whether the matching score is within the threshold. If it is within the range then generate the Fingerprint Biometric key (512 bit) and Signature Biometric key from the template. The Final key (1024) is generated using the finger biometric key and signature biometric key. Decrypt the encrypted file using the Final Key.

4. Results & Discussions

In this section, we have presented the experimental results of the proposed approach, which is implemented in MATLAB (7.5) and Java (JDK1.6). We have tested the proposed approach with different sets of input images. Initially Fingerprints and Handwritten Signature are scanned using standard Bio-sensor scanner with required resolution. As there can be some imperfection in the capture of images, enhancement has been done followed by extraction of feature (Minutiae points for fingerprint and slant angle, aspect ratio and center of gravity for handwritten signature). Figure 9 illustrate the different stage of the fingerprint image, Handwritten Signature image.

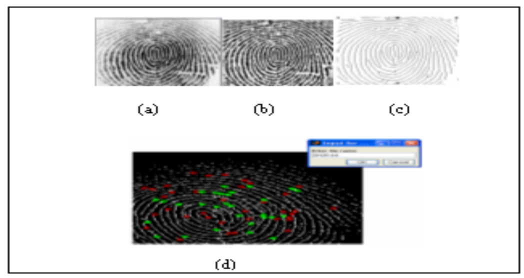


Figure 9: (a) Raw Image, (b) Enhanced Image, (c) Thinned Image and (d) Image with Minutiae points.

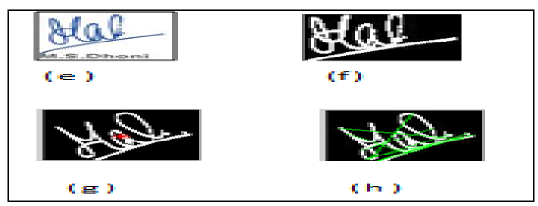


Figure 9: (e) Raw Image (f) Enhanced, Cropped Image (g) Image with Center of gravity (h) Image with Slant Angle

Figure 10 and 12 shows the data that are not match with the template data while figure 11, 13 shows data that are closely matched with the template. Figure 14 shows finger Biometric key generated from template while figure 15 shows Signature Biometric key generated from template and figure 16 shows combine key generation process.

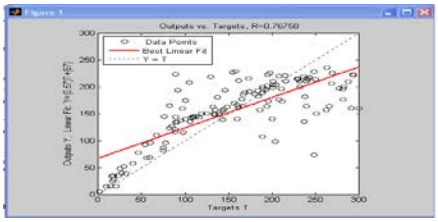


Figure 10: Showing result that is not matched with template (Target output vs. Computed output on Test data)

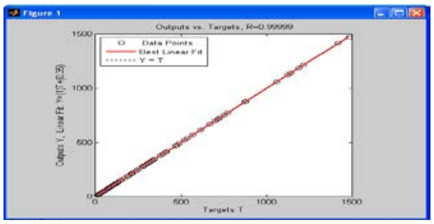


Figure 11: Showing result that is closely matched with template (Target output vs. Computed output on Test data)

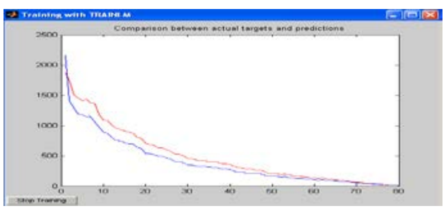


Figure 12: Showing result that is not matched with template (Compersion between Actual data and Predicted data)

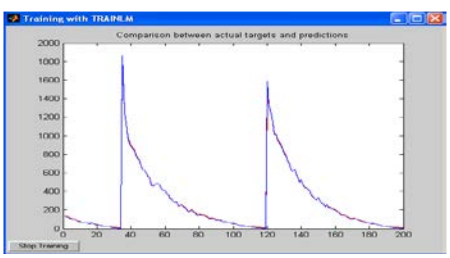


Figure 13: Showing result that is closely matched with template (Compersion between Actual data and Ppredicted data)

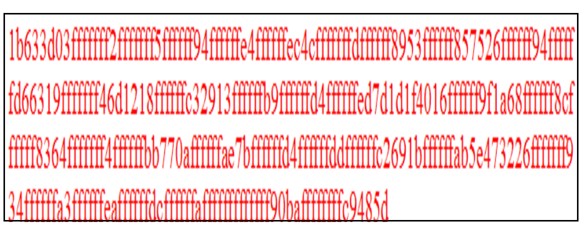


Figure 14: Finger Biometric key generated from template.

Figure 15: Signature key generated from template.

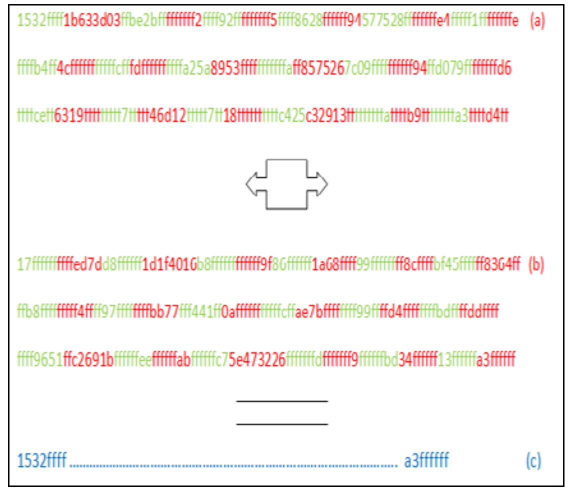


Figure 16: (a) & (b) Combine Key (512 bit) Generations by Crossing over of Signature Key & Finger Print Key (c) Final key (1024 bit)

5. Conclusions

The proposed approach minimizes the shortcomings of Unimodal authentication technique by using ANN. Most of the available applications use full fingerprint but in this approach a portion of the Fingerprint is good enough to generate the biometric key and hence minimizes False Rejection Ratio (FRR).Also in this approach combined the use of handwritten signature technique and fingerprint system is to eliminate the limitation of Unimodal system. So using this approach sensitive data can be made more secure than any traditional technique. Experimental results are also satisfactory. This research has may be further extended using more reliable biometric features.

References

[1] A. K. Jain, A. Ross, and S. Prabhakar, “An introduction to biometric recognition,” IEEE Trans. on Circuits and Systems for Video Technology, vol. 14, pp. 4–20, Jan 2004.

[2] A. Ross and A. K. Jain, “Information fusion in biometrics,” Pattern Recognition Letters, vol. 24, pp. 2115– 2125, Sep 2003.

[3] L. I. Kuncheva, C. J. Whitaker, C. A. Shipp, and R. P. W. Duin, “Is independence good for combining classifiers?,” in Proc. of Int’l Conf. on Pattern Recognition (ICPR), vol. 2, (Barcelona, Spain), pp. 168–171

[4] Zhang Tanghui, Tian Jie, He Yuliang, Yang Xin, "A Combined Fingerprint Matching AlgorithmBased on Similarity Histogram",Chinese Journal of Computers, 2005, Vol.28(10), pp.1728-1733.

[5] Anil Jain, Arun Ross, "Fingerprint Matching Using Minutiae and Texture Features", ICIP, 2001,pp.282-285

[6] A. K. Jain, L. Hong, S. Pantanki and R. Bolle, An Identity Authentication System Using Fingerprints, Proc of the IEEE, vol, 85, no.9,1365-1388, 1997.

[7] Hong, L., Y.Wan and A.K.Jain, 1998. Fingerprint Image Enhancement: Algorithm and performance Evaluation. IEEE. Trabs. PAMI, 20(8): 777-789

[8]Jain, A., Hong, L., Bolle, R.: On-line Fingerprint Verification. IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.19, No.4 (1997) 302–313.

[9]Anil K. Jain, Salil Prabhakar, Lin Hong "A Multichannel Approach to Fingerprint Classification", IEEE Trans on. PAMI, 1999, Vol.21 (4), pp.348-359

[10]Tanmay Bhattacharya, Sirshendu Hore, Ayan Mukherjee ,S. R. Bhadra Chaudhuri “A Novel Highly Secured Session Based Data Encryption Technique Using Robust Fingerprint Based Authentication” Advances in Networks & Communications, CCSIT Part 2 (2011) pp 422-431, Springer

[11] A. Jain, F. Griess, and S. Connell. On-line signature verification. Pattern recognition, 35(12):2963 2972, 2002

[12] V. Nalwa. Automatic on-line signature verification. Proceedings of the IEEE, 85(2):215– 239, 1997

[13] M. Kalera, S. Srihari, and A. Xu. Offline signature verification and identification using distance Statistics, 2004.

[14]B.Jayaraman, C.Puttamadappa ,E.Anbalagan ,E.Mohan and Srinivasarao Madane, Fingerprint Authentication using Back-propagation Algorithm of International Journal of Soft Computing 3(4) :282-287, 2008 ISSN:1816-9503

Fast Hybrid PSO and Tabu Search Approach for Optimization of a Fuzzy Controller

Nesrine TALBI[1,2] and Khaled BELARBI[2]

[1] Electronic Department, Jijel University
Jijel, ALGERIA

[2] Electronic Department, Mentoury University
Constantine, ALGERIA

Abstract
In this paper, a fuzzy controller type Takagi_Sugeno zero order is optimized by the method of hybrid Particle Swarm Optimization (PSO) and Tabu Search (TS). The algorithm automatically adjusts the membership functions of fuzzy controller inputs and the conclusions of fuzzy rules. At each iteration of PSO, we calculate the best solution and we seek the best neighbor by Tabu search, this operation minimizes the number of iterations and computation time while maintaining accuracy and minimum response time. We apply this algorithm to optimize a fuzzy controller for a simple inverted pendulum with three rules.
Keywords: *Particle Swarm Optimization, Tabu Search, Fuzzy Controller, inverted pendulum.*

1. Introduction

Evolutionary algorithms, originally designed to optimize the parameters, have been shown to intervene as early in solving a problem. If in the case of optimizing parameters of a controller, the controller structure is given in advance, many studies show that evolutionary algorithms can be used to automatically obtain this structure. They can intervene at the design phase of the control system and optimize both the parameters and structure of the control system.
The particle swarm optimization (PSO) is a technique that evolutionary uses a "population" of solutions candidate to develop an optimal solution of the problem. The degree of optimality is measured by a fitness function defined by user (Clerc et al. 2001, Dutot et al. 2002, Ji et al. 2007, Kennedy and al. 1995 and 2001). The PSO is different from other methods of evolutionary computation in order that members of the public called " particles " are scattered in space of the problem (Kennedy et al. 1995 and 2001).
The behavior of the swarm must be described from the point of view of a Particle (Kennedy et al. 2001, Omran 2004, Van den Bergh 2002, Venter et al. 2002).
From the algorithm, a swarm is randomly distributed in space research each particle also has a random speed.
PSO is easy to implement and there are few parameters to adjust, but it suffers from slow convergence in refined search stage (weak local search ability). In this context, we tried to solve the problem of convergence time with the combination of PSO and Tabu Search(TS) which is easy to implement and has the advantage of saving all the old solutions visited using the principle of memory to avoid backtracking (cyclical).
In this paper, we propose a hybrid PSO-TS algorithm to optimize a fuzzy controller for Takagi-Sugeno zero-order. The controller inputs are the error and derivative of error and the output is the control itself. The algorithm automatically adjusts the triangular membership functions of input and fuzzy singletons of output. This fuzzy controller is used for stabilization of a simple inverted pendulum.
This paper is organized as follows: Section 2 presents the general structure of fuzzy controller to optimize, In Section 3, the PSO is explained. Section 4 summarizes Tabou Search, and the hybrid algorithm of the hybrid Particle swarm optimization and Tabu Search is presented in section 5. Simulation results are presented in Section 6. Finally, Section 7 outlines our conclusions.

2. Fuzzy Logic Controller (FLC) to be optimized

A FLC is composed of a knowledge base, that includes the information given by the expert in the form of linguistic control rules, a fuzzification interface, which has the effect of transforming crisp data into fuzzy sets, an inference system, that uses them together with the knowledge base to make inference by means of a reasoning method and a defuzzification interface, that translates the fuzzy control action thus obtained to a real control action using a defuzzification method [1].
In this paper we will optimize a fuzzy controller Takagi-Sugeno zero-order. The controller inputs are the error$e(t)$ and the derivative of the error $\Delta e(t)$, and its output is the command $u(t)$.

2.1 Fuzzification

For flexibility in the implementation of the regulator, we must limit the universe of input and output at intervals determined by the normalization of input and output [-1,1], to do this, we use gains of adaptations to have the desired dynamic.

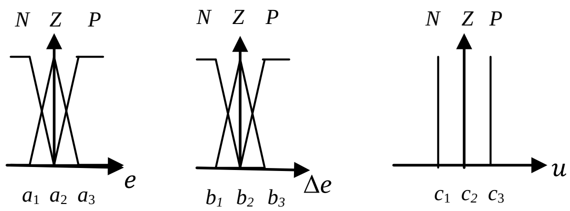


Fig. 1 Membership functions of inputs and output

The triangular membership functions are used for the Fuzzification of inputs (see Fig. 1). For example, the input $e(t)$ is fuzzified by following membership functions:

$$\mu_N(x) = \begin{cases} 1 & if\ x < a_1 \\ \dfrac{a_2 - x}{a_2 - a_1} & if\ a_1 \le x < a_2 \\ 0 & if\ x \ge a_2 \end{cases}$$

$$\mu_Z(x) = \begin{cases} \dfrac{x - a_1}{a_2 - a_1} & a_1 \le x < a_2 \\ \dfrac{a_3 - x}{a_3 - a_2} & a_2 \le x < a_3 \\ 0 & otherwise \end{cases} \quad (1)$$

$$\mu_P(x) = \begin{cases} 0 & x < a_2 \\ \dfrac{x - a_2}{a_3 - a_2} & a_2 \le x < a_3 \\ 1 & x \ge a_3 \end{cases}$$

2.2 Fuzzy Inferences

The fuzzy rule base consists of three fuzzy rules according the Table 1. For the mechanism of inference, we used the method "min - max".

Table 1. Fuzzy Rule Base

de \ e	N	Z	P
N	N	-	-
Z	-	Z	-
P	-	-	P

2.3 Defuzzification

we use the center of gravity for the defuzzification.

3. Particle Swarm Optimization (PSO)

PSO is a population-based stochastic optimization technique developed by Eberhart and Kennedy [2] that was inspired by social behavior of bird flocking or fish schooling. In PSO, each single solution is a “bird” in the search space. We call it “particle”. All the particles have fitness values that are evaluated by the fitness function to be optimized, and have velocities that direct the flying of the particles. The particles fly through the problem space by following the current optimum particles [3].

In every iteration, each particle is updated by following two “best” values. The first one is the best solution and has achieved so far. This value is called *pbest*. Another “best” value that is tracked by the particle swarm optimizer is the best value, obtained so far by all particles in the population. This best value is a global best and called *gbest* [3].
During the iteration time *t*, the update of the velocity from the previous velocity to the new velocity is determined by Eq. (2). The new position is then determined by the sum of the previous position and the new velocity by Eq. (3):

$$V(t+1) = wV(t) + c_1 * R1 * \big(pbest(t) - p(t)\big) + c_2 * R2 * \big(gbest(t) - p(t)\big) \quad (2)$$

$$p(t+1) = p(t) + V(t+1) \quad (3)$$

V is the particle velocity, The variable w is called as the inertia factor, p is the current solution, and $pbest$ and $gbest$ are defined as stated before. $R1$ and $R2$ are the random numbers uniformly distributed within the interval [0,1]. The variables c_1, c_2 are learning factors.
In order to guide the particles effectively in the search space, the maximum moving distance during one iteration is clamped in between the maximum velocity $[-Vmax, Vmax]$ given in Eq. (4) and similarly for its moving range given in Eq. (5) [4]:

$$V(t) = sign(V(t))\min(|V(t)|, Vmax) \quad (4)$$
$$p(t) = sign\big(p(t)\big)\min(|p(t)|, pmax) \quad (5)$$

4. Tabu search (TS)

Tabu search is a higher level heuristic algorithm for solving combinatorial optimization problems. It is an iterative improvement procedure that starts from any initial solution and attempts to determine a better solution. TS was proposed in its present form a few years ago by Glover [5]. It has now become an established optimization approach that is rapidly spreading to many new fields. The flowchart of TS algorithm procedure is shown in Fig. 2 [6].

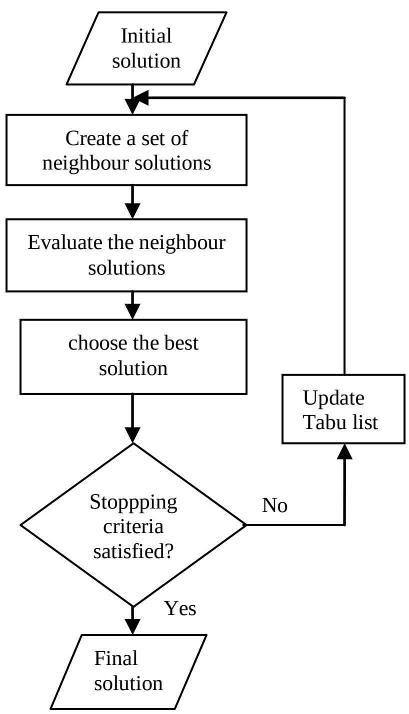


Fig. 2 Flowchart of a standard TS algorithm

5. Hybrid PSO-TS learning algorithm

To accelerate the convergence of PSO, it was proposed to combine the PSO algorithm with Tabu search to find a better solution in a minimum computation time and accuracy.

At each iteration of PSO, we calculate the best solution, then we will search its best neighbor by TS on minimizing a certain criterion (objective function) is the mean square error (MSE) calculated by the following equation:

$$MSE = 1/nT \sum_{k=1}^{n} e(k)^2 \quad (6)$$

Where: n is the total number of samples and T the sampling time, $e(k) = r(k) - y(k)$ is the difference between the value of the desired output $r(k)$ and the value of the measured output $y(k)$ process under control.

Before turning to the hybrid algorithm PSO-TS, we must first introduce the particle on which the parameters are encoded to optimize. These are the parameters of each membership function of inputs, and conclusions of fuzzy rules.

According to Fig. 3 the particle will consist of nine parameters which are the modal values of membership functions of input and output fuzzy singletons, respecting the following constraint:

$$\begin{cases} a_1 < a_2 < a_3 \\ b_1 < b_2 < b_3 \\ c_1 < c_2 < c_3 \end{cases}$$

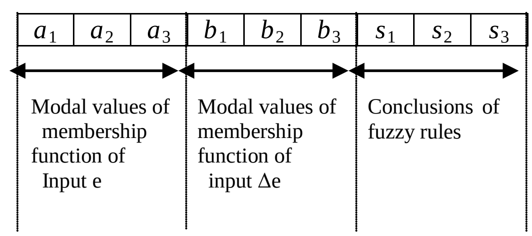


Fig. 3 *Particle* structure of PSO-TS

The pseudo-code of the procedure PSO-TS is as follows:

Begin PSO
  Initialize $R1, R2, c_1, c_2, w$
  Initialize $Vmin, Vmax, pmin, pmax$
  Initialize *Particles* randomly

  Do
    For each particle
      Calculate fitness value by Eq. (6)
          If the fitness value is better than the best fitness value ($pbest$) in history
      set current value as the new $pbest$
    End
     Choose the particle with the best fitness value of all the particles as the $gbest$
    For each particle
      Calculate particle velocity, V, according Eq. (2)
      Normalize V by Eq. (4)

Update particle position,p, according Eq. (3)
Normalize p by Eq. (5)
End

Begin Tabu Search
Set the iteration counter $k = 0$
Initialize the tabu list $T = 0$
Initialize the aspiration function
Do
For each particle
Generate solutions randomly
Evaluate each neighbor by Eq. (6), and choose the best neighbor p
Update the Tabu list and aspiration function
End
While the list tabu is not full or iteration is not reached
End

While maximum iterations is not reached
End

6. Application

We will apply the hybrid PSO-TS algorithm to optimize a fuzzy controller Takagi-Sugeno zero-order. This controller will be used to control a simple inverted pendulum modeled by Eq (7).

$$\ddot{\theta} = \frac{-u\, cos\theta + b\; cos\theta \dot{x} - m\, l\, \dot{\theta}^2 sin\theta - (M+m)\, g\, sin\theta}{l\left(\frac{4}{3(M+m)} - m\, cos^2\theta\right)} \quad (7)$$

Where θ is the angular position and u the control signal.

$g = 9.8\, m/s^2$ is the gravity constant, m_c (mass of the cart) 1 kg, m_p (mass of the pendulum) 0.1 kg, l (length of the pendulum) 0.5 m.

The objective is to maintain the broom in the upright position by means of the force f for any initial position of the broom without regard to the position and velocity of the cart.

The block diagram (see Fig. 4) shows the strategy optimization of the controller:

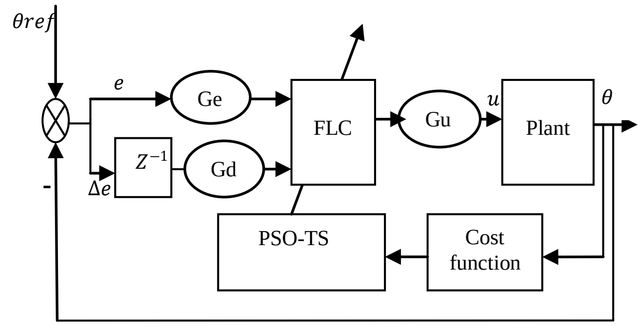


Fig. 4 Optimization of Fuzzy Controller

The cost function is calculated by Eq. (6) with a nominal model. The initial conditions are:$\theta(0) = 0.22\, rad, \dot{\theta}(0) = 0\, rad/sec$.

The PSO-TS algorithm was run for 2 generations. After each generation of PSO, TS was run for 5 iterations, The last generation contained a set of stable and performance satisfying solutions with controllers having three rules extracted for analysis.

Fig. 6 shows the form of membership functions of inputs and output after optimization and Fig. 7 shows the evolution of the cost function. The PSO-TS algorithm converges in a computation time equal to 2.88 sec with a response time of 0.2 sec < ts < 0.8 sec.
To demonstrate the robust ability of the selected fuzzy controller with proposed method method, different initial conditions (0.22 rad < θ < 0.8 rad) are shown in Fig. 5. The simulation result shows that the inverted pendulum is successfully controlled by the PSO-TS algorithm, and the angle of pole is quickly converged toward the balance.

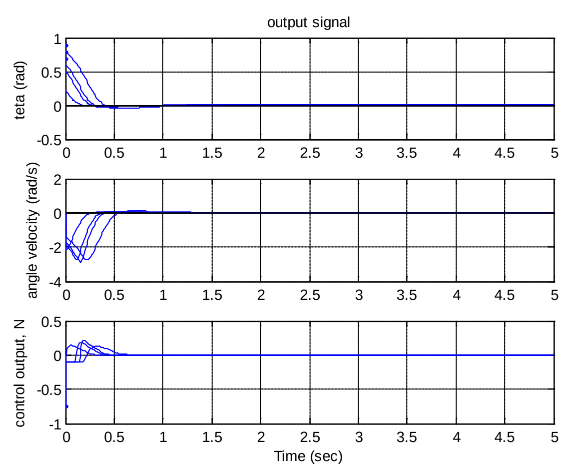


Fig. 5. the pendulum response to optimal fuzzy controller for differents initial conditions.

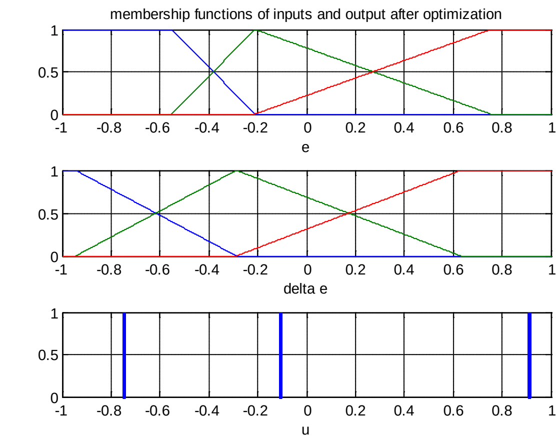


Fig. 6 Arrangement and forms of membership functions of the premises and conclusions after optimization

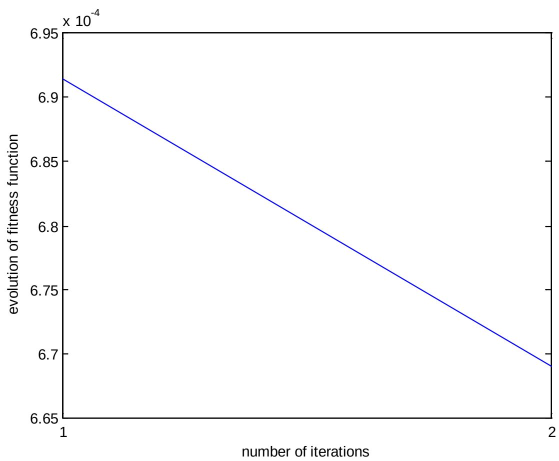


Fig. 7. Fitness function during iterations

7. Conclusion

In this paper, the PSO- TS hybrid algorithm is used to set the parameters of membership functions of inputs and conclusions of rules of fuzzy controller for Takagi-Sugeno zero-order, the latter is used to stabilize a inverted pendulum.

Simulation results have shown that the application of the proposed algorithm has improved the calculation time and the response time while optimizing the accuracy and simplifying the structure of this controller. We used only three fuzzy rules and a minimum number of generations for the implementation of this algorithm. The robustness test has proven the reliability of fuzzy controller optimized.

References

[1] T.J. Ross, "Fuzzy Logic With Engineering Applications", McGraw-Hill, New York, 1995.

[2] RC. Eberhart, J. Kennedy, " A new optimizer using particle swarm theory," Proceedings of the sixth international symposium on micro machine and human science. IEEE service center. Piscataway, NJ, Nagoya, Japan. 1995. p. 39–43.

[3] A. Allahverdi, "A PSO and a Tabu search heuristics for the assembly scheduling problem of the two-stage distributed database application", Computers and Operations Research, 2006.

[4] H. Liu, A. Abraham and W.Zhang, "A Fuzzy Adaptive Turbulent Particle Swarm Optimization", Int. J. Innovative Computing and Applications, Vol. 1, No. 1, 2007

[5] F. Glover, "Tabu search-part I", ORSA J. Comput. 3 (1), 1989, pp. 190–206.

[6] D.T. Pham, D. Karaboga, Intelligent Optimisation Techniques: Genetic Algorithms, Tabu Search, Simulated Annealing, and Neural Networks, Springer, Berlin, 2000

Nesrine TALBI was born in Jijel, ALGERIA in 1978, received her engineer and M.Sc degrees in Electronics, option: control systems from the Mentoury University, Constantine, Algeria. She is currently working toward the Ph.D. degree in control engineering at the University of Constantine. She is a Lecturer in Department of Electronic, Faculty of Sciences and Technology, Jijel University, ALGERIA.

Khaled BELARBI obtained his engineer degree polytechnic school, Algiers, Algeria, and MSc and PhD in control engineering both from Control System Center UMIST, Manchester, UK. He is currently a professor of electrical engineering with the Department of Electronics, Faculty of Engineering, University of Mentoury, Constantine, ALGERIA, and with the Laboratory of Automatic and Robotic at the same Department.His current interests are in predictive control, fuzzy control, intelligence control and neural control.

Study of the effect DCT and DWT domains on the imperceptibility and robustness of Genetic watermarking

Abduljabbar Shaamala[1], Shahidan M. Abdullah[2] and Azizah A. Manaf[3]

[1] **faculty of Computer Science and Information Technology, Universiti Teknologi Malaysia (UTM), Johor, Malaysia**

[2] **Advanced Informatics School (AIS), Universiti Teknologi Malaysia (UTM), Kuala Lumpur, Malaysia**

[3] **Advanced Informatics School (AIS), Universiti Teknologi Malaysia (UTM), Kuala Lumpur, Malaysia**

Abstract

Watermarking using genetic algorithm for the optimization of the tread-off between the watermarking requirements has attacked the attention of researchers; amongst the watermarking requirements, the imperceptibility and robustness is the main requirements. Watermarking embedded in frequency domain using DWT or DCT can affect the imperceptibility and robustness of watermarking, this paper studies the effect of embedding domain on the imperceptibility and robustness in genetic watermarking. Results of watermark image quality and attacks based on peak signal-to-noise ratio (PSNR) numerical correlation (NC) is analyzed through the paper sections, the DWT results showed more robustness high imperceptibility than DCT in watermarking based on GA.

Keywords: *watermarking, genetic algorithm. DWT, DCT.*

1. Introduction

Since digital multimedia have become progressively advanced in the rapidly growing field of internet application, data securities, including copyright protection and data integrity detection, have become a vast concern. One key for achieving information security is digital watermarking, which embeds hidden information or secret data in the image [1]. This technology works as a suitable tool for identifying the source, creator, owner, distributor, or authorized consumer of a document or image. Also watermarking can be used to detect a document or image is illegally distributed or modified [2].

Watermark techniques can be divided into two groups: Visible and invisible, the visible watermark is used if embedded watermark is intended to be seen by human eyes, For example, a logo inserted into corner of an image. While the invisible watermark is embedded into a host image by sophisticated algorithms and is invisible to the human eyes [3]

Watermarking techniques also can be classified according to its robust as robust, semi-fragile and fragile [3], Robust watermarks are designed to survive intentional (malicious) and unintentional (non-malicious) modifications of the watermarked image [4-6], Semi-fragile watermarks are layout for detecting any unauthorized alteration, and allowing in the same time some image processing operations [7, 8]. On the contrary, a watermarking technique that cannot robust against noise or attacks is called fragile technique[3]. Fragile watermarking techniques are concerned with complete integrity verification. Furthermore, watermarking techniques can be classified as blind and non-blind, Blind watermarking [9] techniques don’t require access to the original un-watermarked data (image, video, audio, etc.) to recover the watermark. In contrast, non-blind watermarking technique requires the original data [3, 9] needed for extraction of the watermarked. In general, the non-blind scheme is more robust than the blind watermark as it is obvious that the watermark can be extracted easily by knowing the un-watermarked data.

According to the embedding, watermarking techniques divided into two embedding domain, spatial domain and frequency domain [3, 9].The main concept of spatial domain [10] is to insert a watermark into an image by modifying the gray value of certain pixels in the image [11, 12]. The classical methods are to modify the last significant bits (LSB) of specific pixels of the host image based on the watermark bits [3].For frequency domain, the main concept to insert a watermark into frequency coefficients of the transformed image using the discrete cosine transform (DCT), the discrete wavelet transform (DWT) [13], or other kind of transforms techniques [3, 9]. There are requirements and constraints in design effective watermarking algorithms the three fundamental amongst it are:

- Imperceptibility: should the difference between the watermarked image and the original image not noticeable and visible by human eyes,
- Robustness: is the ability of watermarking to survive and withstand any intentional or unintentional attacks,
- Capacity: is the number of bits embedded into the original image.

The above watermarking requirements are conflicting with each other. If watermark is embedding bits into higher frequency coefficient would change the image as little as possible and achieve the imperceptibility. However, that would reduce the robustness since the watermarked image may experience filtering and the hidden watermark may be vanished. Also if watermark is Embedding bits into lower frequency coefficient would increase the robustness. However, this would sacrifice the imperceptibility [14],[15]. The watermarking problem can be viewed as an optimization problem. Therefore, genetic algorithm (GA) can be used for solving such problem [16], [17].

In this paper we present the effectiveness of embedding domain in the robustness of genetic watermarking. Section 2 briefly describes DWT and DCT embedding domain. Then an overview about genetic algorithm (GA) and some related watermark using genetic algorithm are briefly reviewed. In section 3 we disuse some result of previous works to obtain imperceptibility results and compare attacks results of it in order to identify the robust embedding domain in watermarking using GA.

2. Watermarking embedding domain:

2.1 Discrete Cosine Transform (DCT):

Discrete cosine transform (DCT) is a general orthogonal transform for digital image processing and signal processing, with such advantages, as high compression ratio, small bit error rate, good information integration ability and good synthetic effect of calculation complexity. DCT is a widely used mechanism for image transformation and has been adopted by JPEG to compress images; discrete cosine transform (DCT) is a Fourier-related transform similar to the discrete Fourier transform (DFT)[18]. Discrete cosine transform (DCT) turn over the image edge to make the image transformed into the form of even function[19]. It's one of the most common linear transformations in digital signal process technology. The DCT allows an image to be broken up into different frequency bands, making it much easier to embed watermarking information into the middle frequency bands of an image. The middle frequency bands are chosen such that the most visual important parts of the image (low frequencies) is to be avoided without over-exposing it to removal through compression and noise attacks (high frequencies)[20].

In DCT domain, DC component is more suitable to embed watermark than AC component (AC). Firstly, DC component has larger perceptual capacity. so, after embedding watermark it doesn't cause obvious change for visual quality of original image; secondly, signal processing and noise interference have smaller influence for DC component than AC component[21].

The DCT coefficients for output image $T(u,v)$ are computed according to the input $f(x,y)$ as equation.1. Where f is the input image with size $M \times N$ pixels, M is the raw and N is the column of the image, whereas $T(u,v)$ is the DCT matrix.

$$T(u,v) = \alpha_u \alpha_v \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y).\cos\frac{(2x+1)u\pi}{2M}.\cos\frac{(2y+1)v\pi}{2N} \quad (1)$$

where

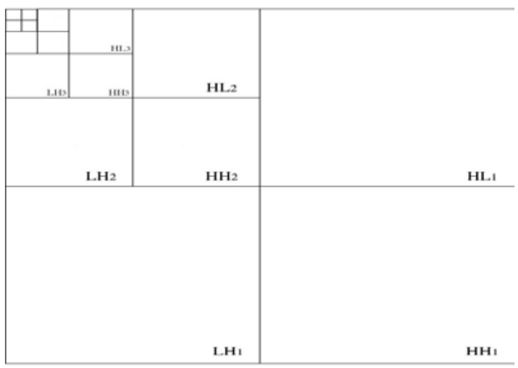
$$\alpha_u = \alpha_v = \begin{cases} \sqrt{\frac{1}{M}} & u = v = 0 \\ \sqrt{\frac{2}{N}} & u \neq v \neq 0 \end{cases}$$

The image recreated by applying inverse DCT according to equation 2.

$$\alpha_u \alpha_v \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} T(u,v).\cos\frac{(2x+1)u\pi}{2M}.\cos\frac{(2y+1)}{2N} \quad (2)$$

2.2 Discrete Wavelet Transform (DWT):

The wavelet transformation is a mathematical tool that can examine an image in time and frequency domains, simultaneously [22]. Discrete wavelet transform (DWT) is simple and fast transformation approach that translates an image from spatial domain to frequency domain. The DWT provides a number of powerful image processing algorithms including noise reduction, edge detection, and compression [23]. The transformed image is obtained by repeatedly filtering for the image on a row-by-row and column-by-column basis. An example of decomposing an image by a 2-level wavelet transformation is shown in Fig. 1. Then after applying the 2-level analysis filter bank a



four sub-band images will be obtained (LL, LH, HL, and HH),

Fig. 1 DWT decompose an image by 2-level

2.3 Advantages of DWT over DCT

According to[24] and [25], there is the DWT advantage over DCT as:

1. No need to divide the input coding into non-overlapping 2-D blocks, it has higher compression ratios avoid blocking artifacts.
2. Allows good localization both in time and spatial frequency domain.
3. Transformation of the whole image introduces inherent scaling
4. Better identification of which data is relevant to human perception higher compression ratio

3. Genetic Algorithms

Genetic Algorithms (GAs) introduced by Holland [26].GA is most widely used amongst the artificial optimization intelligent techniques. A GA is a stochastic searching algorithm based on the mechanisms of natural selection and genetics. GAs has been proven to be very efficient and stable in searching for global optimum solutions

In general, GAs start with some randomly selected population, called the first generation. Each individual in the population called chromosome and corresponds to a solution in the problem domain. An objective called fitness function is used to evaluate the quality of each chromosome. The next generation will be generated from some chromosomes whose fitness values are high. Reproduction, crossover and mutation are the three basic operators used to repeat many time until a predefined condition is satisfied or the desired number of iteration is reached. According to the applications for optimization, designers need to carefully define the necessary elements for dealing with the GA. Then, the fitness function in addition to the terminating criteria is evaluated with the natural selection, crossover, and mutation operations [27].

3.1 Watermarking based on GA related works

Researchers used GA to optimize the watermarking requirements, Wang et al [28] presented watermarking based on Genetic algorithm. They used bit substitution method. Huang et al [29] proposed watermarking method based on GA and DCT domain. They embedded watermark with visually recognizable patterns into image by selection modifying the middle frequency parts of the image. The GA is applied to search for the locations to embed into DCT coefficient block. In addition, Hsiang et al [16] proposed a robust watermarking based on DCT and GA. They tried to design a particle fitness function to solve the tread-off between the three watermarking matrices. On the other hand, they have considered the capacity to be constant. Moreover, Hsiang et al [30] have proposed watermarking based wavelet packet transform (WPT). They have assumed watermarked consists of 0's and 1's all bits of the watermark are embedded into host image. Also, Promcharoen and Rangsanseri [31] presented new approach for watermarking based on DCT. The authors used fuzzy C-mean (FCM) to classify the 8*8 block to texture or non-texture region. They used GA to find out the optimized parameter. As well as, Patra et al [32] proposed the digital watermarking scheme based on singular value decomposition (SVD). The authors used GA to optimize the conflict between quality and robustness. They used Sun et al algorithm for quantization embedding. Furthermore, Li et al [33] proposed watermarking based on DWT domain. They used Arnold transform and GA to improve the performance of watermarking algorithm.

4. Result Analyses

This section studies the effect of watermarking using GA on the embedding domains. Many of researchers have used Lena picture as the host image. They applied some types of attacks on that image after watermark embedding to prove the quality of their works. We chose some pervious works [34],[35],[36],[37],[38],[15],[31]and [39] to obtain robustness results. As well as, we analysis some pervious works [40],[41],[42],[31],[43],[38],[35],[44],[37] and [36] in order to get imperceptibility results. Therefore, we were analyzed their works and study the imperceptibility of image after embedded and how the attacks were affected by embedded domains.

The results were obtained through PSNR and NC measurements. PSNR (Peak signal-to-noise ratio) is measure the quality of the watermarked image by calculating the distortion between the watermarked and original image. While, NC (normalized correlation) It calculates the difference between the embedded and extracted watermark.

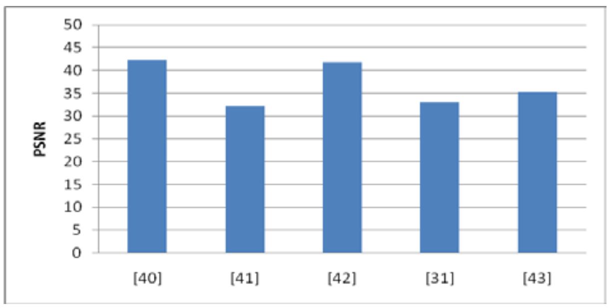


Fig. 2 PNSR value of image in DCT domain

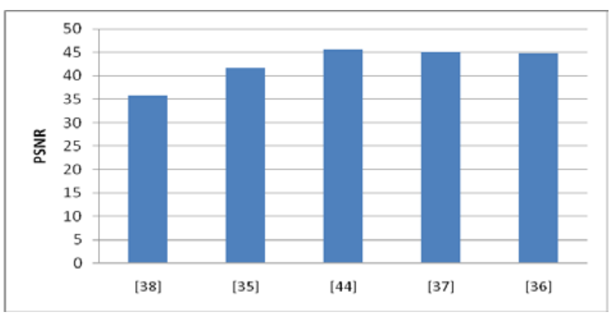


Fig. 3 PNSR value of image in DWT domain

It is concluded from comparing the obtained results of pervious work that, by calculating the difference between the original image and watermark image using PSNR measurement. Above figures show imperceptibility results for both embedded domains. The figure 3 shows the PSNR value according DWT embedding domain, while figure 3 shows according DCT embedding domain. The most of the results of DWT were found to be higher than 40, whereas most DCT result less than that. Therefore, the DWT embedding domain is better for imperceptibility than DCT embedding domain.

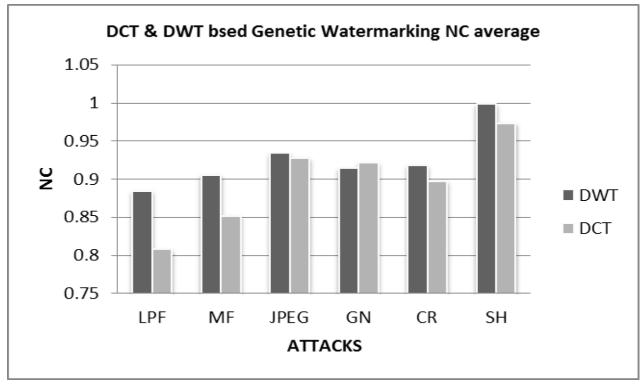


Fig.2 NC result after attack

The figure shows the attacks effect of on DWT and DCT domains. It shows in image processing operation like low-pass filtering (LPF) and medium filtering (MF) that the DWT domain is better than The DCT domain. Other attacks like JPGE, Cropping (CR) and sharpening (SH) almost have the same results with some advantages of DWT. Gaussian noise (GN) give DCT better result more than DWT. Therefore, the DWT embedding domain is more robust than DCT embedding domain

In the brief, DWT domain is better than the DCT domain for embedding in watermarking based on Genetic algorithm.

5. Conclusions

In this paper, we have proposed watermarking based on genetic algorithm and studied the effect of DWT and DCT embedding domain on imperceptibility and robustness of watermarking. As the result of the analysis obtained results by using PSNR and NC measurement. It is clarify the DWT is better than DCT for both imperceptibility and robustness of watermarking using genetic algorithm. In future work will study the affect of others optimization techniques and watermarking requirements with some testing experiments.

Acknowledgments

This work is part of a project supported by University

Technology Malaysia research grant (RUG Tier2 PY/2011/00997).

References

[1] Y.-T. Wu, and F. Y. Shih, “Digital watermarking based on chaotic map and reference register,” *Pattern Recognition,* vol. 40, no. 12, pp. 3753-3763, 2007.

[2] F. Shih, and Y.-T. Wu, “Information Hiding by Digital Watermarking,” *Information Hiding and Applications*, vol. 227, pp. 205-223, 2009.

[3] F.-H. Wang, J.-S. Pan, and L. Jain, "Intelligent Techniques," *Innovations in Digital Watermarking Techniques,* Studies in Computational Intelligence, pp. 27-44: Springer Berlin / Heidelberg, 2009.

[4] I. J. Cox, J. Kilian, F. T. Leighton *et al.*, “Secure spread spectrum watermarking for multimedia,” *Image Processing, IEEE Transactions on*, vol. 6, no. 12, pp. 1673-1687, 1997.

[5] P. L. Lin, “Oblivious digital watermarking scheme with blob-oriented and modular-arithmetic-based spatial-domain mechanism,” *Journal of Visual Communication and Image Representation,* vol. 12, no. 2, pp. 136-151, 2001.

[6] F. Deguillaume, S. Voloshynovskiy, and T. Pun, “Secure hybrid robust watermarking resistant against tampering and copy attack,” *Signal Processing,* vol. 83, no. 10, pp. 2133-2170, 2003.

[7] S. Agreste, G. Andaloro, D. Prestipino *et al.*, “An image adaptive, wavelet-based watermarking of digital images,” *Journal of Computational and Applied Mathematics,* vol. 210, no. 1-2, pp. 13-21, 2007.

[8] S.-k. Je, C.-j. Seo, J.-y. Lee *et al.*, "Self-organizing Coefficient for Semi-blind Watermarking," *Web Technologies and Applications*, Lecture Notes in Computer Science X. Zhou, M. Orlowska and Y. Zhang, eds., pp. 596-596: Springer Berlin / Heidelberg, 2003.

[9] F. Shih, “Digital Watermarking and Steganography FANDAMENTALS AND TECHHNIQUES,” *CRC Press*, 2008.

[10] W. N. Cheung, "Digital image watermarking in spatial and transform domains." pp. 374-378 vol.3.

[11] M. Ramkumar, “Data Hiding in Multimedia - Theory and Applications,” Ph.D.Thesis, New Jersey Institute of Technology, Newark, New Jersey, 2000.

[12] K. A. Navas, M. Sasikumar, and S. Sreevidya, "A Benchmark for Medical Image Watermarking." pp. 237-240.

[13] M. a. K. Vetterli, *Wavelets and Subband Coding*: Prentice Hall, 1995.

[14] A. Phadikar, and S. P. Maity, "Multibit QIM watermarking using M-ary modulation and lifting." pp. 1-5.

[15] H.-C. Huang, C.-M. Chu, and J.-S. Pan, “Genetic Watermarking for Copyright Protection,” *Information Hiding and Applications,* vol. 227, pp. 1-19, 2009.

[16] C.-M. C. a. J.-S. P. Hsiang-Cheh Huang, “The optimized copyright protection system with genetic watermarking,” *A FUSION OF FOUNDATIONS, METHODOLOGIES AND APPLICATIONS,* vol. 13, 2008.

[17] W. Shih-Hao, and L. Yuan-Pei, “Wavelet tree quantization for copyright protection watermarking,” *Image Processing, IEEE Transactions on,* vol. 13, no. 2, pp. 154-165, 2004.

[18] C.-M. P. a. I.-T. Lam, “Fingerprint Watermark Embedding by Discrete Cosine Transform for Copyright Ownership Authentication,” *INTERNATIONAL JOURNAL OF COMMUNICATIONS,* vol. 3, no. 1, pp. 8, 2009.

[19] M. Jiansheng, L. Sukang, and T. Xiaomei, “A Digital Watermarking Algorithm Based On DCT and DWT,” *International Symposium on Web Information Systems and Applications (WISA’09)*, 2009.

[20] I. El-Fegh, D. Mustafa, Z. S. Zubi *et al.*, “Color image watermarking based on the DCT-domain of three RGB color channels,” in Proceedings of the 10th WSEAS international conference on evolutionary computing, Prague, Czech Republic, 2009, pp. 36-39.

[21] M. Eyadat, and S. Vasikarla, “Performance evaluation of an incorporated DCT block-based watermarking algorithm with human visual system model,” *Pattern Recognition Letters,* vol. 26, no. 10, pp. 1405-1411, 2005.

[22] M. Ouhsain, and A. B. Hamza, “Image watermarking scheme using nonnegative matrix factorization and wavelet transform,” *Expert Syst. Appl.,* vol. 36, no. 2, pp. 2123-2129, 2009.

[23] C.-Y. Chang, H.-J. Wang, and S.-W. Pan, “A robust DWT-based copyright verification scheme with Fuzzy ART,” *Journal of Systems and Software,* vol. 82, no. 11, pp. 1906-1915, 2009.

[24] m. S. J. a. N. B. Hingoliwala H.A., “An image comperession by using haar wavelet transform,” *Advances in Computer Vision and Information Technology*, 2008.

[25] D. P. R. K. V.Srinivasa rao, G.V.H.Prasad, M.Prema Kumar, S.Ravichand, “Discrete Cosine Transform Vs Discrete Wavelet Transform: An Objective Comparison of Image Compression Techniques for JPEG Encoder,” *International Journal of Advanced Engineering & Applications,*, Jan 2010.

[26] J. H. Holland, “Adaptation in Natural and Artifcial Systems,,” *The University of Michigan Press, Ann Arbor, MI*, 1975.

[27] M. Gen, and R. Cheng, *Job-Shop Scheduling Problems*: John Wiley & Sons, Inc., 2007.

[28] R.-Z. Wang, C.-F. Lin, and J.-C. Lin, “Image hiding by optimal LSB substitution and genetic algorithm,” *Pattern Recognition,* vol. 34, no. 3, pp. 671-683, 2001.

[29] C.-H. Huang, and J.-L. Wu, "Watermark optimization technique based on genetic algorithms." pp. 516-523.

[30] H. Hsiang-Cheh, C. Yueh-Hong, and L. Guan-Yu, "DCT-Based Robust Watermarking with Swarm Intelligence Concepts." pp. 464-467.

[31] S. Promcharoen, and Y. Rangsanseri, "Genetic watermarking based on texture analysis in DCT domain." pp. 25-28.

[32] J. C. Patra, J. E. Phua, and C. Bornand, “A novel DCT domain CRT-based watermarking scheme for image authentication surviving JPEG compression,” *Digital Signal Processing,* vol. 20, no. 6, pp. 1597-1611, 2010.

[33] H.-f. Li, N. Chang, and X.-m. Chen, "A study on image digital watermarking based on wavelet transform," *The Journal of China Universities of Posts and Telecommunications*, vol. 17, no. Supplement 1, pp. 122-126, 2010.
[34] T. Kumaran, and P. Thangavel, "Watermarking in Contourlet Transform Domain Using Genetic Algorithm." pp. 257-262.
[35] Z. Ning, H. Zunwen, K. Jingming *et al.*, "An Optimal Wavelet-Based Image Watermarking via Genetic Algorithm." pp. 103-107.
[36] C. Yueh-Hong, and H. Hsiang-Cheh, "Genetic Watermarking Based on Wavelet Packet Transform." pp. 262-265.
[37] S.-C. Chu, H.-C. Huang, Y. Shi *et al.*, "Genetic Watermarking for Zerotree-Based Applications," *Circuits, Systems, and Signal Processing*, vol. 27, no. 2, pp. 171-182, 2008.
[38] Y. Lu, J. Han, J. Kong *et al.*, "A Novel Color Image Watermarking Method Based on Genetic Algorithm and Hybrid Neural Networks," *Rough Sets and Current Trends in Computing*, Lecture Notes in Computer Science S. Greco, Y. Hata, S. Hirano *et al.*, eds., pp. 806-814: Springer Berlin / Heidelberg, 2006.
[39] M. Rafigh, and M. E. Moghaddam, "A Robust Evolutionary Based Digital Image Watermarking Technique in DCT Domain." pp. 105-109.
[40] L. Jiann-Shu, C. Fong-Ping, and T. Kuen-Horng, "Image Watermarking Based on Multiple Objective Genetic Algorithm." pp. 102-105.
[41] C.-S. Shieh, H.-C. Huang, F.-H. Wang *et al.*, "Genetic watermarking based on transform-domain techniques," *Pattern Recognition*, vol. 37, no. 3, pp. 555-565, 2004.
[42] C. Chi-Ming, H. Hsiang-Cheh, and P. Jeng-Shyang, "An Adaptive Implementation for DCT-Based Robust Watermarking with Genetic Algorithm." pp. 19-19.
[43] S. Promcharoen, and Y. Rangsanseri, "Genetic Watermarking with Block-Based DCT Clustering." pp. 346-351.
[44] T. Kumaran, and P. Thangavel, "Genetic algorithm based watermarking in double-density dual-tree DWT." pp. 585-590.

A Robust Watermarking Algorithm Based on Image Normalization and DC Coefficients

T. Sridevi[1], V.Vijaya Kumar[2]

[1] Associate Professor, Department of CSE, Chaitanya Bharathi Institute of Technology, Gandipet, Hyderabad, Andhra Pradesh, India

[2] Dean, Department of CSE , Godavari Institute of Engg and Technology, Rajamundry Andhra Pradesh, India

Abstract

Digital watermarking algorithms are prune to different attacks. Of these geometric attacks are recognized as one of the strongest attacks for digital watermarking technology. Normalization watermarking algorithm is resistant to geometrical distortions and DC coefficients are perceptually most significant and more robust to many unintentional attacks (signal processing) and intentional attacks (unauthorized removal). In this paper we propose a robust non-blind watermarking algorithm based on DC coefficients and normalization. By applying discrete wavelet transformation technique (DWT) followed by block based Discrete Cosine Transformation (DCT) technique, DC components are obtained. The cover image is normalized using normalization algorithm in different frequency bands using wavelet decomposition and then block based DCT is applied. In the DC coefficients of the sub-band of the image, watermark is embedded. The algorithm resistant to geometrical distortions (rotation, scaling) and different other attacks (histogram equalization, noise) on different contrast images. The quantitative measurements like SNR (Signal to Noise Ratio), MSE (Mean Squared Error), RMSE (Root Mean Squared Error) and PSNR (Peak Signal to Noise Ratio) are tabulated.

Keywords: *Digital Watermarking, Normalization, Discrete Wavelet Transform (DWT), Discrete Cosine Transform (DCT).*

1. Introduction

Generally information could be hidden either by directly modifying the intensity value of pixels or frequency coefficients of an image. The former technique is called spatial domain technique and later is called frequency domain technique. In transform domain casting of watermark can be done in full frequency band of an image or in specific frequency band such as in low frequency band or in high frequency band or in middle frequency band. If perceptually insignificant coefficients are selected for embedding then the watermark may be lost by common signal processing operations. Geometric attacks are thought of as one of the most dangerous attacks in the digital watermarking world. Although several watermarking schemes that handle geometric attacks have been introduced each of them has problems. A little distortion such as rotation, scaling, translation, shearing, change in aspect ratio [1] can defeat most of the existing watermarking algorithms. Geometrical robustness problems persist due to the ease of performing rotation, scaling, and translation (RST) attacks on images without much quality compromise. A RST invariant domain watermark offers robustness against rotation, scaling, and translation attacks using invariant domains. This paper describes a digital image watermarking method that resists rotation, scaling, aspect ratio, histogram equalization, jpeg, noise attacks. The method exploits normalization. J' Ruanaidh and T. Pun [2] are the first to suggest a Fourier-Mellin transform based watermarking scheme to handle geometric attacks, such as rotation, scaling and translation (RST).The algorithm seems workable theoretically, but proved to be difficult in implementation. C. Y. Lin and M. Wu, etc [3] proposed an improvement to the implementation difficulty [2] by embedding the watermark into a 1-dimensional signal obtained by projecting the Fourier-Mellin transformed image onto log-radius axis. Such algorithm can embed only one bit information, i.e. presence or absence of the watermark, and the implementation is still a headache and far from practical application. Pereira and Pun [4] proposed an approach to embed a template into the DFT domain besides the intended watermark. Moment based image normalization has been used in computer vision for pattern recognition for a long time [5].Parameters of affine geometric attacks are estimated through the detection of template. The scheme handles flipping, scaling and rotation attacks, and it is only used to embed 1-bit information. Thus determining the place of watermark is a conflict between robustness and fidelity and it is purely application dependent.

In [6] the image is divided into many block of size 8x8 and it is block transformed using DCT technique. These transformed blocks are randomly shuffled to decor relate and to spread the watermark across the entire image. The mid band blocks are selected from the permuted blocks to embed watermark. In [7] the cover is decomposed into four bands. The high frequency band is inverse transformed to obtain high frequency image and it is SVD decomposed to embed watermark by modifying high frequency components. Results show that the system is withstanding certain attacks including geometric attacks. In [8] Image is transformed by DWT technique to K level. The middle frequency band LH and HL are SVD transformed and watermark is hidden. Similarly in low frequency and high frequency band the watermark is embedded using distributed discrete wavelet transform method (DDWT). In [9], three level decomposition of DWT is applied on image to get ten bands of frequencies. All ten bands of frequency coefficients are SVD transformed to embed watermark. A new watermarking scheme for images based on Human Visual System (HVS) and Singular Value Decomposition (SVD) in the wavelet domain is discussed. Experimental results show its better performance for compression, cropping and scaling attack. In [10] two level decomposition of DWT is applied to transform an image into bands of different frequency and a particular band is selected and converted into blocks of size 4x4 for embedding data. Each of those the blocks are SVD transformed and watermark is hidden into diagonal matrix of every block. The similarity between the original watermark and the extracted watermark from the attacked image is measured using the correlation factor NC. The algorithm shows that when DWT is combined with SVD technique the watermarking algorithm outperforms than the conventional DWT algorithm with respect to robustness against Gaussian noise, compression and cropping attacks. Based on the review performed many works are existing for embedding watermark by combining DWT and SVD technique for intensity images. In the proposed work the image the image is first normalized using BNA algorithm and the watermark is embedded in the DC components of transformed image. In order to increase the robustness, the low frequency band can be selected. But to increase the capacity of the watermark full band can also be used. The selected band is divided into block of size 4x4 which in turn DCT transformed to obtain only DC coefficients. These DC coefficients are SVD transformed to embed watermark in singular values.

2. Watermarking Algorithm

In this paper, we implement a normalization algorithm (BNA) [11] based watermarking in DWT and DCT.

2.1 Normalization Algorithm

The NA [11] consists of two components. The first is a Rotate and Scale (RnS) step that rotates the signal by a fixed angle θ followed by scale normalization. The second component is the computation of the orientation indicator index (OII) defined below. The algorithm iterates repeatedly the RnS and OII steps. After a finite number of iterations, the signal corresponding to the maximum value of the OII is chosen as f^{θ}.

1. Rotate and Scale (RnS):
In this step, the signal is rotated by a angle θ and is scaled to the normalized the size by {1/α , 1/β }, where α and β are the dimensions of the region of support for the signal f(x, y), and are defined as

$$\alpha = \left|max_x\{x: f(x,y) \neq 0\} - min_x\{x: f(x,y) \neq 0\}\right| \quad (1)$$

and

$$\beta = \left|max_y\{y: f(x,y) \neq 0\} - min_y\{y: f(x,y) \neq 0\}\right| \quad (2)$$

If $f(x, y)$ is the input to the RnS step, the outcome $f(x_d, y_d)$ is obtained by a coordinate transform,

$$\begin{bmatrix} x_d \\ y_d \end{bmatrix} = \begin{bmatrix} 1/\alpha & 0 \\ 0 & 1/\beta \end{bmatrix} \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad (3)$$

2. Orientation Indicator Index (OII):
OII [1] of $f(x,y)$ is defined as

$$OII_f = \sqrt{{\mu_x}^2 + {\mu_y}^2} \quad (4)$$

Where

$$\mu_x = \int_R (x - g_x)^3 . f(x,y)\,dxdy$$

$$\mu_y = \int_R (y - g_y)^3 . f(x,y)\,dxdy$$

(g_x, g_y) is the center of the mass for $f(x,y)$,and defined as

$$g_x = \int_R x . f(x,y)\,dxdy \quad (5.a)$$

$$g_y = \int_R y . f(x,y)\,dxdy \quad (5.b)$$

3. BNA Algorithm

The BNA[1] is slightly different version from [12].Because of the periodic property of OII, a full rotation interval of θ = [0,2π] is time consuming and not necessary. Other orientations with max OII can be find by adding multiple π /2.

The NA algorithm used in this paper is summarized as following.

1) Recentering: Recenter $f^d(x, y)$ with respect to the center of mass (g_x, g_y) defined by eq .5.
2) Scale normalization: Compute {α, β} from equation eq. 1,2, and scale the pixel coordinates (x, y) to (x/α, y/β).
3) Calculate OII for each rotated position
4) Iteration: Repeat steps 3 and 4 for rotation interval θ = [0, π /2].
5) Normalization: Selects the signal corresponding to the maximum value of the OII, and stores it as $f^d(x,y)$

2.2 Discrete Cosine Transform

DCT is another important transformation technique which is widely used due to its energy compaction and decorrelation properties. DCT technique is faster than discrete Fourier transform since the bases are cosine function for the former technique and complex function for the later technique The transformed matrix consists of both AC and DC coefficients. If the DCT technique is applied on block of size NxN then it is called block DCT. In DCT transformed block the left top corner element is called as DC coefficient which is perceptually significant and the remaining coefficients are called AC coefficients which are perceptually insignificant. These coefficients are zigzag scanned to obtain frequency components of an image in decreasing order. These DC and AC components are modified to embed watermark in it [7][13]. Equ. 1 and 2 are used for taking transformation and inverse transformation of an image.

2.3 Discrete Wavelet Transform

DWT is a transformation technique is used to represent an image in a new time and frequency scale by decomposing the input image into low frequency, middle and high frequency bands. The value of low frequency band is the averaging value of the filter whereas the high frequency coefficients are wavelet coefficients or detail values. [4]. The DWT can be used to decompose image as a multistage transform. In the first stage, an image is decomposed into four subbands LL1, HL1, LH1, and HH1, where HL1, LH1, and HH1 represent the finest scale wavelet coefficients, while LL1 stands for the coarse level coefficients, i.e., the approximation image. Fig.1 shows the one level wavelet decomposition of an image [14].

LL	LH
HL	HH

Figure 1: One level decomposition

2.4 Watermarking Scheme

a. Find the normalized image using BNA algorithm as described in section 2.3.
b. Apply DWT to decompose of an image into various frequency bands.
c. Divide the middle frequency band and apply DCT. Divide DCT transformed band into smaller blocks of size 4X4.
d. Extract the DC coefficients $\delta_{i,j}$ from every DCT transformed blocks.
e. Modify the DC coefficients with the watermark coefficients W i,j .

$$\widetilde{\delta_{i,j}} = \delta_{i,j} + \alpha\ W_{i,j}$$

f. Replace the original DC's $\delta_{i,j}$ by the modified DC's $\widetilde{\delta_{i,j}}$ in each block. Then apply inverse DCT and inverse DWT.
g. Step b to step f is repeated for hiding watermark in other bands of a cover image.
h. Inverse wavelet transformation technique is applied to get the watermarked.
i. Restore the watermarked image to the original orientation and size.

Figure 2: Watermark Embedding

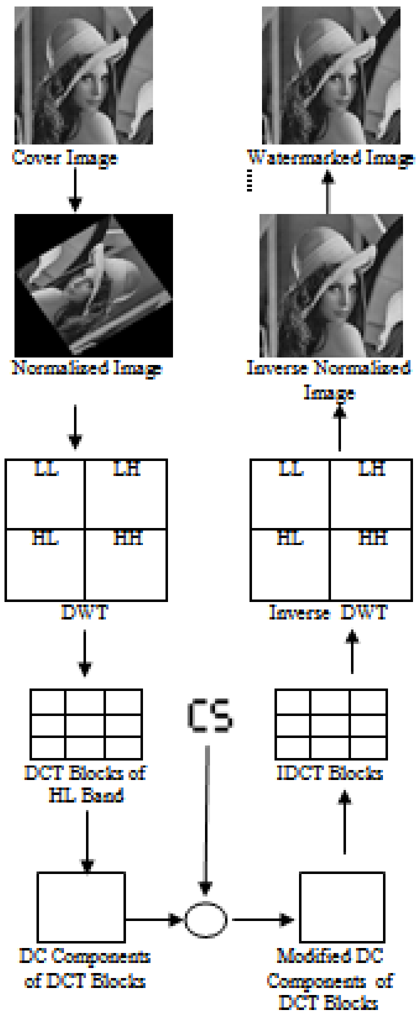


Figure 2: Watermark embedding

2.5 Decoding Procedure

Find the normalized image of watermarked image using BNA algorithm.

a. Apply DWT to decompose of watermarked image into various frequency bands.
b. Divide the middle frequency bands and apply DCT. Divide DCT transformed band into smaller blocks of size 4X4.
c. Extract the DC coefficients $\widetilde{\delta_{i,j}}$ from every DCT transformed blocks.
d. Modify the DC coefficients with the watermark coefficients W I,j .
e. $$W_{i,j} = (\widetilde{\delta_{i,j}} - \delta_{i,j})/\alpha$$
f. Reconstruct the watermark image. In the proposed method, to extract the watermark from all frequency bands, it uses original cover image. So, this algorithm is can be classified as non-blind watermarking technique.

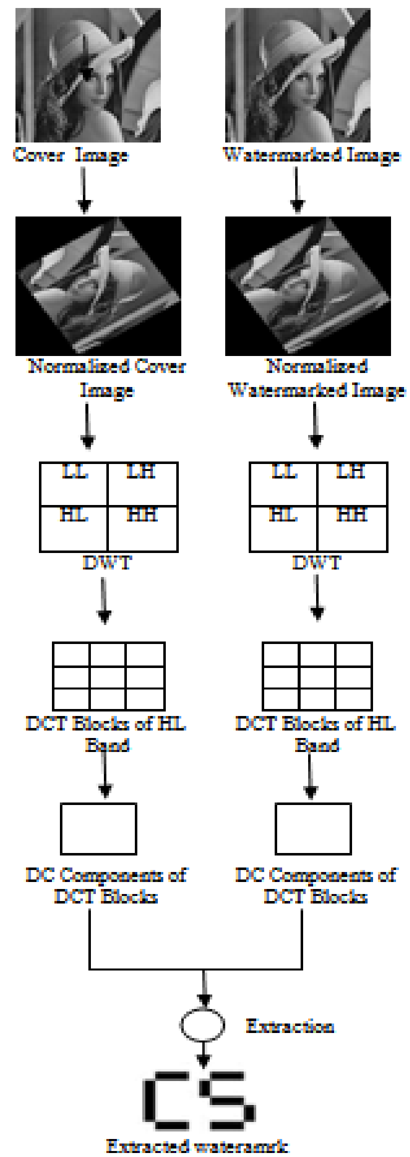


Figure3: Watermark Extraction

3. Results

In this proposed algorithm, Lena, peppers, man images of size 256x256 are taken as test images and the watermark is the CS logo of size 64x64. Embedding intensity value is varied from 0.1 to 1 for all frequency bands. The original images are shown in figure 4 and watermark is shown in figure 5. The watermarked images are shown in Fig 6. Similarly the extracted watermarks are shown in Fig.7.

Figure 4(a-c): Sample cover images

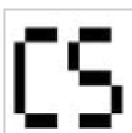


Figure 5: Watermark image

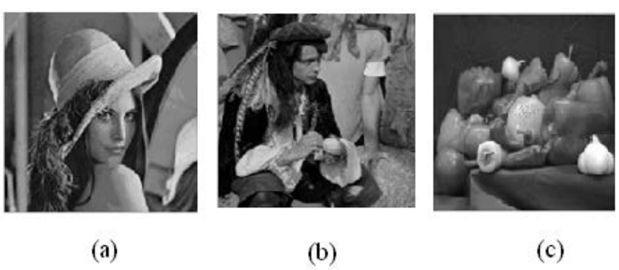


Figure 6(a-c): Watermarked images

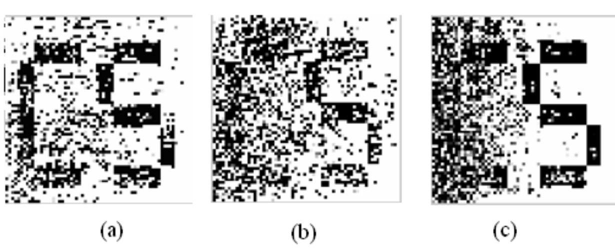


Figure 7(a-c): Extracted watermarks

In order to test the quality of the extracted watermark and cover data both subjective and objective measurements are used. The objective criteria are measured through (5), (6) and (7), (8).

$$MSE = \frac{1}{M\ X\ N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \left(f(i,j) - g(i,j)\right)^2 \quad (5)$$

$$RMSE = \frac{1}{M\ X\ N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \left(f(i,j) - g(i,j)\right)^2 \quad (6)$$

$$SNR = \frac{\sum_{i=0}^{M-1} \sum_{j=0}^{N-1} g(i,j)^2}{\sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \left(g(i,j) - f(i,j)\right)^2} \quad (7)$$

$$PSNR = 10 * \log_{10}\left(\frac{255^2}{MSE}\right) \quad (8)$$

The SNR, MSE, RMSE and PSNR values for watermark embedding are shown in Table 1.

Table 1: Watermark embedding (Man)

Frequency Band	k	Measures			
		SNR	MSE	RMSE	PSNR
LL	0.1	1.3316	2.754e+003	52.484	13.7642
	0.5	1.3161	2.761e+003	52.553	13.7528
	1	1.2948	2.802e+003	52.934	13.6901
LH	0.1	-0.0668	26.6574	5.1631	33.9066
	0.5	-0.0852	40.0348	6.3273	32.1404
	1	-0.1007	70.3027	8.3847	29.6951
HL	0.1	-0.0672	26.7334	5.1704	33.8943
	0.5	-0.0847	40.9852	6.4020	32.0385
	1	-0.0995	71.5244	8.4572	29.6203
HH	0.1	-0.0668	26.1553	5.1142	33.9892
	0.5	-0.0864	37.2286	6.1015	32.4560
	1	-0.1022	62.4746	7.9041	30.2078

The SNR, MSE, RMSE and PSNR values for watermark extraction are shown in Table 2.

Table 2: Watermark Extraction (Man)

Frequency Band	*k*	Measures			
		SNR	MSE	RMSE	PSNR
LL	0.1	-43.9509	3.362e+005	579.8931	-7.1022
	0.5	-29.9779	1.624e+004	127.4558	6.0576
	1	-24.0186	6.316e+003	79.4777	10.1599
LH	0.1	-4.9816	2.886e+003	53.7218	13.5618
	0.5	4.1096	2.894e+003	53.7957	13.5498
	1	7.3074	2.958e+003	54.3950	13.4536
HL	0.1	-3.9755	2.857e+003	53.4512	13.6057
	0.5	3.8456	2.901e+003	53.8669	13.5384
	1	7.3965	2.942e+003	54.2422	13.4780
HH	0.1	1.1388	2.984e+003	54.6329	13.4157
	0.5	12.1995	2.986e+003	54.6442	13.4139
	1	14.4298	3.002e+003	54.7945	13.3901

Attacks like Gaussian noise, rotation, Histogram equalization, Contrast increase, contrast decrease and scaling are applied on the watermarked image. The watermark is extracted from the attacked images. The SNR, MSE, RMSE and PSNR values for extraction after applying attacks for embedding strength of 0.5 are shown is Table 3.

Table 3: Watermark extraction after attacks

Band	Attack	Measures			
		SNR	MSE	RMSE	PSNR
LL	Noise	-24.106	6.383e+003	79.8937	10.1145
	Rotation	3.1679	2.584e+003	50.8418	14.0404
	Hist Eq	3.3647	2.608e+003	51.0758	14.0005
	C Inc	3.5755	2.657e+003	51.5432	13.9214
	C Dec	4.2709	2.7067e+003	52.0264	13.8403
	Scaling	2.7894	2.555e+003	50.5516	14.0901
LH	Noise	-3.9794	3.058e+003	55.3006	13.3102
	Rotation	6.5626	2.910e+003	53.9466	13.5255
	Hist Eq	4.1696	2.798e+003	52.9136	13.6935
	C Inc	5.2821	2.867e+003	53.5506	13.5895
	C Dec	6.4701	2.910e+003	53.9481	13.5253
	Scaling	4.6206	2.836e+003	53.2559	13.6374
HL	Noise	-4.6146	3.071e+003	55.4168	13.2920
	Rotation	6.5458	2.905e+003	53.9049	13.5322
	Hist Eq	4.1949	2.800e+003	52.9225	13.6920
	C Inc	6.0581	2.884e+003	53.7029	13.5648
	C Dec	6.5879	2.911e+003	53.9557	13.5241
	Scaling	5.2452	2.855e+003	53.4383	13.6077
HH	Noise	4.9786	3.041e+003	55.1518	13.3336
	Rotation	10.9554	2.977e+003	54.5662	13.4263
	Hist Eq	8.3710	2.951e+003	54.3281	13.4643
	C Inc	10.1274	2.973e+003	54.5250	13.4329
	C Dec	12.2796	2.990e+003	54.6832	13.4077
	Scaling	7.5581	2.932e+003	54.1518	13.4925

4. Conclusions

Algorithm based on Image normalization and DC coefficients is implemented with different embedding strengths. By looking at the results LH, HL bands has shown robustness to the attacks implemented. LL band is the most vulnerable band.

References

[1] Ping Dong, Nikolas P. Galatsanos, "Affine Transformation Resistant watermarking based on image normalization," at IPL conference at Illinois Institute of Technology, IL, May 2007.
[2] F.A.P. Petitcolas, R.J. Anderson, and M.G. Kuhn, "Attacks on copyright marking systems," in Workshop on Information Hiding, Portland, OR, 15-17 April, 1998 .
[3] J. O'Ruanaidh and T. Pun, "Rotation, scale and translation invariant spread spectrum digital image watermarking, " Signal Processing, vol. 66, no. 3, pp. 303-317, 1998.
[4] C. Y. Lin, M. Wu, etc, "Rotation, Scale, and Translation Resilient Public watermarking for images," in IEEE trans. on image processing, vol. 10, no.5, May 2001.
[5] Ingemar J. Cox, Matthew L. Miller, Jeffrey A. Bloom, Digital Watermarking, Morgan Kaufmann publishers, 2001.
[6] Liu Quan, AI Qingsong "A combination of DCT based and SVD based watermarking, ICSP proceedings of IEEE International conference on signal processing, pp. 873-876, 2004.
[7] Liu Liang and Sun Qi "A new SVD-DWT composite watermarking", ICSP proceedings of IEEE International conference on signal processing .2006
[8] Jung-Chun Liu, Chu-Hsing Lin, and Li-Ching Kuo" A Robust full band image watermarking scheme" Proceedings on IEEE 2006.
[9] 10Qiang Li, et al, "Adaptive DWT-SVD Domain Image Watermarking Using Human Visual Model" proceedings of 9th international conference on advanced communication Technology, Volume 3, pp:1947 - 1951, Feb.2007.
[10] Ruth Buse Dili, Elijah Mwangi, "An Image Watermarking method based on the singular value transformation and the wavelet transformation "Proceedings on IEEE 2007.
[11] Victor H. S. Ha and Jose M. F. Moura, "Affine invariant wavelet transform," Acoustics, Speech, and Signal processing, 2001, proceedings. May, 2001.
[12] Masoud Alghoniemy and Ahmed H. Tewfik, "Geometric distortion correction through image normalization," Multimedia and Expo, 2000, ICME 2000.
[13] Ben Wang, Jinkou Ding, et all. "An Image Watermarking Algorithm Based on DWT DCT and SVD "Proceedings of IC-NIDC2009.
[14] Victor H. S. Ha and Jose M. F. Moura, "Affine invariant wavelet transform," Acoustics, Speech, and Signal processing, 2001, proceedings. May, 2001.

T.Sridevi received B.E. from Osmania University and M.Tech.(CSE) degree from JNT University in 2002. She is pursuing her Ph.D. from Osmania University in Computer Science under the guidance of Dr V. Vijaya Kumar. She has 14 years of teaching/industry experience. She joined as Assistant Professor in Chaitanya Bharathi Institute of Technology, Hyderabad, India in 2002. Presently she is Associate Professor, CBIT, and Hyderabad. She has published 10 research publications in various international and national conferences and journals. Her research areas include Digital Water Marking, Image Processing and Soft Computing. She is a life member of IETE.

Vakulabharanam Vijaya Kumar received integrated M.S. Engg, degree from Tashkent Polytechnic Institute (USSR) in 1989. He received his Ph.D. degree in Computer Science from Jawaharlal Nehru Technological University (JNTU) in 1998.He has served the JNT University for 13 years as Assistant Professor and Associate Professor and taught courses for M.Tech students. He has been Dean for Dept of CSE and IT at Godavari Institute of Engineering and Technology since April, 2007.His research interests includes Image Processing, Pattern Recognition, Digital Water Marking and Image Retrieval Systems. He is a life member for CSI, ISTE, IE, IRS, ACS and CS. He has published more than 50 research publications in various National, Inter National conferences, proceedings and Journals.

Architectural Model of Localization in Multi-Scale Communication for Wireless Sensor Networks

S. Swapna Kumar[1], **Dr M. Nanda Kuma**[2], **Dr V.S Sheeba**[3]

[1] **Department of Electronics & Communication Engineering, Anna University Coimbatore, T.N, India.**

[2] **Department of Electrical Engineering, Calicut University, Kerala, India.**

[3] **Department of Electronics & Communication Engineering, Calicut University, Kerala, India.**

Abstract

There are several range-based localization algorithms that identify sensor's positions for different wireless sensor networks. In this paper we propose the approach that highlights a new method of localization scheme to achieve high localization accuracy in the presence of radio irregularity and obstacles effects for wireless sensor networks. We present the evaluation of complex localization scenario, by applying convex localization estimation algorithm with possible orientation of obstacle using Euclidean and Interior point algorithms for non-ideal transmission networks. It has been observed in the simulation results that the proposed scheme significantly improve the localization accuracy. The range-free localization algorithms evaluate the node ranging error accuracy in anisotropic networks to solve the localization inequalities problem when the feasible set is empty. Study demonstrates the simulation results for the effectiveness of this algorithm. Furthermore, this scheme evaluate infeasible points caused by a complex radio various rigid statistical analysis to validate the results.

Keywords: *Convex optimization Euclidean algorithms, Interior point algorithms, Localization, Mobile anchor, Wireless sensor network.*

1. Introduction

Localization is a process of finding location of the sensor nodes, which is very essential requirement for various reasons such as finding faulty node, re-configuration of the software or application requirements for target tracking etc. Numerous localization techniques have been proposed and implemented by various researchers [1], [2] in WSN. Localization techniques are basically divided into two categories (i) Range Based Localization and (ii) Range Free Localization [3]. The range based schemes estimate the distance between nodes for obtaining the node's location by using an additional hardware, where these types of protocol generate well-controlled events like light spots or sound waves in the network which is not economically feasible. On the other hand Range -free techniques do not estimate absolute distance, these types of protocols uses the radio connectivity to infer proximity, relative proximity, and location. These two localization categories have different trade-offs when both are considered by: hardware requirements, localization accuracy, communication overhead and the infrastructure [16].

The GPS (Global Positioning System) based localization techniques is one of the widely used method by most of the WSN applications like range to several satellites by evaluating the difference in the time of arrival (TOA) of signal from different satellites. Though GPS based localization scheme contains a soaring level of accuracy, but it considers that a line of sight exists among the devices to be localizes and satellites as well as the need of additional hardware which consumes the momentous power makes it infeasible in terms of economic viability and technical accuracy in various conditions. In contrast the basic approach in Range free Localization is on assumptions of close proximity by the radio connectivity of nodes.

This approach is applicable in various scenarios of WSN as it basically depends on radio connectivity. In recent past various range free localization techniques has been proposed by various researchers [12]. Many such correlated work is described in the section related Work. In various aspect due to power exhaustion of the nodes, animus interference, deployment failures or because of physical obstacles such as valley ,mountains, water or building a non connected or a WSN with obstacle synergy arises [4] [5] [15]. Among various other reasons to have obstacle in radio communication one of the obvious reason is a non-uniform distribution due to random deployment of WSN apart from other reasons of non - uniformity into network such as unbalanced power consumptions which leads communication failure between node to node , where most of the existing range free localization cease its accuracy.

In this proposed paper, Architectural Model of Localization in Multi-scale Communication for Wireless Sensor Networks it can achieve a high accuracy by introducing a mobile anchor node into an obstacle existence WSN. The node localization inequality is mapped as a convex optimization algorithm and a simulation is performed.

The rest of this paper is organized as follows. We discuss related work in Section 2. The Mobile Anchor assisted Localization and the convex optimization algorithm for localization is described in Section 3. Section 4 presents the Simulation Graphical user Interface and experimental results. Finally, Section 5 concludes this paper with some discussions on the future work.

2. Related Work

The research in WSN started in the 80s by DARPA [13]. The earlier sensor networks were used in the military application for example Sound Surveillance System (SOSUS). The various applications such as Physical security (Detecting intruders), Medical (Patients in a hospital, Habitat monitoring (Wildlife, plants), Environmental (Tracking forest fires, pollution, Smart buildings, Air traffic control and Surveillance, most of the applications require location of the sensor.

The TOA(Time of Arrival)[14] example GPS which uses a satellite constellation of at least 24 satellites with atomic clocks, where Satellites broadcast precise time and Estimate distance to satellite using signal TOA using Trilateration. The AOA(Angle of arrival)[15] uses antenna array to measure direction of neighbors , where special landmarks have compass + GPS, broadcast location and bearing is used and flood beacons, update bearing along the way then once bearing of three landmarks is known, it calculates the position.

The method, basic APIT scheme Anchors are location aware sensors in the sensor network. APIT employs area based approach to isolate triangular regions between beaconing nodes. Once the area is known the COG calculation is performed for the location. In the Range Free Schemes the Centroid Localization is a simple and easy to implement where nodes receive beacon from anchor nodes. In another method such as DV -Hop localization which maintains a running hop count from beacon nodes and finds the average hop length and finally using trilateration unknown nodes are localized. In Amorphous localization which is similar to DV-Hop algorithm except that different approach used to estimate the average distance of a single hop [13].

3. Proposed System

3.1 Mobile Anchor assisted localization

The preface of Mobile Anchor assisted localization is described in this section [6]. In wireless sensor networks, node can be determined by accepting the beacon signal from the one-hop anchor depending upon whether the node falls in the transmission range or not. The Anchor node can adjust its range according to the adjustment of transmission radius by tuning the transmission power [7]. The algorithm for the protocol assumes L level for transmission power and related transmission range ($t_{ri,\ where\ i}$= 1… L), for the Anchor nodes. It is being assumed that the location of the mobile anchor node is known either by GPS receiver or by any other means. While movement of Anchor nodes, it transmits the beacon signals uninterruptedly with the attachment of its ID, Current Status, Transmission Radius and Transmission Power by the variation of power level, as shown in figure number 2(b). Based on the information received from beacon the unknown position sensor node's effective parameters for its state or location can be evaluated as shown in Figure 2(c), 2(d), 2(e).

3.2 Localization Algorithm Using Convex optimization

In lack of isotropic properties and heterogeneity of devices the actual transmission range differs, because of this reason the communication are intervallic among the nodes in the instance when the nodes distance is within their superlative transmission radius. The other possibilities of communication of two nodes arise when the node's comparative distance is larger than their transmission range. The consequence of any localization algorithm limits its assurance for full coverage due to radio indiscretion and obstacles and hence yields to an infeasible solution [8]-[9].To achieve optimized localization accuracy in both the cases of wireless sensor network with an obstacle and without an obstacle, a narrative approach of Range Free Based Convex Method has been proposed.

TABLE 1 NOTATION AND ACRONYMS

s_p	Unknown position of sensor
c_p	Current Position of Mobile Anchor
t_r	Transmission radius of Mobile Anchor
r	Internal Radius

3.3 Formulation of Problem and Mapping with Convex Optimization:

Let C_p be the current position for the Mobile Anchor with transmission radius is t_r, and the unpredicted-positioned sensor receives beacon signals at position $S_{p,}$ It will be satisfying the equation number (1) else equation number (2). The unpredicted -positioned sensors can obtain a set of dissimilarity by varying transmission power levels of Mobile Anchor node at different positions S_p, should satisfy equation number (3).

$$||S_p - C_p|| \leq tr \quad (1)$$

$$||S_p - C_p|| > tr \quad (2)$$

$$r_i < ||S_p - C_p|| \leq tri, \ i = 1,2,\ldots\ldots\ldots,L \quad (3)$$

Therefore Mobile Anchor assisted localization approach with multi-scale communication range can be mapped as a problem of a set of quadratic intolerance.

The localization approach based on the Mobile Anchor with variable transmission power can be successfully converted into the problem of solving a group of quadratic in tolerance (3). Whereas due to convoluted transmitted environment situation of set of intolerance may have a solution or it may not have a solution. For the scenario like network with obstacle in the boundary [10] and without obstacle a localization algorithm mapped as convex optimization is proposed, where the conceptualization of architecture checks the condition where, r < Mode (unknown position of sensor – Current Position of Mobile Anchor) < Transmission radius of Mobile Anchor and calculate the optimal position at circumference of a circle having center point as current location of MA and radius as an average of transmission radius of Mobile Anchor and Internal radius. Further non convex minimization for localization is estimated by estimate by equation number (4).

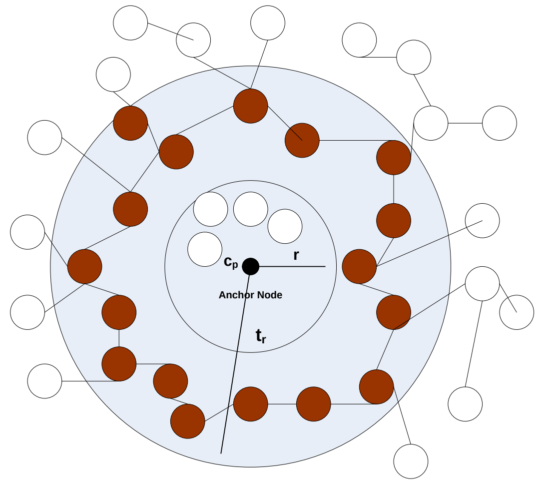
$$(||S_p - C_p|| - r)^2 + (||S_p - C_p|| - t_r)^2 \quad (4)$$

The optimal localization is evaluated under the inequalities constraints by equation number (5). Min @ Sp for all i

$$\sum [(||S_p - C_p|| - r_i)^2 + (||S_p - C_p|| \, t_{ri})^2] \quad (5)$$

The equation (5) remains a non-convex, which needs to be transforming as a problem of convex as it cannot be approximated by implementing convex relaxation techniques. The equation number is turned into a convex problem which is finally transformed to the epigraph as convex cone programming problem, which is being solved by Euclidian distance algorithm and interior point algorithm. The architectural network model is shown in Figure-(1).

The range free algorithm proposed in [11], author claims it as the only algorithm using convex optimization to solve the problem of localization when feasible set is empty. In this paper an addition possible orientation of obstacle is placed in simulation and simulation is performed with 100 nodes and two different algorithms Euclidian distance and interior point has been implemented with and without obstacles. The GUI is shown in section 4.



Node Not in Range of Anchor Node	Node in the Range of Anchor Node	Anchor Node

Fig -1 Architectural computational Model

4. Simulation Work

The simulation in this section proposed has been carried on Intel Pentium Dual Core E2160 CPU with 1.8 GHz and 2 GB RAM. The designed using Matlab 7.2 is considered where 100 sensor nodes will be distributed randomly with specific transmission radii (R) and transmission area (r) deployed for mobile anchor nodes.

The Simulation is preformed in the area of A X A with random deployment of 100 nodes as shown in figure number 2(a). The provision of changing anchor node transmission radius is provides with an initial value of 15.

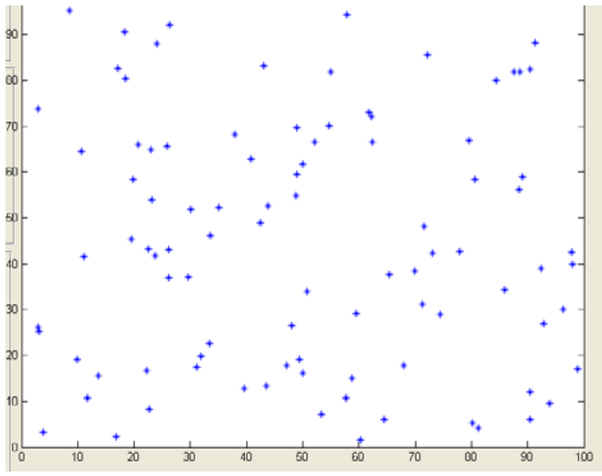


Fig -2(a) Deployment of the node (Network Deployment)

The Anchor node has been given a random path to move across the network as shown in Figure 2(b). The simulation is performed with both Euclidian distance and interior point with varying option of non- physical obstacle and with physical obstacle, which has been shown in Figure 2(c), 2(d), 2(e) and 2(f).

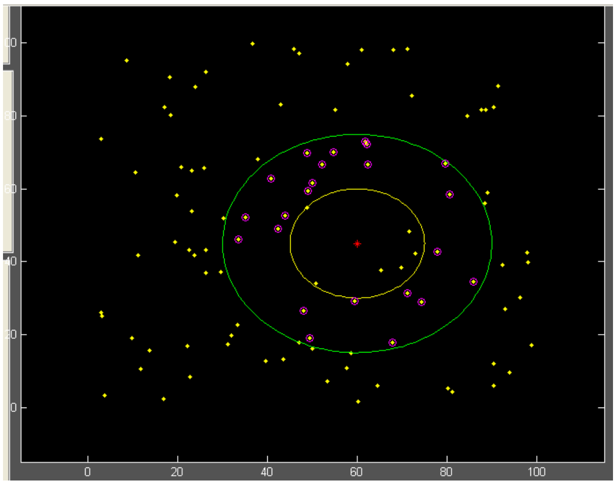


Fig - 2(b) Anchor Node Deployment

4.1 Simulation Scenario one:

The Euclidian Distance algorithm is implemented without obstacle as shown in the Figure-2(c) and the result set is shown in Table no 2.

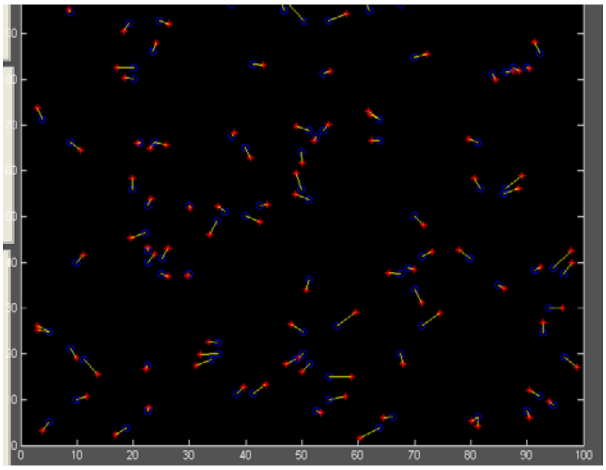


Fig -2(c): Localization using Euclidian Distance Algorithm without obstacle

TABLE 2

No. of Nodes	***Algorithm***	***Obstacle***	***Simulation Time (in Sec)***	***Localization error (%)***
100	Euclidian Distance	No	33.5013	0.14996

4.2 Simulation Scenario two:

The interior point algorithm is implemented without obstacle as shown in the Figure-2(d) and the result set is shown in Table no 3.

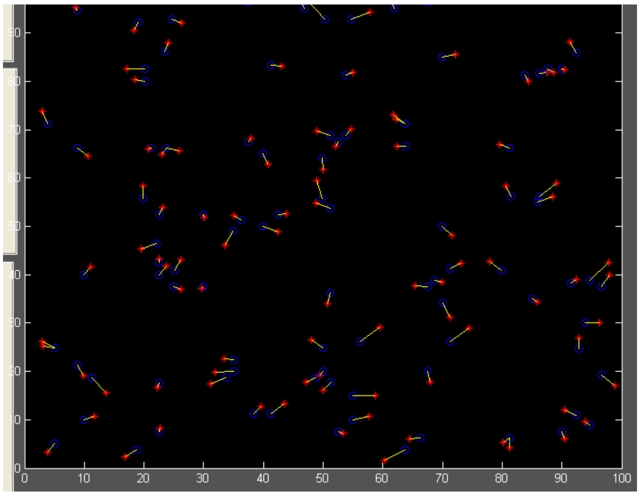


Fig -2(d): localization using interior distance algorithm without obstacle

TABLE 3

No. of Nodes	***Algorithm***	***Obstacle***	***Simulation Time (in Sec)***	***Localization error (%)***
100	Interior Point	No	10.2275	0.14997

4.3 Simulation Scenario three:

Implementation of Euclidian distance algorithm with obstacle at location of 40, 60 of size 20as shown in the Figure-2(e) and the result set is shown in Table no 4.

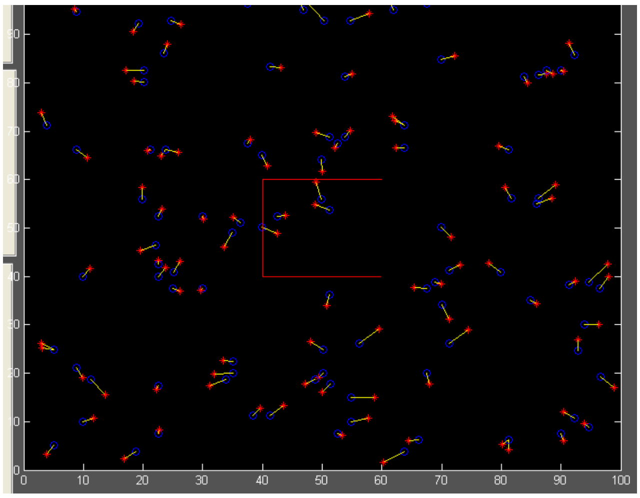


Fig -2(e): Localization using Euclidian distance algorithm with obstacle

TABLE 4

No. of Nodes	*Algorithm*	*Obstacle*	*Simulation Time (in Sec)*	*Localization error (%)*
100	Euclidian Distance	yes	34.1901	0.14996

4.4 Simulation Scenario four:

The interior point algorithm is implemented with the obstacle at location of 40,60 of size 20 as shown in Figure-2(f) and the result set is shown in Table no 5.

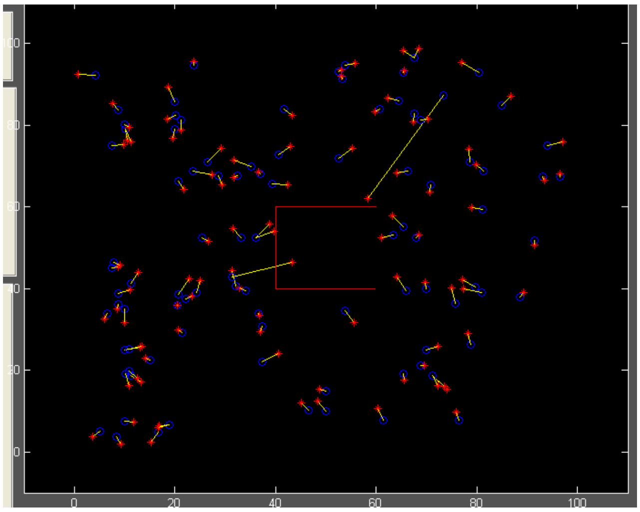


Fig -2(f): Localization using interior point algorithm with obstacle

TABLE 5

No. of Nodes	*Algorithm*	*Obstacle*	*Simulation Time (in Sec)*	*Localization error (%)*
100	Interior Point	yes	9.8746	0.18091

Graph generation for the %error for proposed method form the simulation data by Euclidian method with obstacle and without obstacle along with Interior point method with obstacle and without obstacle. The graph for all data is given in Figure-3(a).

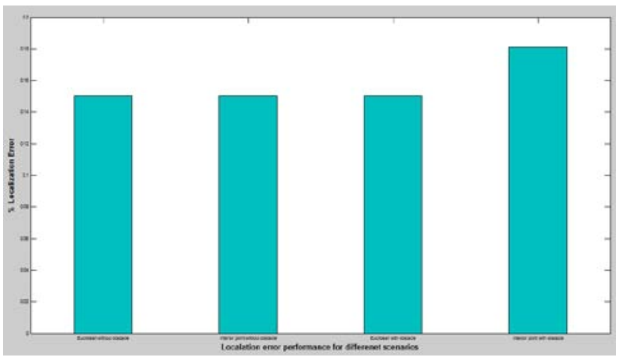


Fig 3(a): Localization error performance in different scenario

Graph generation for the simulation time (sec.) for proposed method form the simulation data by Euclidian method with obstacle and without obstacle along with Interior point method with obstacle and without obstacle. The graph for all data is given in Figure-3(b).

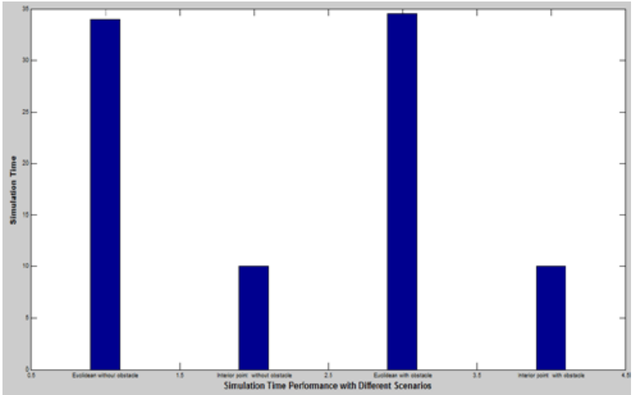


Fig 3(b): Simulation Time Performance with different scenarios

5. Conclusion and Future works

The mobility aided mobile anchor node based WSN in the presence of obstacles has been implemented for new accommodating localization system. A convex localization algorithm has been presented to address the effects of non-ideal transmission of radio signals. It has been shown in the simulation results that the proposed accommodative localization system can appreciably get better the localization accurateness by including a mobile element. In future work, we intend to verify and improve the proposed accommodative localization system using real sensors in a mobility assisted wireless sensor networks for the design of energy efficient architecture.

Acknowledgments

The author would like to thank the Dr. C.N Khairnar of Military College of Telecommunication Engineering, India and Mr. Ranjit Patro-Project Team coordinator of R&D-Samsung India for their valuable guidance and support. We would like to thank the research centre Toch of the Anna University Coimbatore and anonymous reviewer for the productive feedback to form value added paper.

References

[1] Mayuresh, M. Patil, Umesh Shaha, U. B. Desai, S. N. " Merchant, "Localization in Wireless Sensor Networks using Three Masters," IEEE ICPWC.INDIAN INSTITUTE OF TECHNOLOGY, BOMBAY, December 3, 2008.

[2] Radu Stoleru, John A. Stankovic, Sang Son, "Robust Node Localization for Wireless Sensor Networks," IEEE ACM. June 25-26, 2007
[3] Radu Stoleru, John A. Stankovic, and Sang H. Son, "On Compos ability of Localization Protocols for Wireless Sensor Networks", IEEE PCM.
[4] Xu Li, Nathalie Mitton, Isabelle Ryl and David Simplot, "Localized Sensor Self-deployment with Coverage Guarantee in Complex Environment," http://www.springerlink.com/content/l4598140135653xp
[5] Yuan XU, Shubo QIU, Men g HOU, "Reconfigure ZigBee Network Based on System Design," Scientific Research Publishing, doi:10.4236/wsn.2009.13027 October 2009.
[6] Kuo-Feng Ssu; Chia-Ho Ou; Jiau, H.C, "Localization with mobile anchor points in wireless sensor networks," IEEE Vehicular Technology Society. 23 May 2005.
[7] Sohail Jabbar, Muhammad Zubair Aziz, Abid Ali Minhas, Danial Hussain, "A Novel Power Tuning Anchors Localization Algorithm for Mobile Wireless Sensor Nodes," IEEE International Conference on Computer and Information Technology. 29 -July, 2010.
[8] http:// academicearth.org/courses/convex- optimization-ii. "Convex optimization,"
[9] Ming-Yong LIU, Wen-Bai LI and Xuan PEI, "Convex Optimization Algorithms for Cooperative Localization in Autonomous Underwater Vehicles," T.C.A.O.A, Elsevier B.V. 2010.
[10] Michael P. Vitus, Steven L. Waslander, and Claire J. Tomlin. Locally optimal decomposition for autonomous obstacle avoidance with the tunnel-MILP algorithm. submitted to IEEE Transactions on Robotics, 2010.
[11] Hongyang Chen, Qingjiang Shi, Pei Huang, H.Vincent Poor, and Kaoru Sezaki, "Mobile Anchor Assisted Node Localization for Wireless Sensor Networks," arXiv: 0908.0515v1 [cs.IT], 4 Aug 2009.
[12] Yi Zhou, Bao-jun Gu, Kai Chen, Jian-bo Chen and Hai-bing Guan, "An range-free localization scheme for large scale underwater wireless sensor networks," 2009.
[13] Tian He, Chengdu Huang, Brian M. Blum, John A. Stankovic, Tarek F. Abdelzaher, 'Range-free localization and its impact on large scale sensor networks', ACM Transactions on Embedded Computing Systems (TECS), Volume 4 Issue 4, November 2005.
[14 B. H. Wellenhoff, H. Lichtenegger and J. Collins, "Global Positioning System: Theory and Practice. Fourth Edition, Springer Verlag," 1997.
[15] Dragos Niculescu and Badri Nath. "Ad-Hoc Positioning System (APS) Using AoA," IEEE InfoCom, 2003
[16] Mo Li, Yunhao Liu, 'Rendered path: range-free localization in anisotropic sensor networks with holes', Journal IEEE/ACM Transactions on Networking (TON) archive, Volume 18 Issue 1, February 2010.

PROFILE

Mr. S. Swapna Kumar is a Professor working in the department for Electronics & Communication Engineering branch in AXIS College of Engineering & Technology. He is doing his research study in Anna University Coimbatore in the area of wireless sensor network. He attended various workshops and presented paper in the International conference. He is the member of IEEE.

Dr. M. Nanda Kumar is an Associate Professor in Electrical Engineering, in Government Engineering College, India. He completed his doctorate degree from IIT Madras. He is having several years of teaching experience and several publications.

Dr. Sheeba V.S is a Professor in Electronics & Communication Engineering, in Government Engineering College. She completed her doctorate degree from NIT Calicut. She is having several years of teaching experience. She published papers in national & international journals.

Classification of EEG data using FHT and SVM based on Bayesian Network

V. Baby Deepa[1] and Dr. P. Thangaraj[2]
[1]Dept. of Software Engnineering, M.Kumarasamy College of Engineering
Karur, Tamil Nadu, 639 113, India

[2]Dept. of Computer Science and Engineering, Bannari Amman Institute of Technology
Sathyamangalam, Erode, Tamil Nadu, 638 401, India

Abstract
Brain Computer Interface (BCI) enables the capturing and processing of motor imagery related brain signals which can be interpreted by computers. BCI systems capture the motor imagery signals via Electroencephalogram or Electrocorticogram. The processing of the signal is usually attempted by extracting feature vectors in the frequency domain and using classification algorithms to interpret the motor imagery action.

In this paper we investigate the motor imagery signals obtained from BCI competition dataset IVA using the Fast Hartley Transform (FHT) for feature vector extraction and feature reduction using support vector machine. The processed data is trained and classified using the Bayes Net.

Keywords: *Brain Computer Interface (BCI),Electroencephalogram (EEG), Fast Hartley Transform (FHT), Bayes Net (BN)*

1. Introduction

A Brain Computer Interface [1], also known as Direct Neural Interface or a Brain Machine Interface, is a direct communication pathway between a human or animal brain (or brain cell culture) and an external device. BCIs are focused on assisting, augmenting or repairing human cognitive or sensory-motor functions. BCIs would act in two ways. In the case of one-way BCIs, computers either accept commands from the brain or send signals to it but not both. Two way BCIs allow brains and external devices to exchange information in both directions but it is not yet successful in the aspect of implantation.

Brain-machine interfaces [2] help paralyzed patients by re-routing movement-related signals around damaged parts of the nervous system. With recent advancement in technology and knowledge, the researchers now conceivably attempt to produce BCIs that augment human functions rather than simply restoring them.

In this paper, the motor imagery signals obtained from BCI competition dataset IVA are investigated. They are further exposed to Fast Hartley Transform and Support Vector Machine for feature vector extraction and feature reduction respectively. In this paper, Section I gives an introduction to BCI, Section II describes the dataset and EEG data, Section III and IV explain the feature vector extraction and feature reduction using FHT and SVM, Section V briefs about the classification using Bayes Net and is followed by the experimental results and conclusion.

2. Data Set and EEG

The IV A dataset used in the brain computer interface competition provided by Intelligent Data Analysis Group has been taken for investigation. It consists of recordings from five healthy subjects who sat in a chair with arms resting on armrests. Visual cues indicated for 3.5 s which of the following 3 motor imageries the subject should perform: (L) *left* hand, (R) *right* hand, (F) right *foot*. The presentation of target cues was intermitted by periods of random length, 1.75 to 2.25 s, in which the subject could relax. Given are continuous signals of 118 EEG channels and markers that indicate the time points of 280 cues for each of the 5 subjects (*aa*, *al*, *av*, *aw*, *ay*). Subject aa was used in our study.

The abnormalities related to electrical activity of the brain could be detected using a test called electroencephalogram (EEG) [3] which tracks and records brain wave patterns. Electrodes (Small metal discs with thin wires) are placed on the scalp, which send signals to a computer to record the results. A recognizable pattern of EEG is identified as a normal electrical activity in the brain. EEG is used to detect abnormal patterns of brains that indicate seizures and other problems. EEG is also performed to diagnose monitor seizure disorders, sleep disorders and other changes in behavior. It is also used to evaluate brain activity after a severe head injury or before heart or liver transplantation.

3. Feature Vector Extraction

In this paper the feature vector extraction is performed on the dataset using Fast Hartley Transform (FHT). FHT [4] is a technique helps to extract the feature vector efficiently. A discrete Hartley transform (DHT) is a Fourier-related transform of discrete, periodic data similar to the discrete Fourier transform (DFT), with analogous applications in signal processing and related fields. Its main distinction from the DFT is that it transforms real inputs to real outputs, with no intrinsic involvement of complex numbers. As the DFT is the discrete analogue of the continuous Fourier transform, the DHT is the discrete analogue of the continuous Hartley transform.

Since there are fast algorithms for the DHT analogous to the fast Fourier transform (FFT), the DHT was proposed as a more efficient computational tool in the common case where the data are purely real. It was subsequently argued, however, that specialized FFT algorithms for real inputs or outputs can ordinarily be found with slightly fewer operations than any corresponding algorithm for the DHT. The DHT analogue of the Cooley-Tukey algorithm is commonly known as the Fast Hartley Transform (FHT) [5] algorithm.

Discrete Hartley transform is an analogue of discrete Fourier transform for real data. The Hartley transform takes a real sequence as an input. The result is also a real sequence:

$$H_k = \sum_{n=0}^{N-1} x_n \cdot \left(\cos\left(\frac{2\pi nk}{N}\right) + \sin\left(\frac{2\pi nk}{N}\right) \right)$$

It was considered, for sometime, that Hartley transform can be a faster alternative to the real Fourier transform, but later it was found out that there are FFT algorithms, which are a little more efficient than the corresponding FHT [6] algorithms. An integral transform which shares some features with the Fourier transform, but which (in the discrete case), multiplies the integral kernel by

$$\cos\left(\frac{2\pi k n}{N}\right) - \sin\left(\frac{2\pi k n}{N}\right) \quad (1)$$

instead of

$$e^{-2\pi i k n/N} = \cos\left(\frac{2\pi k n}{N}\right) - i \sin\left(\frac{2\pi k n}{N}\right). \quad (2)$$

The Hartley transform produces real output for a real input, and is its own inverse. It therefore can have computational advantages over the discrete Fourier transform, although analytic expressions are usually more complicated for the Hartley transform.

The discrete version of the Hartley transform can be written explicitly as

$$\mathcal{H}[a] \equiv \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} a_n \left[\cos\left(\frac{2\pi k n}{N}\right) - \sin\left(\frac{2\pi k n}{N}\right)\right] \quad (3)$$

$$= \Re \mathcal{F}[a] - \Im \mathcal{F}[a], \quad (4)$$

where $\mathcal{F}$ denotes the Fourier transform. The Hartley transform obeys the convolution property

$$\mathcal{H}[a * b]_k = \tfrac{1}{2}\left(A_k B_k - \overline{A}_k \overline{B}_k + A_k \overline{B}_k + \overline{A}_k B_k\right), \quad (5)$$

where

$$\overline{a}_0 \equiv a_0 \quad (6)$$

$$\overline{a}_{n/2} \equiv a_{n/2} \quad (7)$$

$$\overline{a}_k \equiv a_{n-k}. \quad (8)$$

Like the fast Fourier transforms, there is a fast version of the Hartley transform. Decimation in time algorithm makes use of

$$\mathcal{H}_n^{\text{left}}[a] = \mathcal{H}_{n/2}[a^{\text{even}}] + X \mathcal{H}_{n/2}\left[a^{\text{odd}}\right] \quad (9)$$

$$\mathcal{H}_n^{\text{right}}[a] = \mathcal{H}_{n/2}[a^{\text{even}}] - X \mathcal{H}_{n/2}\left[a^{\text{odd}}\right], \quad (10)$$

where X denotes the sequence with elements

$$a_n \cos\left(\frac{\pi n}{N}\right) - \overline{a}_n \sin\left(\frac{\pi n}{N}\right). \quad (11)$$

Decimation in frequency algorithm makes use of

$$\mathcal{H}_n^{\text{even}}[a] = \mathcal{H}_{n/2}\left[a^{\text{left}} + a^{\text{right}}\right] \quad (12)$$

$$\mathcal{H}_n^{\text{odd}}[a] = \mathcal{H}_{n/2}\left[X\left(a^{\text{left}} - a^{\text{right}}\right)\right]. \quad (13)$$

The discrete Fourier transform

$$A_k \equiv \mathcal{F}\,[a] = \sum_{n=0}^{N-1} e^{-2\pi i k n/N}\, a_n \tag{14}$$

can be written

$$\begin{bmatrix} A_k \\ A_{-k} \end{bmatrix} \quad \sum_{n=0}^{N-1} \underbrace{\begin{bmatrix} e^{-2\pi i k n/N} & 0 \\ 0 & e^{2\pi i k n/N} \end{bmatrix}}_{\mathsf{F}} \begin{bmatrix} a_n \\ a_n \end{bmatrix}$$

So $$= \sum_{n=0}^{N-1} \underbrace{\frac{1}{2}\begin{bmatrix} 1-i & 1+i \\ 1+i & 1-i \end{bmatrix}}_{\mathsf{T}^{-1}} \underbrace{\begin{bmatrix} \cos\left(\frac{2\pi k n}{N}\right) & \sin\left(\frac{2\pi k n}{N}\right) \\ -\sin\left(\frac{2\pi k n}{N}\right) & \cos\left(\frac{2\pi k n}{N}\right) \end{bmatrix}}_{\mathsf{H}} \underbrace{\frac{1}{2}\begin{bmatrix} 1+i & 1-i \\ 1-i & 1+i \end{bmatrix}}_{\mathsf{T}} \begin{bmatrix} a_n \\ a_n \end{bmatrix},$$

$$\mathsf{F} = \mathsf{T}^{-1}\, \mathsf{H}\, \mathsf{T}. \tag{15}$$

4. Feature Reduction

A support vector machine (SVM) [7] is a concept in computer science for a set of related supervised learning methods that analyze data and recognize patterns, used for classification and regression analysis. The standard SVM takes a set of input data and predicts, for each given input, which of two possible classes the input is a member of, which makes the SVM a non-probabilistic binary linear classifier. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

A support vector machine [8] constructs a hyperplane or set of hyperplanes in a high- or infinite- dimensional space, which can be used for classification, regression, or other tasks. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training data points of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier. The original problem may be stated in a finite dimensional space; it often happens that the sets to discriminate are not linearly separable in that space. For this reason, it was proposed that the original finite-dimensional space be mapped into a much higher-dimensional space, presumably making the separation easier in that space.

To keep the computational load reasonable, the mapping used by SVM [9] schemes are designed to ensure that dot products may be computed easily in terms of the variables in the original space, by defining them in terms of a kernel function $K(x,y)$ selected to suit the problem. The hyperplanes in the higher dimensional space are defined as the set of points whose inner product with a vector in that space is constant. The vectors defining the hyperplanes can be chosen to be linear combinations with parameters α_i of images of feature vectors that occur in the data base. With this choice of a hyperplane, the points x in the feature space that are mapped into the hyperplane are defined by the relation:

$$\sum_i \alpha_i K(x_i,x) = constant$$

if $K(x,y)$ becomes small as y grows further from x, each element in the sum measures the degree of closeness of the test point x to the corresponding data base point x_i. In this way, the sum of kernels above can be used to measure the relative nearness of each test point to the data points originating in one or the other of the sets to be discriminated. The set of points x mapped into any hyperplane can be quite convoluted as a result allowing much more complex discrimination between sets which are not convex at all in the original space.

5. Classification Using Bayes Net

Bayesian networks (BNs) [10], also known as *belief networks* (or Bayes nets for short), belong to the family of probabilistic *graphical models* (GMs). These graphical structures are used to represent knowledge about an uncertain domain. In particular, each node in the graph represents a random variable, while the edges between the nodes represent probabilistic dependencies among the corresponding random variables. These conditional dependencies in the graph are often estimated by using known statistical and computational methods. Hence, Bayesian Networks combine principles from graph theory, probability theory, computer science, and statistics.

Graphical Models with *undirected edges* are generally called *Markov random fields* or *Markov networks*. These networks provide a simple definition of independence between any two distinct nodes based on the concept of a *Markov blanket*. Markov networks are popular in fields such as statistical physics and computer vision.

BNs [11] correspond to another GM structure known as a *directed acyclic graph* (DAG) that is popular in the statistics, the machine learning, and the artificial intelligence societies. BNs are both mathematically rigorous and intuitively understandable. They enable an effective representation and computation of the joint probability distribution (JPD) over a set of random variables.

The structure of a DAG is defined by two sets: the set of nodes (vertices) and the set of directed edges.
The nodes represent random variables and are drawn as circles labeled by the variable names. The edges represent direct dependence among the variables and are drawn by arrows between nodes. In particular, an edge from node *Xi* to node *Xj*

represents a statistical dependence between the corresponding variables. Thus, the arrow indicates that a value taken by Variable *Xj* depends on the value taken by variable *Xi*, or roughly speaking that variable *Xi* "influences" *Xj* . Node *Xi* is then referred to as a *parent* of *Xj* and, similarly, *Xj* is referred to as the *child* of *Xi*.

An extension of these genealogical terms is often used to define the sets of "descendants" – the set of nodes that can be reached on a direct path from the node, or "ancestor" nodes – the set of nodes from which the node can be reached on a direct path. The structure of the acyclic graph guarantees that there is no node that can be its own ancestor or its own descendent. Such a condition is of vital importance to the factorization of the joint probability of a collection of nodes as seen below. Note that although the arrows represent direct causal connection between the variables, the *reasoning process* can operate on BNs by propagating information in any direction.

A BN [12] reflects a simple conditional independence statement. Namely that each variable is independent of its non descendents in the graph given the state of its parents. This property is used to reduce, sometimes significantly, the number of parameters that are required to characterize the JPD of the variables. This reduction provides an efficient way to compute the posterior probabilities given the evidence.
In addition to the DAG structure, which is often considered as the "qualitative" part of the model, one needs to specify the "quantitative" parameters of the model.

The parameters are described in a manner which is consistent with a Markovian property, where the conditional probability distribution (CPD) at each node depends only on its parents. For discrete random variables, this conditional probability is often represented by a table, listing the local probability that a child node takes on each of the feasible values – for each combination of values of its parents. The joint distribution of a collection of variables can be determined uniquely by these local conditional probability tables (CPTs).

Bayesian networks are used to represent essential information in databases in a network structure. The network consists of edges and vertices, where the vertices are *events* and the edges *relations* between events. A simple Bayesian network is illustrated in figure where symptoms are dependent on a disease, and a disease is dependent on age, work and work environment. Bayesian networks are easy to interpret for humans, and are able to store *causal relationships*, that is, relations between causes and effects. The networks can be used to represent domain knowledge, and it is possible to control inference and produce explanations on a network.

A simple usage of Bayesian networks is denoted naive Bayesian classification. These networks consist only of one parent and several child nodes. Classification is done by considering the parent node to be a hidden variable (H in the figure) stating which class (child node) each object in the database should belong to. An existing system using naive Bayesian classification is AutoClass.

The theoretical foundation for Bayesian networks is Bayes rule, which states:

$$P(H \mid e) = \frac{P(e \mid H)P(H)}{P(e)}$$

where *H* is a hypothesis, and *e* an event. $P(e \mid H)$ is the posterior probability, and $P(H)$ is the prior probability. To give a formal definition of Bayesian networks, we introduce some terminology which is taken from:

If a subset of *Z* nodes in a graph *G* intercepts all paths between the nodes *X* and *Y* (written), then this $\langle X \mid Y \mid Z \rangle_G$ corresponds to *conditional independence* between *X* and *Y* given *Z*:

$$\langle X \mid Y \mid Z \rangle_G \Rightarrow I(X, Z, Y)_M$$

conversely:

$$I(X, Z, Y)_M \Rightarrow \langle X \mid Y \mid Z \rangle_G$$

with respect to some dependency model M.

A Directed, Acyclic Graph (DAG) *D* is said to be a *I-map* of a dependency model *M* if for every three disjoint sets of vertices, *X*, *Y* and *Z* we have:

$$\langle X \mid Y \mid Z \rangle_G \Rightarrow I(X, Z, Y)_M$$

A DAG is a minimal I-map of *M* if none of its arrows can be deleted without destroying its I-mapness. Given a probability distribution *P* on a set of variables *U*, a DAG is called a *Bayesian Network* of P if $D = (U, \vec{E})$ and only if *D* is a minimal I-map of *P*.

A Bayesian network is shown in Fig, representing the probability distribution P:

$$P(x_6 \mid x_5)P(x_5 \mid x_2, x_3)P(x_4 \mid x_2, x_1)P(x_3 \mid x_1)P(x_2 \mid x_1)P(x_1)$$

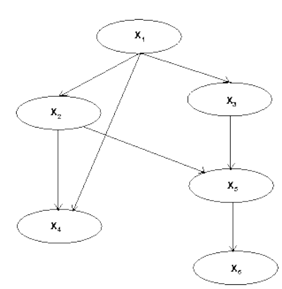


Fig. 1: A Bayesian Network Representing the Distribution P.

6. Experimental Result

Bayes Net

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	103	61.3095 %
Incorrectly Classified Instances	65	38.6905 %
Kappa statistic	0.2178	
Mean absolute error	0.4364	
Root mean squared error	0.497	
Relative absolute error	87.4669 %	
Root relative squared error	99.5084 %	
Total Number of Instances	168	

=== Detailed Accuracy By Class ===

TPRate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0.5	0.284	0.615	0.5	0.552	0.631	hand
0.716	0.5	0.612	0.716	0.66	0.631	foot
Weighted Avg.						
0.613	0.397	0.613	0.613	0.608	0.631	

=== Confusion Matrix ===

a	b	<-- classified as
40	40	a = hand
25	63	b = foot

7. Conclusion

In this paper feature vector was extracted from the BCI competition IVA dataset using Fast Hartley Transform. Sub set selection of the obtained features after normalization was achieved using Support Vector Machine. Decision tree and logistic regression based on Bayes probability was used to train and classify the extracted sub features. Results show that the classification accuracy is over 60% in Bayesian Network. Further investigation has to be done to improve the classification accuracy on a small number of attributes

References

[1]. E.E. Sutter, The brain response interface: communication through visually induced electrical brain responses 1992, Journal of Microcomputer Applications, v. 15, pp. 31-45.

[2]. Chang S. Nam, Gerwin Schalk, Melody Moore Jackson: Current Trends in Brain-Computer Interface (BCI) Research and Development. Int. J. Human. Computer. Interaction 27(1): 1-4 (2011)

[3]. Michael Adamaszek, Sebastian Olbrich and Jürgen Gallinat, The Diagnostic Value of Clinical EEG in Detecting Abnormal Synchronicity in Panic Disorder Journal of Clinical EEG & Neuroscience, July, 2011.

[4]. R. N. Bracewell, "The fast Hartley transform," Proc. IEEE, vol. 72, no. 8, pp. 1010-1018, Aug. 1984.

[5]. H. J. Meckelburg and D. Lipka, "Fast Hartley transform algorithm," Electronics Letters, vol. 21, no. 8, pp. 311-313, Apr. 1985.

[6]. Bracewell, R. N. *The Hartley Transform.* New York: Oxford University Press, 1986.

[7]. T. Joachims, SVM Light Support Vector Machine, 2002, D. Koller, M. Sahami, Hierarchically classifying documents using very few words, in: Proceedings of the 14th International Conference on Machine Learning, 1997, pp. 170–178.

[8]. Shailendra Kumar, Shrivastava Preeti Jain Effective Anomaly based Intrusion Detection using Rough Set Theory and Support Vector Machine, International Journal of Computer Applications,IJCA, Journal Number 3 - Article 8 Year of Publication: 2011, doi: 10.5120/2261-2906

[9]. Hui Xue; Songcan Chen; Qiang Yang; Structural Regularized Support Vector Machine: A Framework for Structural Large Margin Classifier, Neural Networks, IEEE Transactions on , Issue Date: April 2011, Volume: 22 Issue:4, On page(s): 573 – 587, ISSN: 1045-9227

[10]. Pearl, J. & Russel, S. (2001). Bayesian networks. Report (R-277), November 2000, in Handbook of BrainTheory and Neural Networks, M. Arbib, ed, MIT Press, Cambridge, pp. 157–160.

[11]. Friedman, N. & Goldszmidt, M. (1996). Learning Bayesian networks with local structure, in Proceedings of the 12th Conference on Uncertainty in Artificial Intelligence, Portland, August 1–4 1996.

[12]. Boutilier, C., Friedman, N., Goldszmidt, M. & Koller, D. (1996). Context-specific independence in Bayesian networks, in Proceedings of the 12th Conference on Uncertainty in Artificial Intelligence, Portland, August 1–4 1996, pp. 115–123.

V.Baby Deepa, received her Bachelors and Masters degree in Computer Science from Barathidasan University, Trichy and received her M.Phil. Degree as well from the same university. She has 14 years of teaching experience. Besides being an Assistant professor in the faculty of software Engineering, she is serving as the head for the same faculty in M.Kumarasamy College of Engineering, Karur. She has presented more than 16 papers on various topics including national, international conference and journals. She is a research scholar of Anna University Chennai and her research area is Fuzzy and Data Mining.

Dr.P.Thangaraj, received Bachelor's and Master's in Mathematics from Madras University in 1981 and 1983. He completed his M.Phil degree in the year 1993 from Bharathiar University and research work on Fuzzy Metric Spaces and awarded Ph.D degree by Bharathiar University. He completed the post graduation in Computer Applications at IGNOU in 2005 and Master of Engineering degree in Computer Science in the year 2007 at Vinayaka Missions University. Currently he is a Professor Head of Computer Science and Engineering in Bannariamman College of Engineering and Technology.

Feature Extraction And Classification Of Oil Spills In Sar Imagery

Radhika Viswanathan[1] and Padmavathi Ganapathi[2]

**[1] Prof & Head, Department of Computer Science,
Sri Krishna Arts And Science College,
Coimbatore-641 008, India.**

**[2] Prof & Head, Department of Computer Science
Avinashilingam University for Women,
Coimbatore-641 043,India.**

Abstract

Synthetic Aperture RADAR (SAR) imaging system is used to monitor the marine system. Oil spill pollution plays a significant role in damaging marine ecosystem. One main advantages of SAR is that it can generate imagery under all weather conditions. In a SAR image dark spots can be generated by number of phenomena. The dark spots may be of algae, low wind areas, coastal areas and oil spills. The detected dark spots are then classified based on the features. The features of dark spot are extracted to discriminate oil spill from look-alikes. The textural and statistical features are extracted and analyzed for oil spill identification. This paper discusses about the different feature extraction and classification method for oil spill detection and their preliminary results.

Keywords: *oil spill, SAR, features, detection, classification, look-alikes.*

1. Introduction

Oil spills seriously damage the marine ecosystem and cause political and scientific concern since they have serious effects on fragile marine and coastal ecosystems. The amount of pollutant discharges and associated effects on the marine environment are important parameters in evaluating sea water quality. Illegal discharges from ships can indeed be eliminated by the strict enforcement of existing regulations and the control, monitoring and surveillance of maritime traffic. Several studies aiming at oil spill detection using SAR images have been implemented [1-5]. Any formation on the image which is darker than the surrounding area has a high probability of being an oil spill and needs further examination.

Although this process seems to be simple for a human operator, it contains three main difficulties if semi-automated or automated methods are used. First, fresh oil spills are brighter than older spills, thus cannot be easily discriminated. Second, areas surrounding dark areas can have various contrast values, depending on local sea state, oil spill type and image resolution. Third, other phenomena may appear as dark areas. Further classification of the dark areas to oil spills and look-alikes is the focus of this work.

Many research works are focused on the development of automated or semi-automated systems for oil spill detection are reported in literature. Kubat et al. (1998) developed a neural network for the classification of dark regions detected in a series of nine SAR images that served as a training set of the system. The complexities of such a system as well as the appropriate actions that have to be taken into consideration by potential tool developers in such fields were analyzed in detail. Input to the classifier was straightforward, though image preprocessing was not automated.

The classifier had an open architecture of rules so that it could embed user experience in several other fields apart from oil detection. Del Frate et al. (2000) also used neural network architecture for semi-automatic detection of oil spills on SAR images using a set of features characterizing a candidate oil spill as input vector. Solberg and Solberg (1996) and Solberg et al. (1999) produced a semi-automated classifier for oil spill detection, in which the objects with a high probability of being an oil spill were automatically detected. Three different categories of probability (low, medium and high) were recognized.

A rational processing procedure was adopted for 84 SAR images utilized. It involved pixel local thresholding based on wind level information, clustering of small pixel objects or partitioning of large pixel objects based on sizing criteria and feeding each individual cluster to a classifier operating on a stochastic processing basis. Ten different object characteristics were identified and classification was based on a Bayesian inference procedure. Fiscella et al. (2000) developed a stochastic classifier based on Mahalanobis statistical tests and classical compound probabilities.

A preprocessing tool was used in order to extract pixel objects from SAR images and classified them according to statistical criteria implemented on a total of 14 different characteristics of extracted clusters. In the present work a fully automated system for the identification of possible oil spills that resembles the expert's choice and decisions has been developed. The system comprises modules of supplementary operation and uses their contribution to the analysis and assignment of the probability of a dark image shape to be an oil spill. SAR images are read, located, land masked, filtered and thresholded so that the appropriate dark areas are extracted. Candidate oil spill objects are classified to determine the likeness of each individual object to be an oil spill. The output images and tables provide the user with all relevant information for supporting decision-making

The remainder of the paper is organized as follows. Section 2 describes SAR imaging of oil spills and lookalikes, Section 3 describes the feature extraction process. The classification problem is discussed in detail in Section 4, while Section 5 contains the experimental results. Section 6 presents conclusion.

2. SAR images of Oil Spills and Look Alikes

SAR systems are mainly used for monitoring the sea to detect dark formation in the sea, as they are not affected by local weather conditions and cloudiness and occupy day to night. SAR systems detect dark spots on the sea surface indirectly, through the modification dark formations cause on the low wind areas – capillary waves. Several manmade and natural ocean phenomena affect the backscattering of the radar signals. For this reason, a dark formation appears dark on SAR imagery in contrast to the surrounding clean sea.

Dark formations can be oil spills, organic film, low wind areas, areas sheltered by land, wind front areas, rain cells, grease ice, current shear zones, internal waves and upwelling zones.

Dark formation on SAR images is due to decrease of back scattering of the sea surface due to oil films and other natural phenomena. Oil spills are either intentional or accidental. Look –alikes are mostly natural ones. May be due to weather or due to sea organisms.

Data set used are ERS -2 , RADARSAT and ENVISAT Images.

2.1 Feature Extraction Process

Fiscella et al. (2000) used 14 features for oil spill classification, Solberg and Theophilopoulos (1997) used 15 features. Solberg et al. (1999) used 11 features, many of them different from their previous studies and in general different from the 11 features used by Del Frate et al. (2000). A different approach was given by Espedal and Wahl (1999), in which wind vector data were used and compared with the spreading and length of the dark formations detected. A more general description of the calculated features was given by Espedal and Johannessen (2000), where texture features were introduced for the first time. Moreover, Keramitzoglou et al. (2005) referred to 14 features without presenting them and Karathanassi et al. (2006) used 13 features several studies have tried to unify all the features used with similar characteristics (Brekke and Solberg 2005, Montali et al. 2006).

The absence of systematic research on features extracted and their contribution to the classification results forces researchers to select features arbitrarily as input to their systems. A previous work (Stathakis et al. 2006) was focused on this issue, trying to bridge this gap and to discover the most useful features in oil spill detection. The lack of systematic research can be attributed to the fact that the existing methodologies for searching into a large number of different compilations have not been fully exploited. Genetic algorithms have been successful in discovering an optimal or near-optimal solution amongst a huge number of possible solutions (Goldberg 1989). Moreover, a combination of genetic algorithms and neural networks can prove to be very powerful in classification problems. The methodology of feature selection for oil spill detection is given in Stathakis et al. (2006). The 25 most commonly used features in the scientific community were grouped and their contribution to the final classification was examined. The methodology explores the opportunity of having two unknown parameters in the genetic internal structure, i.e. the number of input features and the number of hidden neurons. The novelty of this approach is the simultaneous evolution of both features and neural network topology. Previously genetic algorithms have been used either to evolve neural network topology (Stathakis and Kanellopoulos 2006) or to select features (Kavzoglou and Mather 2002) but not both at the same time. Thus, a novel synergy of genetic algorithms and neural networks is deployed in order to determine a near-optimal neural network for the classification of oil spills and lookalikes. The present paper evaluates the robustness of the proposed feature combination. In order to further justify the proposed feature combination robustness a comparison with the results of several commonly used reparability indices, including Euclidian, Fisher and Mahalanobis, was performed.

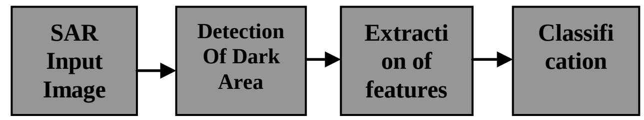
It is very difficult to develop a automatic oil spill detection system as it needs special knowledge to interpret SAR images. Figure 1 shows the steps in detecting the oil spills. The main steps are

1. Preprocess the SAR image to enhance the quality by removing the noise.
2. Detection of Dark areas in SAR Imagery using segmentation process which uses FCM and level Set methods
3. Extraction of Features of Dark Formations
4. Analysis of the feature values by comparing it with oil spill signatures.

In pattern recognition, the *k*-nearest neighbours algorithm (*k*-NN) is a method for classifying objects based on closest training examples in the feature space. *k*-NN is a type of instance based learning, or lazy learning where the function is only approximated locally and all computation is deferred until classification. The *k*-nearest neighbor algorithm is amongst the simplest of all matching learning algorithms: an object is classified by a majority vote of its neighbors, with the object being assigned to the class most common amongst its *k* nearest neighbors. If $k = 1$, then the object is simply assigned to the class of its nearest neighbor.
The same method can be used for regression, by simply assigning the property value for the object to be the average of the values of its *k* nearest neighbors. It can be useful to weight the contributions of the neighbors, so that the nearer neighbors contribute more to the average than the more distant ones. (A common weighting scheme is to give each neighbor a weight of $1/d$, where'd' is the distance to the neighbor. This scheme is a generalization of linear interpolation.)

5. Classification of dark areas (oil spill or lookalike)

Fig(1) Steps In Detection Of Oil Spills In Sar Imagery



Five features of oil spills are extracted for classification which are basically grouped into the 3 categories. They are based on geometric features, physical features and Contextual features

2.2 Classification of Dark Area

After spot detection and feature extraction, dark spots are classified as either oil slicks or look-alikes. This is not an easy task, because slick contrast depends on weather conditions, and the probability of observing look-alikes depends on wind level and other external conditions.

2.2.1 *K Nearest Neighbours (KNN) classification*

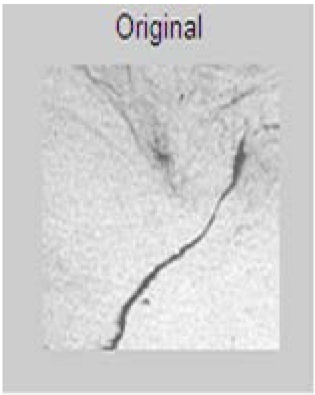
The neighbors are taken from a set of objects for which the correct classification (or, in the case of regression, the value of the property) is known. This can be thought of as the training set for the algorithm, though no explicit training step is required. The *k*-nearest neighbor algorithm is sensitive to the local structure of the data. Using this classification method oil spill classification is done.

3. Experimental Setup and results

The Procedure for classification is described as follows

1. 10 images of each oil spill and look alike are used to form two separate training sets
2. The features like area, perimeter, complexity, shape factor and standard deviation are calculated
3. The values are given as input to the KNN classifier, then the dark area is classified either as oil spill or lookalike based on features.

Sno	Area	Perimeter	Complexity	Shape factor	*Std deviation*	classification *
Image1	1491	233	0.02	6.04	-1.22	1
Image2	1484	366	0.02	9.16	-3.36	1
Image3	1305	395	0.02	9.68	-0.93	1
Image4	1927	395	0.03	10.72	1.48	1
Image5	1040	342	0.03	8.38	0.97	1
Image6	538	243	0.03	10.40	0.91	1
Image7	2382	634	0.03	30.50	-2.42	1
Image8	1625	595	0.03	5.90	1.09	1
Image9	2196	732	0.04	6.71	1.00	1
Image10	0773	778	0.04	6.95	1.68	1
Image11	5159	582	2.00	3.11	-1.75	2 #
Image12	12823	456	1.13	1.14	-2.00	2
Image13	28074	951	1.6	2.02	0.50	2
Image14	12778	513	1.11	2.70	0.72	2
Image15	25857	692	1.13	1.13	.55	2
Image16	14502	614	1.14	2.35	0.63	2
Image17	34319	1196	1.94	1.22	-0.43	2
Image18	32763	824	1.28	1.83	-0.83	2
Image19	21240	796	1.54	1.84	-0.81	2
Image20	11562	666	1.77	2.39	-0.47	2



(a) Original

(b) segmented

4. Conclusions

SAR is the most promising sensor for oil spill detection, as they are not affected by weather conditions. A hybrid method is applied to enhance the image quality. A fusion technique of GMAC with FCM with thresholding is used to detect the dark spots. The features of dark area are extracted. Then they are classified either as oil spill or lookalike. To improve the accuracies more number of features can be considered. The comparison of different classification methods is desirable with same set of data and features.

References

[1]. Solberg, A & .Brekke, C. (2007). Oil spill detection in Radarsat and Envisat SAR images, *IEEE Trans. on Geos. and Rem. Sens.* **45**(3), 7467-55.

[2]. Stathakis, D. Topouzelis, K. & Karathanassi, V. (2006). Large-scale feature selection using evolved neural networks. *13th SPIE Int. Symposium on Remote Sensing*, 11-14 September 2006.

[3]. Karathanassi, V., Topouzelis, K., Pavlakis, P. & Rokos, D. (2006). An object-oriented methodology to detect oil spills. *Int. J. of Rem. Sen.* **27**(23), 5235–5251.

[4]. Keramitsoglou, I., Cartalis, C. & Kiranoudis, C. (2006). Automatic identification of oil spills on satellite images, *Env. Mod. Soft.* **21**(5), 640–652.

[5]. Nirchio, F.,. Sorgente, M, Giancaspro, A., Biamino, W., Parisato, E.,. Ravera, R & Trivero, P. (2005) Automatic detection of oil spills from SAR images, *Int. J. of Rem. Sen.* **26**(6), 1157-1174.

[6]. H. A. Hovland, J. A. Johannessen, and G. Digranes, "Slick detection in SAR images," in *Proc. IEEE Symp. Geosci. Remote Sensing (IGARSS)*,Pasadena, CA, Aug. 1994, pp. 2038–2040.

[7]. A. Solberg and R. Solberg, "A large-scale evaluation of features for automatic detection of oil spills in ERS SAR images," in *Proc. IEEE Symp. Geosci. Remote Sensing (IGARSS),* Lincoln, NE, May 1996, pp. 1484–1486.

[8]. Topouzelis, K., Karathanassi, V. P. Pavlakis and D. Rokos Oil spill detection using rbf neural networks and sar data _____

[9]. Topouzelis, K. "Oil spill detection by SAR images: Dark Formation detection, Feature Extraction and classification algorithms" , Sensors 2008,8,6642-6659; DOI: 10.3390/s8106642

AUTHORS

Mrs V.Radhika received her B.Sc (Phy) ,MCA and M.Phil degree from Bharathiar University in 1993 ,1996 and 2003 respectively. Currently she is pursuing her PhD at Avinashilingam University for Women, Coimbatore. She is the Prof & Head of the Department of Computer Science, Sri Krishna Arts And Science College , Coimbatore. She has 14 years of teaching experience. Her areas of interest include Non-linear filtering techniques in Digital image processing and object oriented technologies. She has more than 10 national and international level publications.

Dr. Padmavathi Ganapathi is the Professor and Head of the Department of Computer Science, Avinashilingam University for Women, Coimbatore. She has 21 years of teaching experience and one year Industrial experience. Her areas of interest include Network security and Cryptography and real time communication. She has more than 150 publications at national and International level. She is a life member of many professional organizations like CSI, ISTE, AACE, WSEAS, ISCA, and UWA. She is currently the Principal Investigator of 5 major projects under UGC and DRDO.

A framework for dynamic indexing from hidden web

Hasan Mahmud[1], Moumie Soulemane[2], Mohammad Rafiuzzaman[3]

[1] Department of Computer Science and Information Technology, Islamic University of Technology, Board Bazar, Gazipur-1704, Bangladesh.

[2] Department of Computer Science and Information Technology, Islamic University of Technology, Board Bazar, Gazipur-1704, Bangladesh.

[3] Department of Computer Science and Information Technology, Islamic University of Technology, Board Bazar, Gazipur-1704, Bangladesh.

Abstract

The proliferation of dynamic websites operating on databases requires generating web pages on-the-fly which is too sophisticated for most of the search engines to index. In an attempt to crawl the contents of dynamic web pages, we've tried to come up with a simple approach to index these huge amounts of dynamic contents hidden behind the search forms. Our key contribution in this paper is the design and implementation of a simple framework to index the dynamic web pages and the use of Hadoop MapReduce framework to update and maintain the index. In our approach, from an initial URL, our crawler downloads both the static and dynamic web pages, detects form interfaces, adaptively selects keywords to generate most promising search results, automatically fill-up search form interfaces, submits the dynamic URL and processes the result until some conditions are satisfied.

Keywords: *Dynamic web pages, crawler, hidden web, index, hadoop.*

.

1. Introduction

Web mining is an application of data mining which aims to discover useful information or knowledge from the web hyperlink structure, page content and usage log. Based on the primary kind of data used in the mining process, web mining tasks are categorized into four main types: (1) Web usage mining, (2) Web structure mining, (3) Web user profile mining and (4) Web content mining [7, 14]. In the past few years, there was a rapid expansion of activities in the web content mining area. This is not surprising because of the phenomenal growth of the web contents and significant economic benefits of such mining. Given the enormous size of the web, the indexed web contains at least 13.85 billion pages [9]. Many users today prefer to access web sites through search engines. A number of recent studies have noted that a tremendous amount of content on the web is dynamic. According to [8] Google, the largest search database on the planet, currently has around eight billion web pages which are already indexed. That's a lot of information. But it's nothing compared to what else is out there. Google can only index the visible web, or searchable web which refers to the set of web pages reachable purely by following hypertext links. But the invisible web or deep web [4, 5, 16, 17, 21], "hidden" behind search forms is estimated to be 500 times bigger than the searchable web. However, a little of this tremendous amount of high quality dynamic contents are being crawled or indexed and in particular, most of them are ignored.

In this paper the focus is on the automatic indexing of dynamic web contents which are the part of deep web. It is same as web content mining as we're extracting the words included in web pages. Here we've tried to come up with a simple approach to crawl the textual portion of dynamic contents hidden behind search forms with the following techniques:

- **Dynamic content extraction:** Extraction of structured data, hidden behind the search forms of Web pages, such as search results. Extracting such data allows one to provide services, so search engines will be benefited if we index dynamic contents of the web pages as most of the time their crawlers avoid those pages.

- **Form detection:** Web form with single general input text field is considered. A site like in [13] uses one single generic text box for form submission. Forms with more than one binding inputs will be ignored.

- **Selection of searching keywords***:* Although the Web contains a huge amount of data, not always an optimized search result is generated for a given keyword. Here the method developed for selecting a candidate keyword for submitting a query will try to generate an optimized search result.

- **Detection of duplicate URLs:** Sometimes two different words may generate same URL twice, which will decrease the efficiency if the same URL is crawled again and again. Detection of duplicate URLs and ignoring them is another try-out of this paper work.

- **Automatic processing:** There is an automation process for crawling. That is recognizing suitable forms, generating keyword for searching, putting the word in the search bar and making or updating an index for the search results; all of these operations will be fully automatic without any human interaction.

This research work only encompasses dynamism in content, not dynamism in appearance or user interaction. For example, a page with static content, but containing client-side scripts and DHTML tags that dynamically modify the appearance and visibility of objects on the page, does not satisfy our definition as well as our objective.

Section 2 of this paper contains some aspects of dynamic web pages including the existing techniques for dynamic web page indexing. Then the proposed approach is presented in Section 3, the framework in Section 4 and its delimitations are discussed in Section 5. The conclusion in Section 6 includes some directions for future work.

2. Defining some aspects of dynamic web pages

Considering the tremendous amount of high quality content “hidden” behind search forms and stored in large searchable electronic database, “Dynamic Web Page” has become a buzzword of current web mining technology. But before continuing our research on dynamic web pages, we’ve to know what a Dynamic Web Page is. A Dynamic Web Page is a template that displays specific information in response to queries. Its *'dynamism'* lies in its resonance and interactivity, both in client-side scripting and server-side scripting. Dynamic web pages can change every time they are loaded (without anyone having to make those changes) and they can change their content based on what user does, like clicking on some text or an image.

Visitors find information in a dynamic site by using a search query. That query can either be typed into a search form by the visitor or already be coded into a link on the home page - making the link a pre-defined search of the site's catalog. In that later case, the portion of the link containing the search parameters is called a 'Query String' .This query is then used to retrieve information from the huge database which is hidden behind the search forms. This whole operation is depicted in Fig.1 below.

2.1 Problems with dynamic pages

All dynamic pages can be identified by the “?” symbol in the URLs, such as

http://www.mysite.com/products.php?id=1&style=a

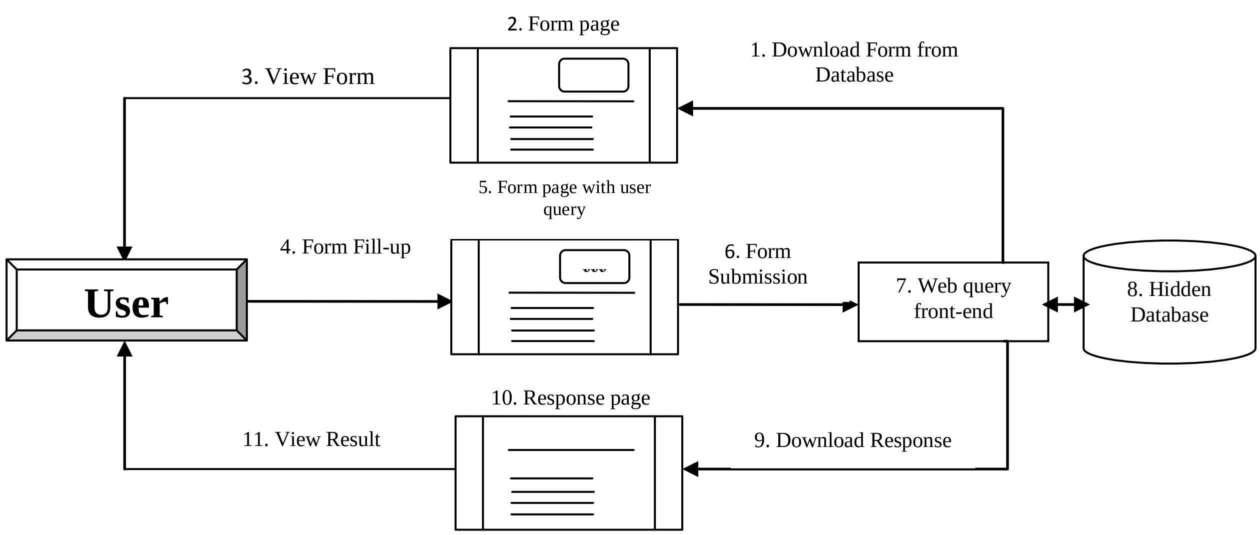


Fig.1: User interactions with search forms to retrieve information from hidden database

Search engines like Google can crawl and index dynamic pages which don't have more than 2 parameters in the URL (the example above has two parameters separated by the "&" symbol). Even so, Google may not crawl dynamic pages for various reasons. Among them, the most common are:

- **Trap:** A spider trap happens when a search engine's web crawler becomes snared in an infinite circle of a website's coding, which produce endless numbers of documents; web pages that are heavily flooded with characters, which may crash spiders programs**.**

- **Insufficient query information:** Search engine spiders have a much tougher time with dynamic sites. Some get stuck because they can't supply the information the site needs to generate the page.

- **Security and privacy issues:** Many of these sites require user authentication and bypassing it automatically may cause violation of privacy or law.

- **Policy preferences:** Some search engines deliberately avoid extensive indexing of these sites.

- **Costly:** It needs expertise in contradiction to the static one that is simple and straight forward.

Moreover, Google will not follow links that contain session IDs embedded in them as in [1].

2.2 Crawling dynamic pages

A web crawler is a relatively simple automated program, or script that methodically scans or "crawls" through Internet pages to create an index of the data it's looking for. Alternative names for a web crawler include web spider, web robot, bot, crawler, and automatic indexer.

When a search engine's web crawler visits a web page, it "reads" the visible text, the hyperlinks and the content of the various tags used in the site, such as keyword rich Meta tags. Using the information gathered from the crawler, a search engine will then determine what the site is about and index the information. In general, it starts with a list of URLs to visit, called the *seeds*. As the crawler visits these URLs, it identifies all the hyperlinks in the page and adds them to the list of URLs to visit, called the crawl-frontier. URLs from the frontier are recursively visited according to a set of policies. Web crawling is done on all text contained inside the hypertext content, tags, or text. The operation of a hidden web crawler is shown in Fig.2.

In practice common crawler algorithms must be extended to address the following issues like in [11]:

- **Speed:** In real life if some HTTP request takes one second to complete, some will take much longer or fail to respond at all. Normally a simple crawler can fetch no more than 86,400 pages per day. At this rate, it would take 634 years to crawl 20 billion pages in a single computation manner. That's why in practice, crawling is carried out using hundreds of distributed crawling machines. Now-a-days Hadoop MapReduce is used over these distributed systems to overcome the bottleneck of a single server computation with higher data processing speed.

- **Politeness:** Unless care is taken, crawler parallelism introduces the risk that a single Web server will be bombarded with requests to such an extent that it becomes overloaded. That's why crawler algorithms are designed to ensure that only one request to a server is made at a time. In order to serve this need a politeness delay is inserted between requests.

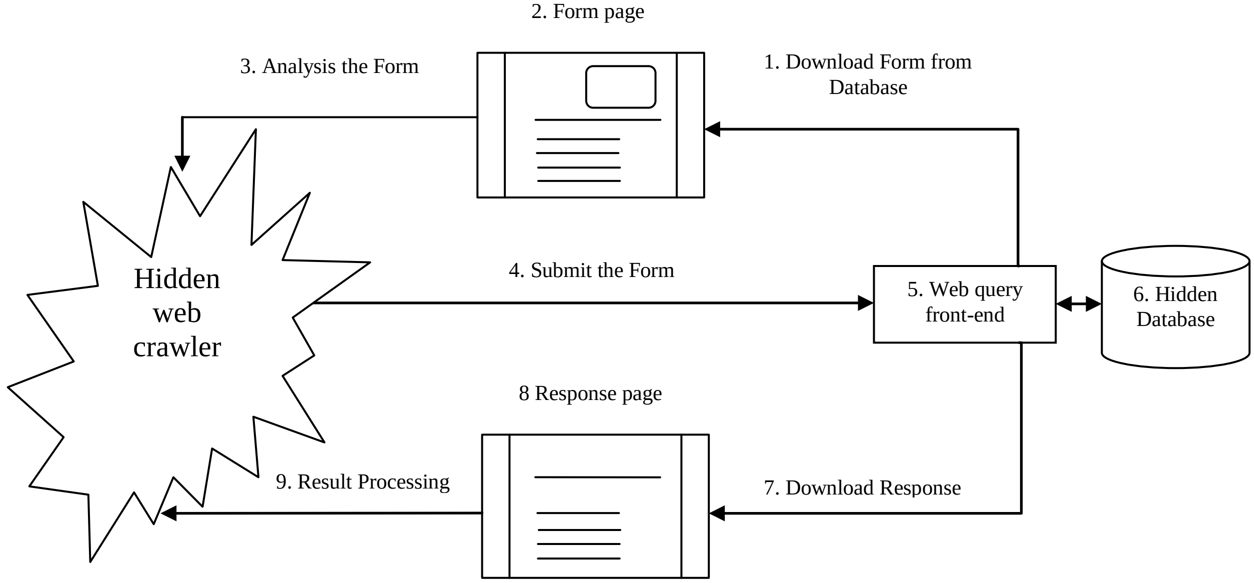


Fig.2: Hidden web crawler

- **Excluded content:** Before fetching a page from a site, a crawler must fetch that site's robots.txt file to determine whether the webmaster has specified to crawl some or the entire site.

- **Duplicate content:** Identical content is frequently published at multiple URLs. But when the page includes its own URL, a visitor counter or a date; more sophisticated fingerprinting methods are needed. Crawlers can save considerable resources by recognizing and eliminating duplication as early as possible because unrecognized duplicates can contain relative links to whole families of other duplicate content.

- **Continuous crawling:** Carrying out a full crawling with fixed intervals would imply slow response to important changes in the Web. For example, submitting the query "current time New York" to the GYM (Google, Yahoo, Microsoft) engines reveals that each of these engines crawls the ***www.timeanddate.com/worldclock*** site every couple of days. However, no matter how often the search engine crawls this site, the search result will always show the wrong time. That's why continuous crawling without any certainty or limit is avoided in most of the current search engines

2.3 Indexing dynamic pages

Indexes are data structures permitting rapid identification of which crawled pages contain like particular words or phrases. To index a set of web documents with the words they contain, we need to have all documents available for processing in a local repository. Creating the index by accessing the documents directly on the Web is impractical for a number of reasons. Collecting "all" web documents can be done by browsing the Web systematically and exhaustively and storing all visited pages. This is done by a crawler and is used by search engines. A similar operation is depicted in Fig.3.

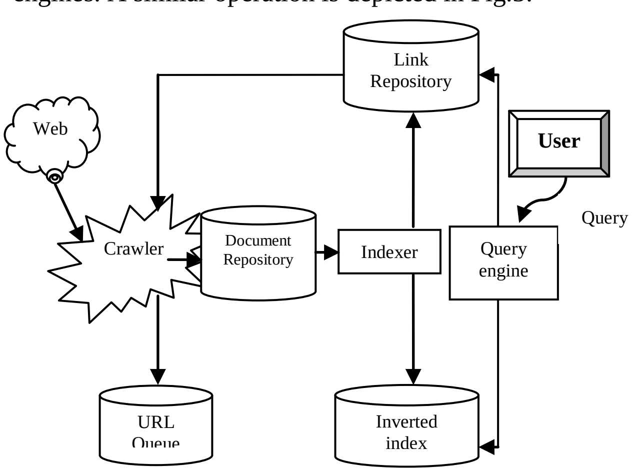


Fig.3: Crawler used in search engines

The purpose of storing an index is to optimize speed and performance in finding relevant documents for a search query. Without an index, the search engine would scan every document in the corpus, which would require considerable time and computing power. For example, while an index of 10,000 documents can be queried within milliseconds, a sequential scan of every word in 10,000 large documents could take hours. The additional computer storage required to store the index, as well as the considerable increase in the time required for an update to take place, are traded off for the time saved during information retrieval. Search engine architectures vary in the way indexing is performed and in methods of index storage to meet the various design factors. Types of indices data structures include:

A. Inverted indices
B. Forward indices

A. Inverted indices

Many search engines incorporate an inverted index when evaluating a search query to quickly locate documents containing the words in a query and then rank these documents by relevance. Because the inverted index stores a list of the documents containing each word, the search engine can use direct access to find the documents associated with each word in the query in order to retrieve the matching documents quickly. TABLE 1 is a simplified illustration of an inverted index.

B. Forward Index

The forward index stores a list of words for each document. TABLE 2 is a simplified form of the forward index.

TABLE 1: Inverted index of words w.r.t. their URLs

Inverted Index Word	***URL***
the	*URL 1, URL 3, URL 4, URL 5*
cow	*URL 2, URL 3, URL 4*
says	*URL 5*
moo	*URL 7*

Table 2: Forward index of words contained in URLs

Forward Index URL	***Words***
URL 1	*the, cow, says, moo*
URL 2	*the, cat, and, the, hat*
URL 3	*the, dish, ran, away, with, the, fork*

2.4 Existing techniques for dynamic web page indexing

One of the deep web crawler architecture is proposed in [2] where a task-specific, human-assisted approach is used for crawling the hidden web. There are two basic problems related to deep web search,

- **Firstly** the volume of the hidden web is very large and
- **Secondly** there is a need of such type of crawlers which can handle search interfaces efficiently, which are designed mainly for humans.

In this paper a model of task specific human assisted web crawler is designed and realized in HiWE (hidden web exposure). The HiWE prototype built at Stanford which crawl the dynamic pages is designed to automatically process, analyze, and submit forms, using an internal model of forms and form submissions. HiWE uses a layout-based information extraction technique to process and extract useful information. The advantages of HiWE architecture is that its application/task specific approach allows the crawler to concentrate on relevant pages only and with the human assisted approach automatic form filling can be done. Limitations of this architecture are that it is not precise with response to partially filled forms and it is not able to identify and respond to simple dependency between form elements. Recently [5] studied the problem of automating the retrieval of data hidden behind simple search interfaces that accept keyword-based queries but did not focus on the detection of search interfaces.

A technical analysis of some of the important deep web search interface detection techniques is done to find out their relative strengths and limitations with reference to current development in the field of deep web information retrieval technology [3] .We found this analysis crucial for the detection of the search interface and it can be a good starting ground for anyone interested in this field. Reference [4] proposed some ways to select keywords for query such random, generic-frequency and adaptive. Meanwhile other usual approach to dynamic indexing is to remove query strings from dynamic URL's, adding dynamic links to static pages, making dynamic links look like static using mod_rewrite available in web server like apache. We also have some paid inclusion programs ,these programs, are premium services for indexing dynamic sites include those of AltaVista, *Inktomi* and *FAST*, to name a few . Also there are Deep web search tools enhance deep Web searching, including *BrightPlanet*, *Intelliseek's Invisible Web*, *ProFusion*, *Quigo*, *Search.com*, and *Vivisimo*. Irregularities highlighted in the existing techniques have leaded us to the proposal of the following approach.

3. Proposed approach

In order to index the dynamic contents hidden behind the search forms, we've come up with an approach which contains the following steps:

3.1 Web pages collection
3.2 Form interface recognition
3.3 Keyword selection
3.4 Query submission
3.5 Result Processing
3.6 Updating the index

3.1 Web pages collection

This part is essentially a static crawling, given the initial URL, the crawler recursively fetches all pages that are linked by it (don't make it recursive, unless you are using functional languages; just use a queue of URLs to be fetched). Test it on the set of web pages created at the beginning. If a page is linked many times, it must be downloaded once. Static crawling is depicted in Fig.4

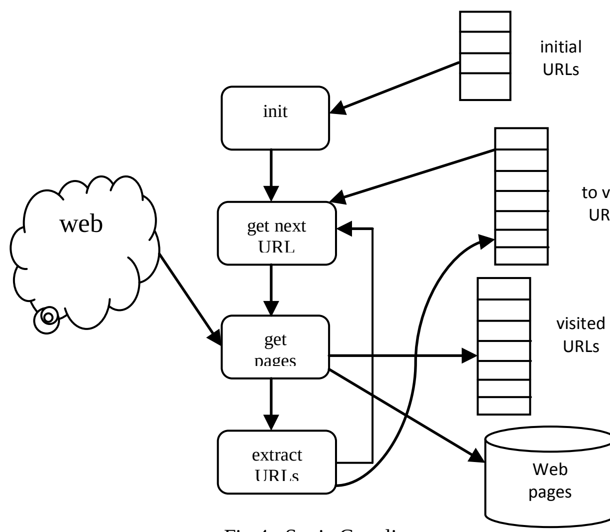


Fig.4: Static Crawling

```
<form action="MAILTO:someone@example.com" method="post"
enctype="text/plain">
Name:<br /><input type="text" name="name" value="your name" /><br />
E-mail:<br /><input type="text" name="mail" value="your email" /><br />
Comment:<br /><input type="text" name="comment" value="your comment"
size="50" />
<br /><br />
<select name="cars">
<option value="volvo">Volvo</option><option value="saab">Saab</option>
<option value="fiat">Fiat</option><option
value="audi">Audi</option></select>
<br /><br /><br /><br /><br /><br />
<input type="submit" value="Send"><input type="reset" value="Reset">
</form>
```

Fig.5: HTML form tag markup for sample Form input controls

3.2 Form interface recognition

Recognizing a web form and its fields is a key point of this approach. A standard HTML web form consists of form tags [6], a start tag <form> and an end tag </form> within which the form fields reside. Forms can have several 'input controls', each defined by an <input> tag and some values considered as domain for those input controls. Input controls can be of a number of types, the prominent ones 'text boxes', 'check boxes', 'selection lists' and 'buttons' (submit, reset, normal or radio).Each of the field is having attributes like label, name and values. The form tag also has attributes like 'method' i.e.: 'get' and 'post' and 'action' which identify the server that will perform the query processing in response to the form submission.

In this study, we focus on the forms with one input control binding to a generic text box. Forms with multiple input controls will be ignored. Fig.6 shows such a form with several input controls and Fig.5 shows the piece of HTML markup that was used to generate this form. Whenever our crawler will encounter forms like this they will be discarded from the crawling operation.

But if the crawler encounters a form tag like depicted in Fig.7 it will consider the form as eligible for crawling and will proceed with its operation. A general search form with single input, often on the top-right of the web page is used in this approach.

Further, as per the HTML specification, forms using ***post*** method for form submission are used whenever submission of the form results in state changes or side effects (e.g. for shopping carts, travel reservation and login). For these reasons we restrict our attention to those forms which are using ***get*** method to submit the form as they tend to produce contents suitable for indexing.

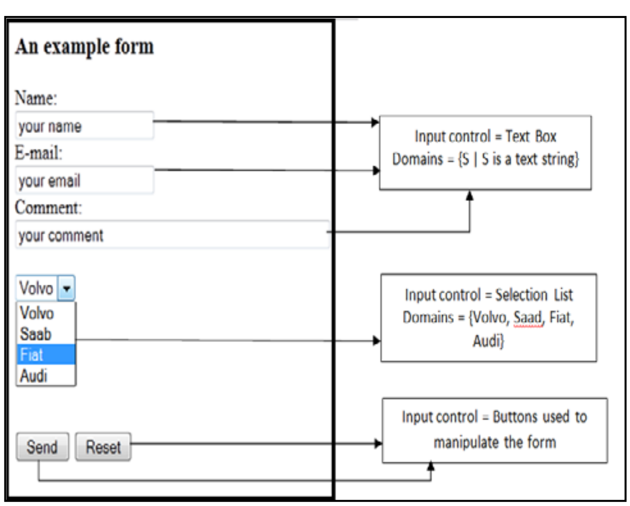


Fig.6: Simple labeled form with several control inputs

```
<form action=http://www.iut.com/department
method="get">
<input name=keyword type=text/>
Input name=searching type=submit value=search/>
</form>
```

Fig.7 HTML form tag markup for a considerable sample form

3.3 Keyword selection

The selection of the most appropriate and relevant value for the input field that can maximize the search is challenging, even though the generic text field generally can accept any keyword. How should a crawler select the queries to issue, given that the goal is to download the maximum number of unique documents from a textual database? Finding the answer for this question is another approach that we've tried to cover here.

In order to solve this question we could select our initial keyword from a dictionary and use it for query submission. But generating a dictionary and searching a keyword within it will be both time and space consuming. Since we want to cover all possible languages, we can't start with from a dictionary of terms. After all an English dictionary will never contain a word which can be used as a keyword for query submission in a Chinese search form. In this case our approach suggests the following aspects:

- **Initial value:** At the very first, keywords are selected from the static content of the web page having the search form interface.
- **Adaptive:** After the generation of the 1st set of results, promising keywords are selected from the successfully retrieved pages. Here keywords for query submission in a search form are selected adaptively from itself.
- **Multilingualism:** By selecting the searching keywords from the web page instead of a predefined location like dictionary or repository our approach also supports multilingualism.

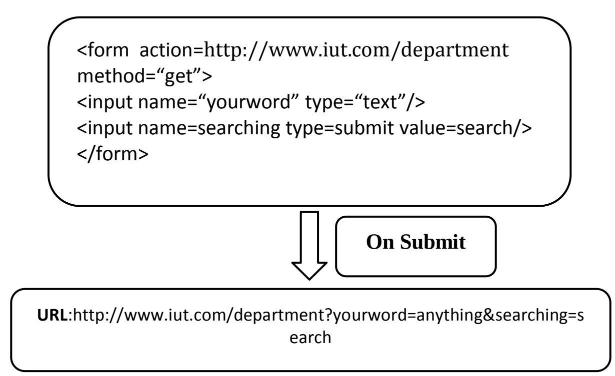


Fig.8: Dynamic URL on querysubmission

- **Limit:** At most **max** submissions per form will take place to prevent the crawler from falling in a trap (infinite loop). Where *max* is a given number representing the maximum number of queries.

The priority in keyword selection is calculated based on the term frequency F_{tf} in our approach as it determines the importance of a word in a set of documents. The term frequency F_{tf} is a measure of how often a word is found in a collection of documents. Suppose a word 'W_i' occurs 'n_p' times within a web page 'P' and there are total of Np words (including the repeated ones) on that page. Then the term frequency,

$$F_{tf} = n_p / Np. \qquad (1)$$

But if we fail to obtain a convenient keyword in that given page, the choice is taken in the repository or at last in the worst case from the dictionary. The selected keywords destined to the query should be compared against the stop words list as these words used to be more frequent.

3.4 Query submission

After selecting the keyword, there is another challenge in submitting the query in the search form automatically, i.e. without any human interaction. Whenever a keyword for a query submission will be selected it'd automatically be submitted in the search form to generate more search results. The action would be something similar depicted in Fig.8.

In this way, whenever a form is submitted a dynamic URL is generated and sent to the database. How many time the query should be submitted and when should it stop? Of course it shall stop when *max* numbers of queries have been submitted.

3.5 Result processing

When our crawler submits a form for processing, different results are possible.

1) The returned page will contain all the data behind the form.

2) The returned page may contain data, but not showing all the data for the query in a single page. Instead, there may be a "next" button leading to another page of data, such as the. In this case, the system will automatically gather all the data on all "next" pages (actually not all, up to a certain limit to avoid a Trap) into a single query result.

3) The query might return data, but only part of the data behind the form because the query is just one of many possible combinations of the form fields. In this case the only returned portion will be processed.

4) The query may return a page that not only contains data, but also contains the original form. Here whatever the result is generated we'll gather information as much as possible.

5) The query may return a page that has another different form to fill in. For this case we'll start with the resultant form from the beginning.

6) Some other error cases might involve a server being down, an unexpected failure of a network connection, or some other HTTP errors.

7) The query may go and return the same form requesting for required field to be filled or to be filled with consistent data. Usually this kind of form contains JavaScript.

8) Successive queries may return redundant result, it is therefore important for similarity detection be verified amount successive queries. After all this, the result should be crawled and indexed.

3.6 Updating the index

After the processing the result an initial index will be created. But as this is a continuous process more and more pages will be crawled to extract more words and will be added to the index in times. As a result a continuous updating of the index is required here which will eventually exceed the capacity of a normal single storage device. That's why multiple storage device is needed and in order to do this we've used "Hadoop-MapReduce" to do the job. Hadoop is an Apache software foundation project as in [20]. It's a scalable, fault-tolerant system for data storage and processing and a framework for running applications on large clusters. Hadoop includes:

- *HDFS* - a distributed file system and
- *Map/Reduce* - offline computing engine.

HDFS splits user data across servers in a cluster. It uses replication to ensure that even multiple node failures will not cause data loss. HDFS breaks incoming files into blocks and stores them redundantly across the cluster. In this approach we're using this splitting and reducing technique to handle the huge amount of index.

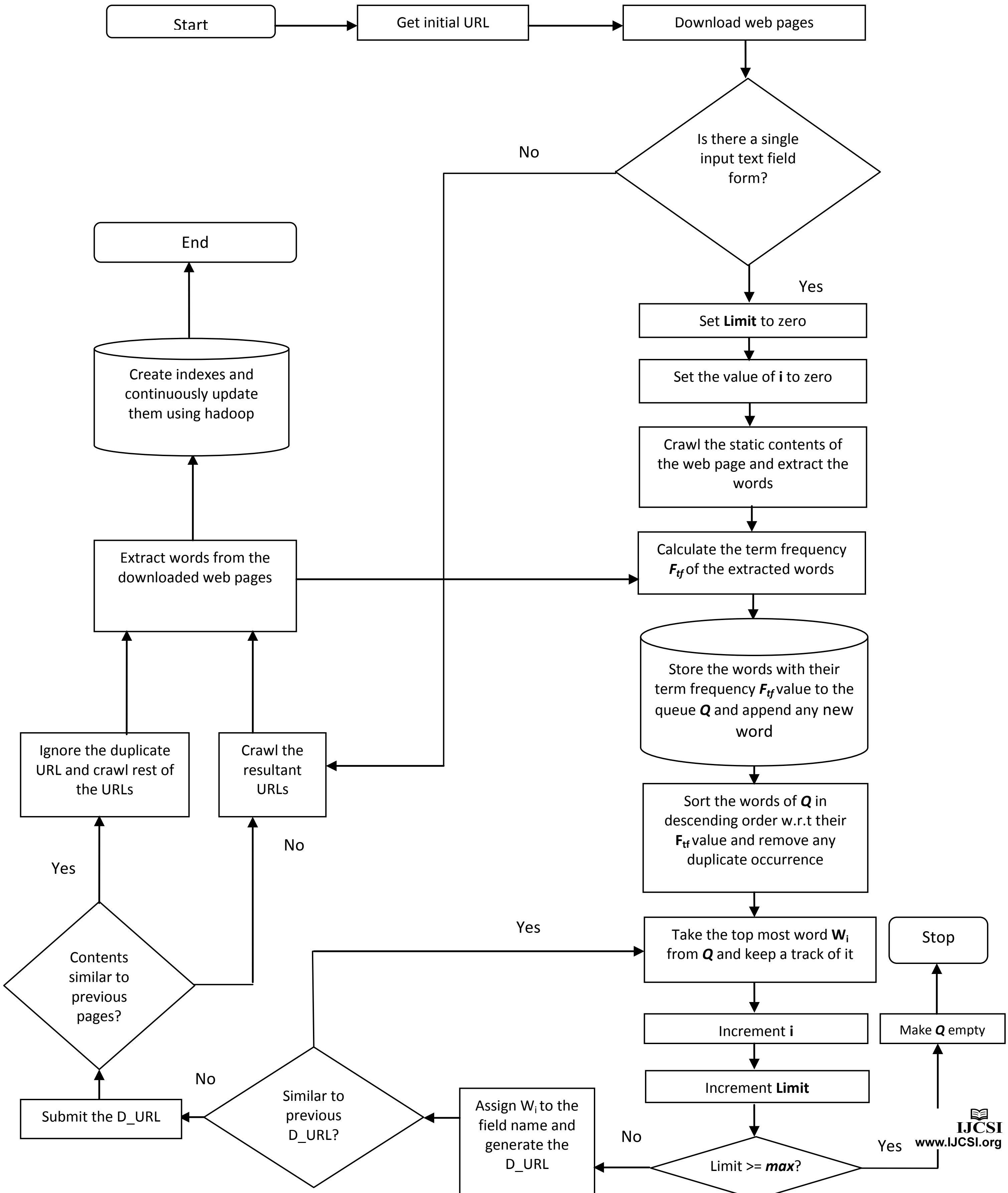

Start
Get initial URL
Download web pages
Is there a single input text field form?
No
Yes
Set Limit to zero
Set the value of i to zero
Crawl the static contents of the web page and extract the words
Calculate the term frequency F_{tf} of the extracted words
Store the words with their term frequency F_{tf} value to the queue Q and append any new word
Sort the words of Q in descending order w.r.t their F_{tf} value and remove any duplicate occurrence
Take the top most word W_i from Q and keep a track of it
Increment i
Increment Limit
Limit >= max?
Yes
No
Make Q empty
Stop
Assign W_i to the field name and generate the D_URL
Similar to previous D_URL?
Yes
No
Submit the D_URL
Contents similar to previous pages?
Yes
No
Ignore the duplicate URL and crawl rest of the URLs
Crawl the resultant URLs
Extract words from the downloaded web pages
Create indexes and continuously update them using hadoop
End

Table 3: legend of the framework

Notations	*Meaning*
Q	*Contains crawled words*
F_{tf}	*Term frequency*
max	*Maximum number of submission=10*
W_i	i^{th} *word*
Limit	*Current number of submission*
i	*Word index*
D_URL	*Dynamic URL*

4. General Framework

According to Fig.9, the web crawler starts with an initial URL to download the page. After that the downloaded web page is crawled and is checked to see if it contains a single input text field for query submission or not. If not, it will simply be crawled and the words and URLs are indexed. If yes, the *limit* and variable 'i' is used respectively to count the number of submissions and searching keywords. After that this web page containing the form is crawled and the term frequency F_{tf} of words extracted from it are calculated and stored in a queue Q. Words in the queue are sorted in descending order of their F_{tf} and the duplicates are removed. The top most word from Q is submitted to the form through a dynamic URL. This process is repeated till the limit reaches the maximum limit *max* when Q is emptied and the crawler stops. Contents of web pages retrieved are detected to see if they are similar to previous ones and the duplicates are deleted, duplicates URLs can be filter as in [18]. These new web pages are then crawled again, sent to the index for updating and term frequency is calculated and submitted as previously until a certain amount of web pages are downloaded. More details can be observed from the Fig.9 and TABLE 3.

5. Delimitations of our approach

In our approach we are not concerned with the following aspects:

1) **Dealing with the form unless it is in the standard format:** If the code is not properly written in a suitable form, our parser will not be able to conveniently extract information from the web page containing that form. Therefore the presence of the form may not be detected.

2) **Handling form that doesn't support passing parameters via URL:** As in [12] the get method append its parameters to the action in the URLs in the form of a dynamic URLs format that are often clearly visible (e.g. http://jobs.com/find?src=bd&s=go). In contrast the post method parameters are sent in the body of the HTTP request and its URL is just simple making it difficult for us to deal with it (e.g., http://jobs.com/find).

3) **Forms with multiple elements:** Because we're focusing in only single input form, any form other than this kind will not be considered for submission.

4) **Forms that span across several pages:** This is the case where the same form is extended over multiple continuous pages.

5) **Forms with JavaScript embedded:** Usually input fields of this type of form have a lot of restriction such the type of input, the format, the length, the syntax. Because we are not going to handle all these, we just prefer to ignore them and discard the form.

6) **Forms that a single input is not a text field:** The single input under consideration must be a text field type

7) **Forms with personal information** indication such as username, password, E-mail will not be considered for privacy raison.

6. Conclusion

In this paper we have studied how to use a hidden web crawler as an approach to dynamic web indexing. We have proposed a complete and automated framework that may be quite effective in practice, leading to an efficient and higher coverage. We have tried to make our design as simple with fewer complexities as possible. Towards the achievement of our goal, we've already developed a prototype for dynamic web page indexing using java, and the website used is [13]. Our future work will include a complete implementation, evaluation and analysis of this approach. We'll also try to compare the performance in both java platform and Hadoop MapReduce.

References

[1] Dan Sisson. *Google SEO secrets, the complete guide*, pp.26–28, 2006.

[2] S. Raghavan, H. Garcia-Molina. Crawling the Hidden Web, in: Proc. of the 27th Int. Conf. on Very Large Databases (VLDB 2001), September 2001.

[3] Dilip Kumar Sharma1, A.k.Sharma2.Analysis of techniques for detection of web search interfaces, *2YMCA University of Science and Technology, Faridabad, Haryana,*

India,http://www.csi-india.org/web/csi/studentskorner-december10, *accessed on June*, 2011.

[4] A.Ntoulas, Petros Zerfos, Junghoo Cho, Downloading Textual Hidden Web Content through Keyword Queries, JCDL '05. Proceedings of the 5th ACM/IEEE-CS Joint Conference, 2005.

[5] Luciano Barbosa, Juliano Freire, siphoning hidden-web data through keyword-based interfaces, Journal of Information and Data management, 2010.

[6] http://www.w3schools.com/html/html_forms.asp, *accessed* on, June 2011

[7] Wiley, *Data Mining the Web Uncovering Patterns.(2007)*.DDU.[0471666556].

[8] Pradeep, Shubha Singh, Abhishek, NewNet- Crawling Deep Web, IJCSNS International Journal of Computer Science and Network Security, VOL.10 No.5, pp. 129-130, May 2010.

[9] http://www.worldwidewebsize.com/, *accessed on June*, 2010.

[10] J Bar-Ilan - Methods for comparing rankings of search engine results-2005,http://www.seo-jerusalem.com/googles-best-kept-secret/,http://www.search-marketing.info/search-algorithm/index.htm, *accessed on June*, 2010.

[11] David Hawking, Web Search Engines-1, pp. 87-88, 2006.

[12] Jayant Madhavan, David Ko, Luc jaKot, Vignesh Ganapathy, Alex Rasmussen, Alon Halevy. “Google's Deep-Web Crawl”, Proceedings of the International Conference on Very Large Databases (VLDB), 2008.

[13] http://www.dmoz.org/, *accessed on June*, 2010.

[14] Brijendra Singh, Hemant Kumar Singh."Web Data Mining Research: A Survey", IEEE, 2010.

[15] http://www.ncbi.nlm.nih.gov/pubmed, *accessed on June*, 2010.

[16] C.H.Chang, M.Kayed, M.R.Girgis, K.F.Shaalan,” A survey of web information extraction systems”. IEEE Transactions on Knowledge and Data Engineering 18(10), pp.1411–1428, 2006.

[17] P.Wu, J.R.Wen, H.Liu, W.Y.Ma,"Query selection techniques for efficient crawling of structured web sources". In: Proc. of ICDE, 2006.

[18] Wang Hui-chang, Ruan,Shu-hua, Tang,Qi-jie."The Implementation of a Web Crawler URL Filter Algorithm Based on Caching”. Second International Workshop on Computer Science and Engineering, IEEE, 2009.

[19] Jeffrey Dean, Sanjay Ghemawat ."MapReduce: Simplified Data Processing on Large Clusters". To appear in OSDI, 2004 http://labs.google.com/papers/mapreduce.html.

[20] http://hadoop.apache.org/, *accessed on june*, 2010.

[21] King-Ip Lin, Hui Chen. "Automatic Information Discovery from the “Invisible Web”", Information Technology: Coding and Computing (ITCC’02), IEEE, 2002.

Hasan Mahmud has received his Bachelor degree in Computer Science and Information Technology (CIT) from Islamic University of Technology (IUT), Bangladesh in 2004. After that he had joined as a faculty member in Computer Science and Engineering (CSE) department at Stamford University Bangladesh.

He did his Master of Science degree in Computer Science (Specialization on NetCentric Informatics) from University of Trento (UniTN), Italy in 2009. He had received University Guild Grant Scholarship for the two years (2007-2009) Master's study and also awarded with early degree scholarship. He has 5 research papers published in different international journals. He is currently working as an Assistant Professor in the department of Computer Science and Information Technology (CIT), IUT, Bangladesh. His current research interests are on web mining, Human Computer Interaction, and Ubiquitous Computing.

Moumie Soulemane did his Higher Diploma in Computer Science and Information Technology (CIT) with specialization in web technology from the Islamic University of Technology (IUT), Bangladesh in 2010. Currently he is the final year student for the B.Sc. degree in Computer Science and Information Technology (CIT) at the same university.

Mohammad Rafiuzzaman is currently in the final year of bachelor degree in Computer Science and Information Technology (CIT) at the Islamic University of Technology.

Energy Efficient Clustering Algorithms in Wireless Sensor Networks: A Survey

Vinay Kumar[1], Sanjeev Jain[2] and Sudarshan Tiwari, IEEE Member[3]

**[1] Dept. of ECE, Motilal Nehru National Institute of technology Allahabad
Allahabad, Uttar Pradesh, India-211004**

**[2] Dept. of ECE, Motilal Nehru National Institute of Technology Allahabad
Allahabad, Uttar Pradesh, India-211004**

**[3] Dept. of ECE, Motilal Nehru National Institute of technology Allahabad
Allahabad, Uttar Pradesh, India-211004**

Abstract

To maximize network lifetime in Wireless Sensor Networks (WSNs) the paths for data transfer are selected in such a way that the total energy consumed along the path is minimized. To support high scalability and better data aggregation, sensor nodes are often grouped into disjoint, non overlapping subsets called clusters. Clusters create hierarchical WSNs which incorporate efficient utilization of limited resources of sensor nodes and thus extends network lifetime. The objective of this paper is to present a state of the art survey on clustering algorithms reported in the literature of WSNs. Our paper presents a taxonomy of energy efficient clustering algorithms in WSNs. And also present timeline and description of LEACH and Its descendant in WSNs.

Keywords: *wireless sensor networks, clustering, energy efficient clustering, LEACH, network lifetime, energy efficient algorithms, energy efficient routing.*

1. Introduction

Energy usage is an important issue in the design of WSNs which typically depends on portable energy sources like batteries for power .WSNs is large scale networks of small embedded devices, each with sensing, computation and communication capabilities. They have been widely discussed in recent years [1-3]. Micro-Electro-Mechanical System (MEMS) sensor technology has facilitated the development of smart sensors, these smart sensors nodes are small devices with limited power, processing and computation resources. Smart sensors are power constrained devices that have one or more sensors, memory unit, processor, power supply and actuator [4]. In WSNs, sensor nodes have constrained in term of processing power, communication bandwidth, and storage space which required very efficient resource utilization. In WSNs the sensor nodes are often grouped into individual disjoint sets called a cluster, clustering is used in WSNs, as it provides network scalability, resource sharing and efficient use of constrained resources that gives network topology stability and energy saving attributes. Clustering schemes offer reduced communication overheads, and efficient resource allocations thus decreasing the overall energy consumption and reducing the interferences among sensor nodes. A large number of clusters will congest the area with small size clusters and a very small number of clusters will exhaust the cluster head with large amount of messages transmitted from cluster members. LEACH protocol is hierarchical routing based on clustering and find the optimal number of clusters in WSNs in order to save energy and enhance network lifetime. In this work, we have surveyed the state-of-art of clustering algorithms in WSNs. We have discussed the advantages and disadvantages of clustering along with a survey of LEACH and its descendant.

Given the importance of clustering for WSNs, rest of the paper is organized in following structure; Section II presents the Challenges and limitations of wireless sensor networks. Section III presents an overview of hierarchical routing in WSNs. Section IV presents a survey on state-of-art of clustering algorithms reported in the literature and section V presents the conclusion of the paper.

2. Challenges and limitation of wireless sensor networks

In WSN sensor nodes have limited processing power, communication bandwidth, and storage space. This gives rise to new and unique challenges in data management and information processing. In-network data processing techniques, such as data aggregation, multicast and broadcast need to be developed. Network lifetime is the key characteristics used for evaluating the performance of any sensor network [6]. A lifetime of the network is determined by residual energy of the system, hence main and most important challenge in WSN is the efficient use of energy resources. Literature shows the energy efficiency is introduced in WSNs using any of the following mechanisms: Energy conservation mechanism, Power conservation mechanism, Energy harvesting mechanism and Energy efficient routing.

2.1 Energy aware routing

The aim of routing in WSNs is to find out and maintain routes in WSNs. Routing challenges with reference to WSNs [19] are Energy consumption without losing accuracy , Node deployment, Link heterogeneity, Data reporting model , Scalability, Network dynamic transmission media, Connectivity, Coverage, Data aggregation, Quality of services.

3. Hierarchical Routing in WSNs

The main target of hierarchical routing or cluster based routing is to efficiently maintain the energy usage of sensor nodes by involving them in multi-hop communication within a particular cluster. Cluster formation is generally based on the energy reserve of sensors and sensors proximity to the Cluster Head (CHs). Clustering plays an important role for energy saving in WSNs. With clustering in WSNs, energy consumption, lifetime of the network and scalability can be improved. Because only cluster head node per cluster is required to perform routing task and the other sensor nodes just forward their data to cluster head. Clustering has important applications in high-density sensor networks, because it is much easier to manage a set of cluster representatives (cluster head) from each cluster than to manage whole sensor nodes. In WSNs the sensor nodes are resource constrained which means they have limited energy, transmit power, memory, and computational capabilities. Energy consumed by the sensor nodes for communicating data from sensor nodes to the base station is the crucial cause of energy depletion in sensor nodes.

3.1 Pros and Cons of clustering in WSNs

The pros of Clustering are that it enables bandwidth reuse thus can improve the system capacity [7]. Due to the fact that within a cluster, all the normal nodes send their data to the CHs so energy saving is achieved by absence of flooding, multiple routes, or routing loops. Due to the fact that clustering enables efficient resource allocation and thus help in better designing of power control and other advantage is due to the fact that any changes of nodes behavior within a cluster affect only that cluster but not the entire network, which will therefore be robust to these changes. There are also several cons of existing clustering schemes in WSNs like in the selection of the cluster heads, some algorithm selects cluster heads only according to the ID number or residual energy of the sensor nodes. Science all the data in sensor network are sent to the base station, the traffic near the base station is higher. The sensor nodes in these areas will therefore run out energy earlier. The base station will then be isolated and as a result, the residual energy stored in the other sensor nodes will be wasted. Another disadvantage is the energy is wasted by flooding in route discovery and duplicated transmission of data by multiple routes from the source to the destination [7].

4. Clustering Algorithms in WSNs

Fig.1 shows the taxonomy of clustering algorithms in WSNs

4.1 CACC: Clustering Algorithm based on Cell Combination [8]

In this paper author proposed a clustering algorithm which based on cell combination for the networks. Sensor nodes are distributed densely and the energy of sensor nodes is always limited. In this clustering algorithm, the monitoring region is divided into hexagonal cells by considering the geographic location information of nodes. Each cluster consists of at least seven hexagon cells. Nodes with the same cluster identity form a cluster and the cluster head in each cluster is elected from the central cell of each cluster. The shape of the cells consider nearly circular in order to improve channel reuse and energy efficiency.

4.2 VAP-E: Energy-Efficient Clustering -Virtual Area Partition [9]

In this authors proposed an energy efficient clustering algorithm which based on virtual area partition in heterogeneous networks environment where the maximal transmission power of each node may be different. Authors found that VAP-E can balance the load between clusters, enhance the energy efficiency of sensor nodes, prolong the lifetime of networks, and improve the efficiency of communications. Authors also compare this algorithm with respect to LEACH and LEACH-E and found that VAP-E can enhance the stability period and network life time with the same simulation condition.

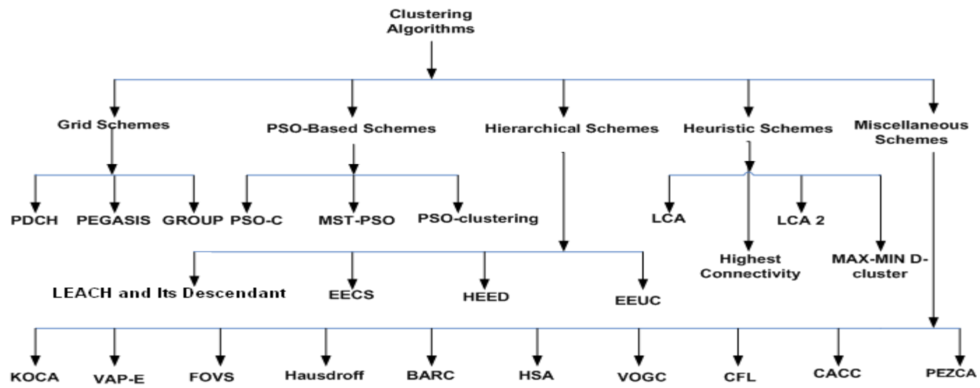


Fig.1: Taxonomy of clustering algorithms in WSNs

4.3 CFL: Clustering for Localization [10]

Authors proposed a clustering algorithm which uses a combined weight function and tries to divide the sensor nodes so that a minimum number of clusters with maximum number of sensor nodes in each cluster could be achieved. The weight functions at each sensor node, which is a combination of different parameters including: residual energy, number of neighbors and transmission power. Basically CFL clustering algorithm is designed for localization in WSNs. It is unable to work when the distribution of sensor nodes are not good.

4.4 FoVs: Overlapped Field of View [11]

Authors proposed a clustering algorithm for wireless multimedia sensor networks based on overlapped Field of View (FoV) areas. The main contribution of this algorithm is finding the intersection polygon and computing the overlapped areas to establish clusters and determine cluster membership. For dense networks, overlapping FoVs causes wasting power of the system because of redundant sensing of the area. The aim of the clustering method is prolonging network lifetime and energy conservation.

4.5 KOCA: K-Hop Overlapping Clustering Algorithm [12]

Authors proposed a clustering algorithm based on K-hop overlapping which is used to overcome the problem of overlapping multi-hop clustering for WSNs. Goal of KOCA algorithm is generating connected overlapping clusters that cover the entire sensor network with a specific average overlapping degree. Authors also found that KOCA produces approximately equal-sized clusters, which allow equally distributing the load evenly over different clusters. In KOCA, clustering formation terminates in a constant time regardless of the network size. Under contention and severe errors, up to 10 percent, KOCA communication overhead is reduced due to the dropped packets. Author's simulation results show that clusters are approximately equal in size. This is requiring achieving load balancing between different clusters.

4.6 PEZCA: Power-Efficient Zoning Clustering Algorithm [13]

Authors proposed a Power-Efficient Zoning Clustering Algorithm (PEZCA) which uses two algorithms: classical LEACH (Low-Energy Adaptive Clustering Hierarchy) and PEGASIS (Power-Efficient Gathering in Sensor Information Systems). In this algorithm, base station consider at a center of the scenario and the scenario area is divided into multiple fan shaped regions and the clusters closer to the base station have smaller sizes than those farther away from the base station. Thus CHs (cluster heads) nearest to the BS (base station) can preserve more energy for inter-cluster data transmission. PEZCA provide more balance in energy consumption and life time of network comparisons to LEACH.

4.7 VoGC: Voting-on-Grid clustering [14]

In this author combined voting method and clustering algorithm, and developed new clustering schemes for secure localization of sensor networks. Authors also found that the newly proposed approaches have very good performances on localization accuracy and the detection rate of malicious beacon signals. In this scheme, malicious beacon signals are filtered out according to the clustering result of intersections of location reference circles. Authors used a voting-on-grid (VOGC) method instead of traditional clustering algorithms to reduce the computational cost and found that the scheme can provide good localization accuracy and identify a high degree of malicious beacon signals.

4.8 BARC: Battery Aware Reliable Clustering [15]

In this clustering algorithm authors used mathematical battery model for implementation in WSNs. With this battery model authors proposed a new Battery Aware

Reliable Clustering (BARC) algorithm for WSNs. It improves the performance over other clustering algorithms by using Z-MAC and it rotates the cluster heads according to battery recovery schemes. A BARC algorithm consists of two stages per round for selection of cluster heads: initialization or setup and steady state. In this formation of cluster, take place by electing a set of CHs. BARC enhances the network lifetime greatly compare to other clustering algorithms.

4.9 Hausdroff Clustering [16]

Authors considered that, once cluster formations take place it's remaining same throughout the network lifetime. This algorithm maximizes the lifetime of each cluster in order to increase the life time of the system. Cluster life time can be enhanced by rotating the role of cluster heads (CHs) among the nodes in the cluster. Cluster heads selection basically based on the residual energy of the sensor nodes and it also used the proximity of neighbors as a secondary criterion for enhancing energy efficiency and further prolong the network lifetime. The Hausdroff clustering algorithm is equally applicable for both uniform and nonuniform sensor node initial energy distribution.

4.10 HSA: Harmony Search Algorithms [17]

This is music based metaherustic optimization algorithm which is analogous with a music improvisation process where musician continue to polish the pitches in order to obtain better harmony. By which it optimizing the energy consumption and minimizing intra-cluster distance of the network. In this the base station computes and allocates nodes into clusters according to the information of their residual energy and location. The operation has two phases: clustering setup and data transmission. This algorithm provides improvement in term of power consumption and network life time over LEACH protocol. With a small network diameter, energy consumption of the network is almost same when using different clustering protocols.

4.11 PEGASIS: Power-Efficient Gathering in Sensor Information System [18]

By this author proposed algorithm PEGASIS that is a chain based protocol provide improvement over LEACH algorithms. In PEGASIS, each node communicates only with a close neighbor and takes turns transmitting to the base station, thus reducing the amount of energy spent per round. Using greedy algorithm, the nodes will be organized to form a chain, after that BS can compute this chain and broadcast it to all the sensor nodes. Energy saving in PEGASIS over LEACH takes place by many stages: First, in the local data gathering, the distances that most of the sensor nodes transmit are much less compared to transmitting to a cluster-head in LEACH. Second, only one node transmits to the BS in each round of communication. PEGASIS outperforms LEACH by limiting the number of transmissions, eliminating the overhead of dynamic.

4.12 Max-Min D-Cluster Algorithm [19]

Authors proposed a clustering algorithm in which no nodes are more than d-hops away from the cluster head. The cluster head selection strategy developed, by having each sensor node initiate a 2d round of flooding, from which results are considered. In order to select the cluster head nodes, follow a set of rule in which 1^{st} d round called flummox, used to propagate largest node IDs and after completion of this round 2^{nd} d round start which is called flagmen. This algorithm is applicable only when two assumptions are made: all nodes that survive the flood max elect themselves cluster heads. During flooding, no node ID will propagate further than d-hops from originating node. This algorithm provides load balancing among the cluster heads.

4.13 PDCH: Pegasis Algorithm Improving Based on Double Cluster Head [20]

Authors proposed an algorithm based on hierarchical chain topology and this algorithm using bottom level cluster head and super level cluster head to improve the load balance. In the hierarchical structure, base station (BS) is the center of a circle. The BS will predefine the number of levels and every node's distance to BS decided the level which it belongs to. Every node receives the signal from the BS, then according to the signal strength to detect the distance to BS. PDCH outperform to PEGASIS algorithm and it is also useful for large networks.

4.14 GROUP [21]

GROUP clustering algorithms based on clustering algorithm that provides scalable and efficient packet routing for large-scale WSNs. Only some parts of total number of sensor nodes participate in formation of cluster heads (CHs). In this, cluster heads are arranged in a grid manner and primary sink (One of the sink), dynamically and randomly builds the cluster grid. Greed Seed (GS) is a node within a given radius from the primary sink. Any queries from sink to nodes are propagated from greed seed to its cluster heads and so on.

4.15 EECS: Energy Efficient Clustering Schemes [22]

Authors proposed an algorithm in which cluster formation is different from LEACH protocol. In LEACH protocol cluster formation takes place on the basis of a minimum distance of nodes to their corresponding cluster head. In

EECS, dynamic sizing of clusters takes place which is based on cluster distance from the base station. The results are an algorithm that addresses the problem that clusters at a greater distance from the sink requires more energy for transmission than those that are closer. Ultimately it provides equal distribution of energy in the networks, resulting in network lifetime. Thus main advantage of this algorithm is the full connectivity can be achieved for a longer duration. So we can say it provides reliable sensing capabilities at a larger range of networks for a longer period of time. It provides a 35 percent improvement in network life time over LEACH algorithm.

4.16 EEUC: Energy Efficient Unequal Clustering [23]

This scheme is distance based scheme similar to EECS and it also required that every node has global identification such as its locations and distances to the base station. Hotspot is the main problem in WSNs because of multi hopping that occurs when CHs closer to the sink tend to die faster compare to another node in the WSNS, because they relay much more traffic than remote nodes. This algorithms partition the all nodes into clusters of unequal size, and clusters closer to the sink have smaller sizes than those farther away from the sink. Thus cluster heads (CHs) closer to the sink can conserved some energy for the inter-cluster data forwarding. Energy consumed by cluster heads per round in EEUC much lower than that of LEACH standard but similar to HEED protocol.

4.17 LCA: Linked Cluster Algorithms and LCA2 [24]

A link cluster algorithm which was one of the oldest clustering algorithms developed for wired sensor networks, but later developed also for wireless sensor networks. In LCA each node has a unique ID number and selection of cluster heads in this algorithm depends upon two factors: Node has the highest ID number in the cluster. If none of its neighbors' are cluster heads. Since LCA used TDMA frame for communication between the nodes, where each frame has slots for each network in the network to communicate. This means that LCA is only applicable for small networks and for larger network LCA impose greater communication delay. Authors proposed LCA2 algorithm, in order to eliminate the election of an unnecessary number of cluster heads, as in LCA.

4.18 Highest-connectivity cluster algorithm [25]

Authors, propose a highest-connectivity cluster algorithm which is similar to LCA. In this algorithm instead of using the ID number for selection of cluster heads authors used connectivity by node. In this the node which is connected more number of nodes is elected as a cluster head. Highest-connectivity cluster algorithm suffers from additional overhead associated with more frequent topology changes. Highest-connectivity cluster algorithm has a slightly larger cluster size than d-hop Max-Min clustering algorithms.

4.19 PSO-Clustering [26]

Authors proposed PSO-clustering which have four variants of PSO: PSO-TVIW (PSO with time varying inertia weight), PSO-TVAC (PSO with time varying acceleration constants), HPSO-TVAC (hierarchical PSO-TVAC) and PSO-SSM (PSO with supervisor student mode) for energy aware clustering in WSNs. This algorithm is applicable only when each node has fixed Omni-directional transmission range, the sensor field should be mapped into a 2-Dimensional space and nodes are randomly distributed. After deployment of the nodes, the nodes are static and the positions of the nodes are known to the base station. The base station runs the clustering algorithm and updates nodes about their cluster-head and all nodes should have same transmission ranges and hardware configurations.

4.20 PSO-C: Centralized-PSO [27]

Authors proposed centralized-PSO algorithms, in which the nodes which have energy above average energy resource are elected as the cluster heads. In this authors also compare this algorithm with LEACH protocol and with LEACH-C. Simulation results show that PSO outperform to LEACH and LEACH-C in term of network life time and throughput etc. It also outperforms GA and K-means based clustering algorithms.

4.21 MST-PSO: Minimum Spanning Tree-PSO [28]

Authors proposed a minimum spanning tree-PSO based clustering algorithm of the weighted graph of the WSNs. The optimized route between the nodes and its cluster heads is searched from the entire optimal tree on the basis of energy consumption. Election of cluster head is based on the energy available to nodes and Euclidean distance to its neighbor node in the optimal tree. Others have concluded that network life time does not depend on the base station location or residual energy of the node. Once the topology decided to then network life time becomes almost settled. Author's shows two techniques for improving network life time: reduce the startup energy consumption of the transmitter and receiver, and optimized the network topology.

4.22 LEACH and Its Descendant

Low Energy Adaptive Clustering Hierarchical Protocol (LEACH) uses the following techniques to achieve the design goals: randomized, self-configuring and adaptive cluster formation, Local control for data transfers and low-energy media access control and application specific data processing. LEACH protocol has many rounds and each round has two phases, a setup phase and steady state phase, in set up phase it provides cluster formation in adaptive manner and in the steady state phase transfer of data takes place. LEACH uses a TDMA or a CDMA MAC to reduce inter-cluster and intra cluster collisions. Cluster formation based on many properties such as the number and type of sensors, communication range and geographical location. The energy consumption of the information gathered by the sensors node to reach the sink will depend on the number of cluster heads and radio range of different algorithms, because the energy consumption can be reduced by organizing the sensor nodes in the clusters [38].

LEACH-F: In this author proposed an algorithm in which the number of clusters will be fixed throughout the network lifetime and the cluster heads rotated within its clusters. Steady state phase of LEACH-F is identical to that of LEACH. LEACH-F may or may not be provided energy saving and this protocol does not provide the flexibility to sensor nodes mobility or sensor nodes being removed or added from the sensor networks.

LEACH-C: LEACH cluster formation algorithm has the disadvantages of guarantee about the number of cluster head nodes and its placement. Since the clusters are adaptive, so there is poor clustering set-up during a round will affect overall performance. However, using a central control algorithm to form the clusters may produce better clusters by distributing the cluster head nodes throughout the network.

LEACH-B : Authors proposed decentralized algorithms of cluster formation in which sensor node only knows about own position and position of final receiver and not the position of all sensor nodes. LEACH-B operates in following phases: Cluster head selection algorithm, Cluster formation and data transmission with multiple accesses. Each sensor node chooses its cluster head by evaluating the energy dissipated in the path between final receiver and itself. It provides better energy efficiency than LEACH.

LEACH-ET: In this cluster will change only when one of the following conditions is satisfied: first, Energy consumed by anyone of the cluster head nodes (CHs) reach energy threshold (ET) in one round. Second, every sensor node should know the energy threshold (ET) value. If in initial phase, anyone of the cluster head nodes dies. If any sensor node acts as a cluster head node (CHs) in a certain round, it should have the energy dissipated value and compares the dissipated value with the energy threshold (ET) value.

Energy–LEACH: it provide improvement in selection of cluster heads of LEACH protocol It makes residual energy of the node as the main factor which decides whether these sensor nodes turn into the cluster head or not in the next round. E- LEACH improves the cluster head election procedure that is chosen in LEACH protocol node cannot be chosen to be a cluster head node. This protocol provides longer network life time and energy saving compared to LEACH protocol

TL-LEACH: its works in three phases, cluster-head casing, Cluster setup and data transmission. In this protocol the author improve the LEACH in which some of the cluster heads elected during setup phase in LEACH were chosen again as the level-2 cluster heads (CHs), which have the communication with the base station

MH-LEACH: In this author proposed a protocol which improves communication mode from a single hop to multi hop between cluster head and base station. In LEACH protocol each cluster head directly communicates with sink and the distance between the sink and cluster head does not matter, if the distance is large it will consume more power. So modified form like MH LEACH protocol which adopt an optimal path between the base station and cluster head and multi hop communication takes place among cluster heads.

ACHTH –LEACH: The author has induced ACHTH - LEACH to improve LEACH and rectify its defects. The clusters are set up based on the Greedy k-means algorithm. The cluster heads are elected by considering the residual energy of sensor nodes. And the cluster heads may adopt two-top transmission to reduce the energy spent on sending data to the BS. The simulation results show that ACHTH-LEACH effectively prolong the lifespan of the network by the balanced clustering approach and the two-hop communication to the BS. The performance of ACHTH-LEACH can be further improved if some parameters and threshold values are optimized in and the percent of nodes alive is less than threshold values are optimized.

MELEACH-L: In this paper the authors enumerated energy-efficient multi-channel routing protocol for wireless sensor networks. With the aid of controlling the size of each cluster and separating CHs from backbone nodes, MELEACH-L manages the channel assignment among neighbor clusters and the cooperation among CHs during the data collection. Analysis and simulations clearly show the validity of the two criteria for large-scale WSNs and the energy-efficiency of MELEACH-L.

Table1: Time line of LEACH and Its Descendant

S.N.	LEACH and Its Descendant	Abbreviation	Year of Publication
1	LEACH [49]	Low energy adaptive clustering hierarchy	2002
2	LEACH-C [49]	Centralized - Low energy adaptive clustering hierarchy	2002
3	LEACH-F [50]	Fixed no. Of Cluster- Low energy adaptive clustering hierarchy	2002
4	LEACH-B [51]	Balanced- Low energy adaptive clustering hierarchy	2003
5	LEACH-ET [52]	Energy threshold- Low energy adaptive clustering hierarchy	2006
6	LEACH-E [53]	Energy- Low energy adaptive clustering hierarchy	2007
7	TL-LEACH [54]	Three Layer- Low energy adaptive clustering hierarchy	2007
8	A -s LEACH [31]	Advanced-solar aware- Low energy adaptive clustering hierarchy	2007
9	S- LEACH [55]	Secure- Low energy adaptive clustering hierarchy	2008
10	Trust Based –LEACH [36]	Trust Based- Low energy adaptive clustering hierarchy	2008
11	LEACH-DCHS-CM [37]	Cluster maintenance - Low energy adaptive clustering hierarchy-DCHS	2008
12	TB- LEACH [38]	Time based- Low energy adaptive clustering hierarchy	2008
13	MAT- LEACH [40]	Mobile agent based- Low energy adaptive clustering hierarchy	2008
14	Armor- LEACH [42]	Advance LEACH routing protocol for micro-sensor networks	2008
15	LEACH-Mobile [45]	Low energy adaptive clustering hierarchy-Mobile	2008
16	A-LEACH [41]	Advanced- Low energy adaptive clustering hierarchy	2008
17	ME- LEACH-L [34]	More energy efficient - Low energy adaptive clustering hierarchy-L	2009
18	Re-cluster- LEACH [46]	Re-cluster- LEACH- Low energy adaptive clustering hierarchy	2009
19	LEACH-H [44]	Low energy adaptive clustering hierarchy-H	2009
20	O- LEACH [43]	Optical- Low energy adaptive clustering hierarchy	2009
21	LEACH-TM [39]	Low energy adaptive clustering hierarchy-trust-minimum transmission	2009
22	MS- LEACH [35]	Combination of **multi-hop** and **single hop**- Low energy adaptive clustering hierarchy	2009
23	Hybrid-LEACH [32]	Hybrid- Low energy adaptive clustering hierarchy	2009
24	LEACH-D [29]	Low energy adaptive clustering hierarchy-D	2010
25	P-LEACH [30]	Low energy adaptive clustering hierarchy-partition	2010
26	ACHTH- LEACH [33]	Adaptive cluster head election and two hop transmission- Low energy adaptive clustering hierarchy	2010
27	MR- LEACH [47]	Multi-hop hop routing- Low energy adaptive clustering hierarchy	2010
28	HPR- LEACH [48]	Heterogeneous- Low energy adaptive clustering hierarchy	2010

MS-LEACH: In this paper the authors have analyzed the problem of energy consumption of the single-hop and multi-hop transmissions in a single cluster. Finally a critical value of the cluster area size is determined. MS-Leach is based on the critical value. Simulation results clearly show that MS-Leach outperforms at most by 200% in term of network lifetime. It is proposed as future work its relationship between multi-hop and single-hop transmissions will be analyzed in-depth in various protocols and new mechanisms of routing will be developed.

Trust-Based LEACH: Since the commonly security solution based on cryptography and other traditional methods which cannot incorporate new challenges from internal attackers, and trust is recognized as a novel approach to defend against such attacks. In this paper the authors have proposed a trust-based LEACH protocol to provide secure routing, while keeping the necessary functionalities of the original protocol. The decision-making is based on the decision trust, evaluated independently and adaptively for different decisions by basic situational trust.

LEACH-DCHS-CM: The authors have presented a LEACH-DCHS-CM algorithm against the characteristic of the frequent formation of the clusters in LEACH-DCHS algorithm. Highlighted the option of using energy balance clustering algorithm when the number of failed nodes up to a certain extent. As a future wok main concentration on the “certain value” settings of nodes also deserve further research.

TB-LEACH: A new protocol of Cluster-Head Selection Algorithm for LEACH based on time (TB-LEACH). Principle of TB-LEACH is stated and the main flowchart and pseudo codes realizing TB-LEACH. Analysis between new protocol and LEACH protocol is done which significantly shows that there is an improvement which is done by formation of a constant number of clusters; TB-LEACH constructs the cluster by using an algorithm based random-timer, which doesn't require any global information. Also the simulation results confirm the TB-LEACH provides the best energy efficiency and the longer network lifetime in comparison to LEACH.

LEACH-TM: In this paper as compared to LEACH which have certain drawbacks a new improved LEACH-TM introduces the concept of Trust, designs the cluster-head adjusting procedure and establishes a multi - path with cluster-heads acting as routers. The simulation confirms that LEACH-TM over comes the deficiency makes in the reliability of data transmission, the distribution of cluster heads and the lifetime of networks.

A-LEACH: LEACH protocol suffers with the problem that Head node spends the more energy in comparison to others. In this paper it is proposed how effectively to process data using a mobile agent technique based LEACH. So that the task of energy saving and reliable data is fulfilled. In this paper, the authors have elaborated how to select the cluster heads in every round which depends both on current state probability and general probability.

Armor-LEACH: This protocol as compared to LEACH, Sec-LEACH the Armor-LEACH provides large sensor networks with high energy saving, and high level of performance, three times better. Simultaneously it produces a higher level of security than it produced by LEACH and TCCA. Final results produced provide a very efficient solution for sensor networks communications.

A-LEACH: In this paper Author has elaborated energy-efficient communication protocol, called optical LEACH (a-LEACH), for hybrid sensor networks that incorporate distributed optical fiber sensor links located at the center and two isolated wireless sensor networks (WSNs) with randomly scattered nodes. Simulation results indicate performance in terms of lifetime is simulated with -30% improvements over LEACH protocol.

LEACH-H: In order to enhance life span of WSN a new protocol LEACH-H is developed which includes pros and cons of both LEACH and LEACH-C. Analysis of LEACH and LEACH-C is working out by authors which are classic WSN cluster routing protocols. It is suitable for applications in the large-scale WSN.

LEACH-Mobile: Protocols such as LEACH suitable for clusters are best suited for routing in wireless sensor networks. LEACH-M which is developed and best suited for mobility centric environments with some modification was suggested in the basic scheme. LEACH-Mobile is one such protocol. The proposed LEACH-M protocol improves in the mobile scenario by ensuring whether a sensor node is able to communicate with its cluster head

Re-Cluster-LEACH: This paper puts forward a new routing protocol on the basis of LEACH protocol: A Re-cluster-LEACH routing protocol based on node density. This protocol improves the cluster head selection algorithm in LEACH, increase cluster-based data fusion and brings forward the algorithm of the second selection of a cluster head. The paper also makes improvements on the most important flaw in LEACH single hop transmission

MR-LEACH: In this paper, it has proposed a multi-hop routing protocol with low energy dynamic cluster hierarchy to minimize the energy consumption of sensor nodes. The performance evaluation section has shown that MR-LEACH performs well compared to similar approaches given that network is divided into an optimal number of layers.

LEACH-HPR: In this paper, LEACH-HPR introduced an energy efficient cluster head election method and using the improved Prim algorithm to construct an inter-cluster routing in the heterogeneous WSN. Simulation results show LEACH-HPR is more efficient to reduce and balance energy consumption and hence enhance the lifetime of WSN

5. Results

We have surveyed the state-of-art of different clustering algorithms in wireless sensor networks along with LEACH and descendant reported in the literature of WSNs till today and presented the comparison of different LEACH descendant. We have found that the some energy efficient algorithms increases the network lifetime Although every effort has been made to provide complete and accurate state of the art survey on energy efficient clustering algorithms along with LEACH and its descendant as applicable to WSNs.

References

[1] S. S. Iyengar, L. Prasad, and H. Min, “Advances in distributed sensor technology,” Englewood Cliffs, NJ: Prentice Hall, 1995.

[2] G. Pottie and W. Kaiser, “Wireless integrated network sensors” ACM Communications, vol. 43, no. 5, pp. 51–58, May 2000.

[3] J. M. Kahn, R. H. Katz, and K. S. J. Pister, “Next century challenges: Mobile networking for smart dust,” in Proceedings of ACM/IEEE Internatonal Conference on Mobile Computing Networks, pp. 271–278, August 1999.

[4] I. F. Akyildiz, S. Weilian, Y. Sankarasubramania and E. Cayirci, “A survey on sensor networks,” IEEE Communications Magazine, vol. 40, no. 8,pp. 102–114, August 2002.

[5] Z. Q. C. Yunxia, “On the lifetime of wireless sensor networks,” Vol. 9, pp. 976–978, 2005.

[6] J. N. Al-Karaki and A. Kamal, “Routing techniques in wireless sensor networks: a survey,” vol. 11, pp. 6–28, Dec. 2004.

[7] C. Hao and S. Megerian, “Cluster sizing and head selection for efficient data aggregation and routing in sensor networks,” In Proceedings of the IEEE on Wireless Communications and Networking, vol. 4, pp. 2318–2323, April 2006.

[8] L. Chang-RI, Z. Yun, Z. Xian-ha, and Z. Zibo "A clustering algorithm based on cell combination for wireless sensor networks" In *Second International Workshop on Education Technology and Computer Science*, 2,74–77.

[9] R. Wang, L. Guozhi, and C. Zheng " A clustering algorithm based on virtual area partition for heterogeneous wireless sensor networks". In *International Conference on Mechatronics and Automation*, 372–376.

[10] S. Zainalie and M. Yaghmaee "CFL: A clustering algorithm for localization in wireless sensor networks". In *International Symposium on Telecommunications* 435–439.

[11] M. Alaei , and J.M. Barcelo-ordinas "Node clustering based on overlapping FOVs for wireless multimedia sensor networks". In *Proceedings of the IEEE wireless communication and networking*, 1–6.

[12] M. Youssef, A. Youssef, and M. Younis, " Overlapping multi-hop clustering for wireless sensor networks" *IEEE transactions on parallel and distributed systems*, 20, 12, 1844–1856.

[13] F. Bai, H. Mu, and J. Sun " Power-efficient zoning clustering algorithm for wireless sensor networks". In Proceedings *of the Information Engineering and computer science*, 1-4.

[14] W. Yang, and W.T. Zhu "Voting-on-grid clustering for secures localization in wireless sensor networks". In *Proceedings of the IEEE International Conference on communication*, 1–5

[15] K. Watfa, O.Mirza, and J. Kawtharani "BARC: A battery aware reliable clustering algorithm for sensor networks". *Journal of Network and Computer Applications*, 32 ,6, 1183–1193.

[16] X. Zhu, L. Shen, and T. Yum "Hausdorff clustering and minimum energy routing for wireless sensor networks". *IEEE transaction on vehicular technology*, 58, 2, 990–997

[17] D. C. Hoang, P. Yadav, R. Kumar, and S. Panda "Node clustering based on overlapping FOVs for wireless multimedia sensor networks". In *Proceedings of the IEEE on Communications Workshops (ICC)*, 1–5.

[18] S. Lindsey,. And C.S. Raghavendra, C.S. "PEGASIS: Power-efficient gathering in sensor information systems". In *IEEE proceeding on aerospace*, 3, 1125-1130.

[19] A. D. Amish, R..prakash, T. vuong, and D. huynh "Max-min d-cluster formation in wireless ad hoc networks". *In Nineteenth IEEE Annual Joint proceeding on Computer and communication societies*, 1, 32–41.

[20] W. Linping, B. Wu, C. Zhen, and W.Zufeng "Improved algorithm of PEGASIS protocol introducing double cluster heads in wireless sensor network". *IEEE International Conference on Computer, Mechatronics, Control and Electronic Engineering*, 148–151.

[21] Y. Liyang, M.W. Neng, Z. Wei, and Z. Chunlei "GROUP: A grid-clustering routing protocol for wireless sensor networks". In *IEEE International conference on Wireless Communications, Networking and Mobile Computing*, 1–5.

[22] Y. Mao, L. Chengfa, C. Guihai, and J. Wu "EECS: An energy efficient clustering scheme in wireless sensor networks". In *Proceedings IPCCC, IEEE 24th International*, 535–540.

[23] C. Li, M. Ye, G. Chen and J. Wu, "An energy-efficient unequal clustering mechanism for wireless sensor networks". In *IEEE International Conference on Mobile Ad-hoc and Sensor Systems*, 604–611.

[24] D.J. Baker and A. Epheremide, "The architectural organization of a mobile radio network via a distributed algorithm". *IEEE transaction on communication* 29,11, 1694–1701.

[25] A.D Amis, R. Prakash, T. Yuong, and D. Huynh, " Max-min d-cluster formation in wireless ad hoc networks". *In Nineteenth IEEE Annual Joint proceeding on Computer and communication societies*, 1, 32–41.

[26] S. Guru, S. Halgamuge and S. Fernando "Particle swarm optimizers for cluster formation in wireless sensor networks". In *IEEE International Conference on sensor*, 319–324.

[27] N.M.A. Latiff, C.C. Tsimenidis, and B.S Sheriff "Energy-aware clustering for wireless sensor networks using particle swarm optimization". In *IEEE International Conference on mobile radio communication*, 1–5.

[28] X. Co, H. Zhang, J. Shi, and G.Cui "Cluster heads election analysis for multi-hop wireless sensor networks based on weighted graph and particle swarm optimization". In *IEEE fourth International Conference on computing*, 7, 599–603.

[29] Y. Liu, Z. Lo, K. Xu and L. chen; "A reliable clustering algorithm base on LEACH protocol in wireless mobile sensor networks" 2nd International Conference on Mechanical and Electrical Technology (ICMET), 2010, pp 692 – 696.

[30] H. Gou and Y. Yoo "An Energy Balancing LEACH Algorithm for Wireless Senso r Networks" In seventh international conference on, Information Technology: New Generations (ITNG), 2010, pp. 822-827.

[31] M. J. Islam, M. M. Islam and M. N. Islam "A-sLEACH: An Advanced Solar Aware Leach Protocol for Ener gy Efficient Routing in Wireless Sensor Networks" Networkin g, Sixth International Conference, 2007, pp.4.

[32] A. Azim, and M. M. Islam "Hybrid LEACH: A relay node based low energy adaptive clu stering hierarchy for wireless sensor networks" Communicatio ns (MICC), 2009 IEEE 9th Malaysia International Conference on , Page(s): 911 – 916,

[33] L. Guo, Y. Xie, C. Yang and Z. Jing; "Improvement on leach by combining adaptive cluster head ele ction and two-hop transmission" Proceedings of the Ninth International Conference on Machine Learning and Cybernetics, Qingdao, July 2010, Vol. 4, pp. 1678 – 1683, 11-14.

[34] J. Chen and H. Shen "MELEACH-L: More Energy-efficient LEACH for Large-scale WSNs" Wireless Communications, Networking and Mobile Computing, 2008. WiCOM '08. pp 1-4,

[35] T. Qiang, W. Bingwen and W.COM Zhicheng "MS-Leach: A Routing Protocol Combining Multi-hop Transmissions and Single-hop Transmissions" Pacific-Asia Conference on Circuits, Communications and Systems, 2009, pp. 107-110,.

[36] Trust based – F. Song, and B. Zhao "Trust-based LEACH Protocol for Wireless Sensor Networks" Second International Conference on Future Generation Communication and Networking, 2008. pp.202-207,

[37] Y. Liu, J. Gao ,Y. Jia and L. Zhu "A Cluster Maintenance Algorithm Based on LEACH-DCHS Protocol" International Conference on Networking, Architecture, and Storage, 2008, pp.165-166.

[38] H. Junping, J. Yuhui and D. Liang "A Time-based Cluster-Head Selection Algorithm for LEACH" IEEE Symposium on Computers and Communications, 2008, pp.1172-1178.

[39] W. Weichao, D. Fei and X. Qijian "An Improvement of LEACH Routing Protocol Based on Trust for Wireless Sensor Networks" 5th International Conference on Wireless

Communications, Networking and Mobile computing, 2009, pp. 1-4 .

[40] MAT-based : J. Hee-Jin, N. Choon-Sung, J. Yi-Seok and S. Dong-Ryeol " A Mobile Agent Based LEACH in Wireless Sensor Networks" 10th International Conference on Advanced Communication Technology, 2008, vol.1, pp.75-78.

[41] M.S. Ali, T. Dey and R. Biswas "ALEACH: Advanced LEACH Routing Protocol for Wireless Microsensor Networks" International Conference on Electrical and Computer Engineering, pp. 909-914.

[42] M. S. Ali, T. Dey, and R. Biswas "ALEACH: Advanced LEACH Routing Protocol for Wireless Microsensor Networks" Proceedings of 17th International Conference on Computer Communications and Networks,2008, pp.1-7

[43] L.-S. Yan, W. Pan, B. Luo, J.-T. Liu and M.-F. Xu Communication Protocol Based on Optical Low Energy-Adaptive-Clustering-Hierarchy (O-LEACH) for Hybrid Optical Wireless Sensor Networks" Communications and Photonics Conference and Exhibition (ACP), 2009 Asia pp. 1-6.

[44] W. Wang, Q. Wang, W. Luo, M. Sheng, W. Wu and Li Hao "Leach-H: An Improved Routing Protocol for Collaborative Sensing Networks" International Conference on Wireless Communications & Signal Processing, 2009, pp.1-5.

[45] G. Santhosh Kumar, V. Paul M V , and K. Poulose Jacob "Mobility Metric based LEACH-Mobile Protocol" 16th International Conference on Advanced Computing and Communications, 2008. ADCOM pp. 248 – 253.

[46] G. Yi, S. Guiling, L. Weixiang and P. Yong "Recluster-LEACH: A recluster control algorithm based on density for wireless sensor network" 2nd International Conference on Power Electronics and Intelligent Transportation System vol.3, pp.198-202, 2009.

[47] M. O. Farooq, A.B. Dogar and G.A. Shah "MR-LEACH: Multi-hop Routing with Low Energy Adaptive Clustering Hierarchy" Fourth International Conference on Sensor Technologies and Applications (SENSORCOMM), 2010, pop. 262 – 268.

[48] L. Han "LEACH-HPR: An energy efficient routing algorithm for Heterogeneous WSN" IEEE International Conference on Intelligent Computing and Intelligent Systems (ICIS), 2010 , vol.2, pp.507-511.

[49] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," IEEE Transaction on Wireless Communications, 2002, vol. 1, no. 4, pp. 660–670.

[50] W. B. Heinzelman. "Application-Specific Protocol Architectures for Wireless Networks". PhD thesis, Massachusetts Institute of Technology, June 2000.

[51] A. Depedri, A. Zanella and R. Verdone, "An Energy Efficient Protocol for Wireless Sensor Networks" In Proc. AINS, 2003, pp. 1-6.

[52] L. Lijun, W. Hunt, and C. Peng, "Discuss in a round rotation policy of hierarchical route in wireless sensor networks," in proceedings IEEE International Conference WiCOM, 2006, pp. 1–5.

[53] X. Fan and Y. Song, "Improvement on leach protocol of wireless sensor network," in In Proceedings of the International Conference on Sensor Technologies and Applications, 2007, pp. 260–264.

[54] D.Zhixiang and Q.Bensheng, "Three-layered routing protocol for wsn based on leach algorithm," in IET Conference on Wireless, Mobile and Sensor Networks,2007, pp. 72–75.

[55] W. Xiao-yun, Y. Li-zhen1, and C. Ke-fei1, "SLEACH: Secure low-energy adaptive clustering hierarchy protocol for wireless sensor networks," 205, vol. 10, pp. 127–131.

Vinay Kumar received the Bachelor's Degree in Electronics and Communication Engineering. From Uttar Pradesh Technical University Lucknow, India in 2006 and the Master's Degree in Digital system from MNNIT (Motilal Nehru National Institute of Technology) Allahabad, India in 2010 and currently pursuing Ph.D. from MNNIT, Allahabad. His fields of interest are WSNs, MANETs, and WMNs. He is author of more than 5 conference papers

Sanjeev Jain received the Bachelor's Degree in Electronics and Communication Engineering from M.B.M Engineering college, Jodhpur, India subsequently M.E. from Malviaya National Institute of Technology, Jaipur, India. Currently pursuing Ph.D., From M.N.N.I.T, Allahabad, India. Simultaneously he is holding the post of Associate Professor at Government Engineering College Bikaner affiliated to Rajasthan Technical University, Kota, India. His areas of interest are WSNs, WMNs.

Sudarshan Tiwari received Ph.D degree in Electronics and Computer Engineering From IIT Roorkee in 1994 and currently working as a Professor of Electronics and communication Engineering Department of MNNIT Allhabad, India. Simultaneously he is holding the post of Dean resource generation and international affairs. He is the author or coauthor of more than 110 technical publications including journal and proceeding papers. His research interests include Communication Engineering & Networking, Wireless communication and networks, Wireless sensor networks, Network Coding, WDM optical Networks and Wireless mesh networks.

Genetic programming approach on evaporation losses and its effect on climate change for Vaipar Basin

K.S.Kasiviswanathan[1,*], R.Soundhara Raja Pandian [2], S.Saravanan[3], Avinash Agarwal [4]

[1] WRDM, IIT Roorkee,
India

[2] IC&SR, IIT Madras,
India

[3] Civil Engineering, NIT Trichy
India

[4] Surface water division, NIH Roorkee
India

Abstract

Climate change is the major problem that every human being is facing over the world. The rise in fossil fuel usage increases the emission of `greenhouse' gases, particularly carbon dioxide continuously into the earth's atmosphere. This causes a rise in the amount of heat from the sun withheld in the earth's atmosphere that would normally radiated back into space. This increase in heat has led to the greenhouse effect, resulting in climate change and rise in temperature along with other climatological parameters directly affects evaporation losses. Accurate modelling and forecasting of these evaporation losses are important for preventing further effects due to climate change. Evaporation is purely non-linear and varying both spatially and temporally. This needs suitable data driven approach to model and should have the ability to take care of all these non-linear behaviour of the system. As such, though there are many empirical and analytical models suggested in the literature for the estimation of evaporation losses, such models should be used with care and caution. Further, difficulties arise in obtaining all the climatological data used in a given analytical or empirical model. Genetic programming (GP) is one such technique applied where the non-linearity exist. GP has the flexible mathematical structure which is capable of identifying the non-linear relationship between input and output data sets. Thus, it is easy to construct 'local' models for estimating evaporation losses. The performance of GP model is compared with Thornthwaite method, and results from the study indicate that the GP model performed better than the Thornthwaite method. Forecasting of meteorological parameters such as temperature, relative humidity and wind velocity has been performed using Markovian chain series analysis subsequently it is used to estimate the future evaporation losses using developed GP model. Finally the effect of possible future climate change on evaporation losses in Pilavakkal reservoir scheme, India has been discussed.

Keywords: *Climate change, genetic programming, green house effect.*

1. Introduction

Climate change is a major issue in this century and it affects directly as well as indirectly the water resources. Increasing amount of CO_2 emission into the atmosphere will increase the global temperature known as global warming [8]. Also it [8] suggests that the water resources system will negatively respond on water availability and water supply. The understanding of the impact of climate change on evaporation could be better explained by accurate modelling and forecasting of the variable.

Evaporation is one of the least satisfactorily explained components of the hydrologic cycle because, unlike say stream flow, it cannot be directly measured. The simplest approaches are mass transfer, energy budget methods to estimate the evaporation. Further more, The National Weather Service class "A" pan is the most widely used evaporation instrument today and its application in hydrologic design and operation is long standing. Pan evaporation has been widely used for estimating lake and reservoir evaporation. These methods are generic and may affect the modelling approach and needs suitable data driven approach.

Numerous researchers have attempted to estimate the evaporation values from climatic variables, and most of these methods require data that are not easily available. Important historical examples include: (a) empirical relationships between meteorological variables ([3], [17], and [19]) and (b) physically-based equations ([15], [16]). While the former methods estimate evaporation based on climate data, the latter methods link evaporation dynamics with the supply of energy and the aerodynamics transport characteristics of a natural surface. Simple methods that are reported (e.g. [14], [18], and [19]) try to fit a linear relationship between the variables. However, the evapotranspiration process is embedded with large variability in both spatial and temporal scales [10].

Black box models like Genetic Programming (GPs), artificial neural networks (ANNs) and fuzzy techniques have been proposed as efficient tools for modeling and their application in water resources problems are reported successfully ([1], [2], [4], [9], [10], [12], [13], [20], [23]). GP has the advantage of providing inherent functional input-output relationships as compared to traditional black box models, which can offer some possible interpretations to the underlying process.Genetic programming (GP) is chosen as the modelling tool owing to its ability to evolve mathematical models from historical data.

In this study, the major objective was to evaluate the potential of GPs for estimating evaporation for Kovilar Reservoir in the Vaipar basin of Tamilnadu, India and its results are compared with Thornthwaite method [21]. Forecasting of the meteorological parameter such as temperature, relative humidity and wind velocity is carried out using Markovian chain series analysis method. Using the forecasted series, the future evaporation loss is estimated with the help of developed GP model. Finally climate change effects were discussed with temperature and evaporation changes.

2. Study area and data used

Pilavakkal reservoir system consists of Periyar and Kovilar reservoirs in Virudhunagar District of Tamilnadu State, India. For this study, Kovilar reservoir has been taken as a study area (Figure 1). It is situated within the geographical coordinates of (9^{o}41'N, 77^{o}23'E) and (9^{o}38'N, 77^{o}32E). Kovilar reservoir is constructed across the non-perennial Kovilar River, which carry only intermittent flash flows depending on the seasonal rainfall. The command area experiences a tropical climate throughout the year. A maximum mean monthly temperature of 38.34^{o}C is observed during May, whereas a minimum mean monthly temperature of 20.04^{o}C occurred in the month of January. Mean annual rainfall of the dam site is 1187 mm against the state average of 945mm. However, the rainfall is highly erratic and is less than 75% of the mean in 20% of the years. Hence the area can be classified as drought prone according to the standards fixed by the Indian Meteorological Department.

Historical monthly average records of hydro meteorological variables such as temperature, wind speed, relative humidity, sunshine hour, precipitation and Evaporation have been data collected for the years 1992-2000 from Kavalur meteorological station, the nearest available meteorological station, about 20 km from the reservoir site. The data for the years 1992-1997 were used in model training and 1998-2000 data were used in model testing.

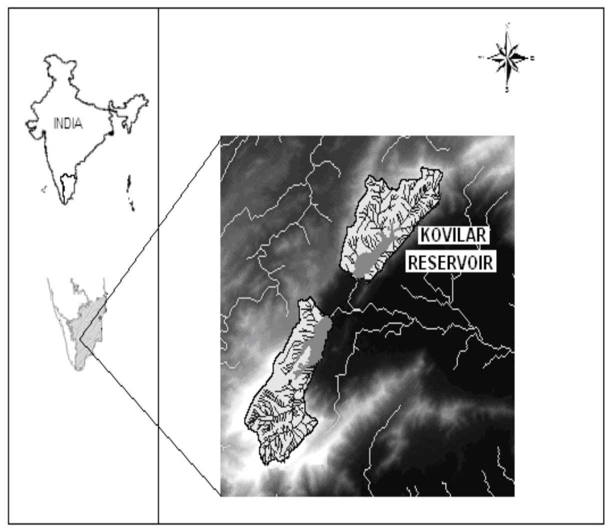


Fig. 1 The study area: Kovilar reservoir, Vaipar basin, India.

3. Materials and methods

3.1 Genetic Programming

Genetic Program (GP) is an automatic programming technique for evolving computer programs to solve, or approximately solve, problems [13]. In engineering applications, GP is frequently applied to model structure identification problems. In such applications, GP is used to infer the underlying structure of either a natural or experimental process in order to model the process numerically. Genetic programming has recently been applied to `real-world' problems [5]. This is a search procedure that uses random choice as a tool to direct a highly exploitative search through the numerical coding of a given parameter space. Goldberg [7] suggested that the probabilities of crossover and mutation are used in the simple genetic algorithm to effectively control the operators of crossover and mutation and they must range from 0 to 1. The final output of the model consists of independent variables and constants, and the functional set consists of basic mathematical operators {+, –, x, /, sin, cosh, log, power …} that may be used to form the model. The choices of operators depend upon the degree of complexity of the problem to be modelled.

The following parameters are considered while creating the dynamic models of Kovilar reservoir.

$$E_{t\,comp} = f\,(T_t, RH_t, Nh_t, V_t, E_{t\,obs}) \quad (1)$$

where,
$E_{t\,comp}$ = Computed Evaporation losses at time t (Mm^3)
$E_{t\,obs}$ = Observed Evaporation losses at time t (Mm^3)
T_t = Temperature at time t (°C)
RH_t = Relative Humidity at time t (%)
Nh_t = Sunshine hour at time t (Hours/day)
V_t = Wind velocity at time t (Kmph)

The mathematical models evolved from GP are presented as

$$E_t = \left(\cos\left\{ T_t \left[\sin\left(\frac{-1}{0.107 V_t} \right) - RH_t \right] \right\} \right)^2 * V_t * T_t \qquad (2)$$

In equation (2), the temperature term explicitly appears. This is mainly because the evaporation losses depend on temperature. The prediction is found to be better in the zone of mean of the evaporation values i.e. about 0.1 Mm^3.

3.2 Thornthwaite model

The Thornthwaite method [22] was used to estimate the potential evaporation of the basin. This method uses air temperature as an index of the energy available for evaporation, assuming that air temperature is correlated with the integrated effects of net radiation and other controls of evaporation, and that the available energy is shared in fixed proportion between heating the atmosphere and evaporation. There is no correction for different vegetation types. It merely requires inputs of mean monthly temperature, precipitation, latitude, and an estimate of soil storage capacity. Thus, the Potential Evaporation (PE) values are calculated based on observed temperature (and day-length) data and, when applied to typical, calibrated paired experimental watersheds where one is treated and the other is a control, may be correlated with and evaluated for any change in runoff quantity and timing. Ultimately, Thornthwaite method is useful for description, classification, management, and research. The method's biggest shortcoming is the minimum time division as month. That may produce a situation where end-of-the-month precipitation should not in fact appear as runoff until the following month, a delay that may be confusing in the computed water balance. However, the mean annual computations do wrap around: for example, mean annual January runoff, is based on mean annual December runoff.

The Thornthwaite's empirical equation is:

$$PE = 16\left(\frac{10t}{I}\right)^a \qquad (3)$$

where,
PE = Potential evapotranspiration in centimeter per month.
t = Mean monthly air temperature (°C)
i = the monthly heat index and it is expressed as:

$$i = \left(\frac{t}{5}\right)^{1.514} \qquad (4)$$

I = Annual heat index and it is given by the equation:

$$I = \sum_{j=1}^{12} i \qquad (5)$$

j = 1, 2, 3 ...12 is the number of the considered months.

$$a = 6.7 * 10^{-7} I^3 - 7.7 * 10^{-5} I^2 + 0.018 I + 49 \qquad (6)$$

The computed monthly potential evaporation in Equation 3 is for a standard month with 360 hours of daylight. It must be corrected for the varying length of day with latitude using the appropriate correction factor. For the evaluation of the actual evaporation the Thornthwaite water balance model [6] was utilized. The required parameters to determine actual evaporation using this model are mean monthly precipitation mean monthly potential evaporation.

3.3 Markovian model

The Markovian model was used to forecast the required meteorological data. After that, these forecasted parameters were used for the model prediction. In general, the forecasted series preserve the statistical properties such as mean and standard deviation of observed series. Errors in estimate of statistical properties due to depth of observed data and it reflects in generated values also. Finally, the forecasting is carried out using the following relationship from the Markovian chain series analysis. The following equation represents m[th] order Markovian model.

$$q_i = \underbrace{b_0 + b_1 q_{i-1} + b_2 q_{i-2} + \dots\dots\dots\dots + b_m q_{i-m}}_{\text{deterministic part}} + \underbrace{e_i}_{\text{random part}} \quad (7)$$

In this study, the below Markovian equation is used to forecast the values

$$y_{t+1} = \bar{y}_t + \frac{\sigma_t}{\sigma_{t-1}} r_{(t-1)-t} (y_{t-1} - \bar{y}_{t-1}) + e\sigma_t \sqrt{1 - r^2_{(t-1)-t}} \quad (8)$$

where,

y_{t+1} = forecasted value at time (t+1)

$\bar{y}_t$ = average of actual value at time (t)

y_{t-1} = actual value at time (t-1)

$\bar{y}_{t-1}$ = average of actual value at time (t-1)

σ_t = standard deviation at time (t)

σ_{t-1} = standard deviation at time (t-1)

r = correlation coefficient between time series (t), (t-1)

e = random part, rectangularly distributed (0, 1)

4. Results and discussion

The coupled data driven-evaluation scheme (Genetic Process and Genetic Algorithm) for evaporation prediction in reservoir was developed and demonstrated. The variable T_t, RH_t, V_t appears in equation 2. The above parameters in the equation shows, the evaporation in Kovilar reservoir is related with meteorological parameter temperature, wind velocity and relative humidity from the given input variables. The velocity term appears in the model shows that velocity of wind imparts main reason for evaporation. This is observed that the study area don't have obstruction, hence free wind movement occurs. And relative humidity of an air-water system is dependent not only on the temperature but also on the absolute pressure of the system of interest. Similar findings were earlier reported by [11]. The term sunshine hour does not appear in the developed model. It is observed that this term indirectly appeared in the form of temperature. During the GP model development, the exponential term was included. But in final model instead of exponential term, the trigonometrical term arrived. This shows that there is a chance for the seepage along with evaporation.

Thornthwaite method shown lesser efficiency and correlation coefficient compare to GP model both in training and testing periods in table 1. This probably due to the Thornthwaite method considers only temperature and precipitation in the model. In GP model, evaporation process depend on many more meteorological parameters such as temperature, relative humidity, and wind velocity which are directly as well as indirectly affecting the process. At the same time as the GP model performs very well because the GP has the ability to learn the non-linearities. The fig 2 shows the observed and computed value of GP during training and testing period. It uses the appropriate functions that are having input-output relationship. Moreover, the optimization takes place both constant and variables which involves in the model generation.

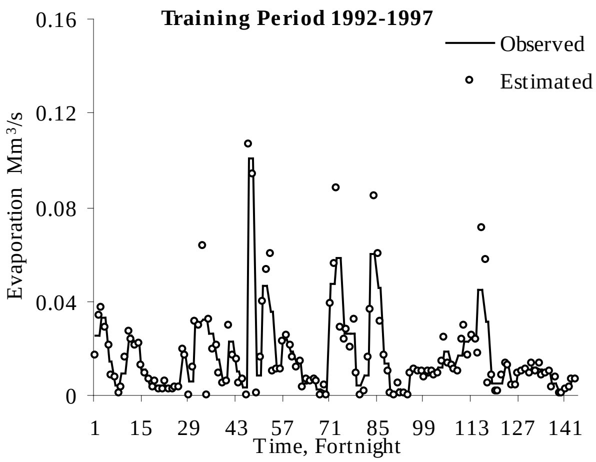
Table 1: Performance evaluation of static GP and Thornthwaite model

Model	**Model performance in different period**			
	Training (1992-1997)		**Testing (1998-2000)**	
	CC (%)	**CE (%)**	**CC (%)**	**CE (%)**
GP	88.7	77.4	84.3	73.9
Thornthwaite	73.1	68.5	67.0	61.3

The developed GP evaporation model has been utilized for finding the possible future climate change in Pilavakkal reservoir. This needs three parameters (i.e. temperature, wind velocity and relative humidity) for the estimation of evaporation at a given time (t). Hence, the forecasting of these three parameters performed using Markovian chain series analysis for the next 50 years (2001-2050).

Markovian chain series analysis determines the relationship between dependent variables (temperature, wind velocity and relative humidity) and independent variable (time). The major advantage of using this method is taking the sessional variation of climatical parameters. It satisfies the statistical conditions by taking care of standard deviation and correlation between the data serious. The randomly generated part minimises the error in the system. Thus, it provides meaningful forecasted values. It is clearly presented in the table 2. It shows the values of temperature variation in the next 50 years. Mean annual temperature and evaporation values are calculated for the year 2000 from the observed data and it is reported as 29.08°C and 30.9 $x10^3$ m^3 respectively. It is fixed as a base period to calculate the increase in temperature and evaporation values in future. Based on this, the variation is presented in table 2. Due to the climate change effect, there is an increase in annual temperature 1.09°C in the year 2025 and 2.14°C in the year 2050 occurred in Kovilar reservoir.

Prediction of evaporation is carried out by obtained forecasted values from Markovian method. Table 3 clearly shows the changes of evaporation in the next 50 years.



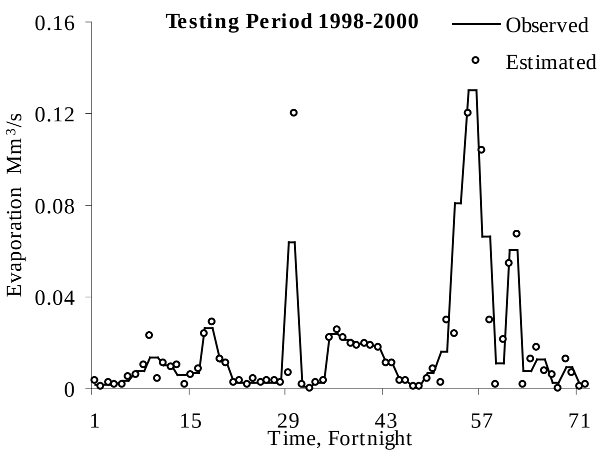


Fig. 2 Observed and estimated evaporation by GP model for Kovilar reservoir.

Table 2: Forecasting of temperature using Markovian chain series analysis

No	Time series	Length of Record in years	Annual Mean Value in °C	Inc. in actual value from base period in °C
1	1992-2000	9	29.08 (2000)	----
2	2001-2025	25	30.17 (2025)	1.09
3	2026-2050	25	31.22 (2050)	2.14

Table 3: Prediction of evaporation using Genetic Programming modeling approach

No	Time series	Length of Record in year	Annual Mean Value in m^3x10^3	Inc. in actual value from base period in m^3x10^3	Inc. in %
1	1992-2000	9	30.90 (2000)	----	----
2	2001-2025	25	34.30 (2025)	3.40	11
3	2026-2050	25	36.13 (2050)	5.23	17

*(Annual mean value at a particular time period)

Evaporation is directly proportional to temperature. From the obtained results, it is verified that mean annual evaporation loss increases upto 17% with 7.35% increase in temperature during they year 2050. Finally, it is observed that the approach adopted here is more rigorous in regional basis. It considers changes in temperature, and evaporation due to climate change scenario.

5. Conclusions

Two approaches (GP and Thornthwaite method) have been applied for Kovilar reservoir. From this study, GP has good potential for the non-linear effects by considering the parameters which having more influence to affect the evaporation process. GP model results are compared with Thornthwaite model. GP model gives satisfactorily better performance than Thornthwaite model. Thornthwaite model underestimates the evaporation this may be due to the empirical nature. These models are calibrated for a certain set of conditions and that it may not be applicable outside of this environment. These models, being developed with the local data alone, offers obvious advantages over the general evaporation models which are very generic. Thus, GP model provides the promising algorithm which can accurately estimate the local evaporation losses. Markovian chain series generates a reliable forecasted value which is used to calibrate the GP model. These models could be a better choice for climate change prediction in regional basis level. This can be coupled with General Circulation Models (GCMs) to predict the global climate change.

References

[1] A. Agarwal, and R. D. Singh, "Runoff modelling through back propagation artificial neural network with variable

rainfall-runoff data", Water resources management, Vol. 18, No. 3, 2004, pp. 285-300.

[2] V. Babovic, and M. Keijzer, "Rainfall runoff modeling based on genetic programming", Nord Hydrol., Vol. 33, 2000, pp. 331–346.

[3] H. F. Blaney, and W. D. Criddle, "Determining water requirements in irrigated area from climatological irrigation data", Soil Conservation Service Technical Paper no. 96., US Department of Agriculture, Washington DC, USA, 1950.

[4] U. K. Chakraborty, and D. G. Dastidar, "Using Reliability Analysis to Estimate the Number of Generations to Convergence in Genetic Algorithms", Information Processing Letters, Vol. 46, 1992, pp.199-209.

[5] L. Davis, "Handbook of Genetic Algorithms", Van Nostrand Reinhold, New York. 1991.

[6] T. Dunne, and L. B. Leopold, "Water in Environmental Planning", W. H. Freeman, New York, 1978.

[7] D.E. Goldberg, Genetic algorithms. In: Search, "Optimization and Machine Learning", Addison-Wesley, New York, 1989.

[8] IPCC, Climate Change 1995: The science of Climate Change. J. T. Houhgton, L. B. Meira Filho, B. A. Callander, N. Harris, A. Kattenberg, and K. Maskell, (eds.). Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, 2001.

[9] A.W. Jayawardena, and D. A. K. Fernando, "Use of radial basis function type artificial neural networks for runoff simulation", Computer-Aided Civil and Infrastructure Engineering, Vol. 13, No. 2, 1998, pp. 91-99.

[10] P. Kamban, A. Elshorbagy, and S. K. Carey, "Modelling the dynamics of the evapotranspiration process using genetic programming", Hydrological Sciences, Vol. 52, No. 3, 2007.

[11] K. S. Kasiviswanathan, and K. G. Jeganathan, "Regional empirical models for predicting evaporation losses from reservoirs", B.Tech thesis work, 2006.

[12] S. T. Khu, S. Y. Liong, V. Babovic, Madsen, and N. Muttil, "Genetic programming and its application in real-time runoff forecasting", J Am Water Resour Assoc., Vol. 37, 2001, pp. 439–451.

[13] J. R. Koza, "Genetic programming: on the programming of computers by natural selection", MIT Press, Cambridge, MA, 1992.

[14] V. Lakshminarayana, and S. P. Rajagopalan, "Optimal cropping pattern for basin India", J. of Irrig. Drain. Division, ASCE 103, (IR1), 1977, pp. 53-71.

[15] J. L. Monteith, "Evaporation and environment in the state and movement of water in living organisms", In: Proc. Society of Experimental Biology., Symposium no. 19, 1965, pp. 205–234, Cambridge University Press, Cambridge, UK.

[16] H. L. Penman, "Natural evaporation from open water, bare soil and grass", Proc. Roy. Soc. London, Vol. 193, 1948, pp. 120–146.

[17] C. H. B. Priestley, and R. J. Taylor, "On the assessment of surface heat flux and evaporation using large scale parameters", Mon. Weather Rev., Vol. 100, 1972, pp. 81–92.

[18] H. Raman, S. Mohan, and N. C. V. Rangacharya, "Decision support for crop planning during droughts", J. of Irrig. Drain. Eng., ASCE Vol. 118, No.2, 1992, pp. 229-241.

[19] J. C. Stephens, and E. H. Stewart, "A comparison of procedures for computing evaporation and evapotranspiration", In: Publication 62, International Association of Scientific Hydrology, 1963, pp. 123–133, International Union of Geodesy and Geophysics, Berkeley, California, USA.

[20] G. Surabhi, and M. C. Deo, "Real-time wave forecasting using genetic programming", Ocean Engineering, Vol. 35, 2008, pp. 1166– 1172.

[21] C. W. Thornthwaite, "An approach toward a rational classification of climate", Geographical Review, Vol. 38, 1948, pp. 55-94.

[22] C. W. Thornthwaite, and J. R. Mather, "Instructions and Tables for Computing Potential Evapotranspiration and the Water Balance", Publ. in Climatology, Vol. 10, No.3, 1957.

[23] B. Zhang, and S. Govindaraju, "Prediction of watershed runoff using Bayesian concepts and modular neural networks", Water Resources Research, Vol. 36, No.3, 2000, 753-762.

***K.S. Kasiviswanathan** has received his master degree in water resources development in IIT Roorkee and currently pursuing his PhD in IIT Madras. His research interest includes application of soft computing tools in water resources with uncertainty quantification.

R. Soundhara Raja Pandian has received his BE (ECE) in Anna University, Chennai, MBA (HR) in Madras University and currently pursuing ME in Embedded System Technologies. His core research area includes Parallel Computing and Cloud Computing in Education.

S. Saravanan obtained his doctorate in IIT Roorkee and currently working as an assistant professor in NIT Trichy. His research area includes watershed management with GIS applications. He has published several papers in reputed international journals and conferences.

Avinash Agarwal is a senior scientist in National Institute of Hydrology (NIH), Roorkee. He has published several international journal papers and presented research papers in conferences. He mostly deals in surface water hydrology with soft computing tools.

A Framework to Enhance Quality of Service for Content Delivery Network Using Web Services: A Review

Manivannan.K[1], Ravichandran.C.G[2]

[1]Research Scholar, Anna University of Technology
Madurai, Tamilnadu- 625002, India

[2] RVS College of Engineering and Technology,
Dindigul, Tamilnadu-624005, India

Abstract

Content Delivery Networks (CDNs) is anticipated to provide better performance delivery of content in internet through worldwide coverage, which would be a fence for new content delivery network providers. The appearance of Web as a omnipresent media for sharing content and services has led to the rapid growth of the Internet. At the same time, the number of users accessing Web-based content and services are growing exponentially. This has placed a heavy demand on Internet bandwidth and Web systems hosting content and application services. As a result, many Web sites are unable to manage this demand and offer their services in a timely manner. Content Delivery Networks (CDNs) have emerged to overcome these limitations by offering infrastructure and mechanisms to deliver content and services in a scalable manner, and enhancing users' Web experience. The planned research provides a framework designed to enhance QoS of Web service processes for real time servicing. QoS parameters of various domains can be combined to provide differentiated services, and allocating dynamically available resources in the midst of customers while delivering high-quality real time multimedia content. While accessing the service by a customer, it is possible to adapt real time streams to vastly changeable network conditions to give suitable quality in spite of factors upsetting Quality of service. To reach these intentions, adaptive web service processes to supply more information for determining the quality and size of the delivered object. The framework includes a section for QoS monitoring and adaptation and QoS faults prediction possibility and convalesce actions in case of failure. The aim of this research is to encourage research about quality of composite services in service-oriented architectures with security measures.

Key words: *Content Delivery, Quality of Service, Web Service, Real Time Service Provisioning.*

1. Introduction

Applications of CDNs can also be found in many communities, such as academic institutions, advertising media and Internet advertisement companies, data centers, Internet Service Providers (ISPs), online music retailers, mobile operators, consumer electronics manufacturers, and other carrier companies. Along with the proliferation, formation, and consolidation of the CDN landscape, new forms of Internet content and services are coming into picture while distribution and management of content is introducing new challenges in this domain. This raises new issues in the architecture, design and implementation of CDNs. The technological trends in this domain need to be explored in order to provide an exclusive research roadmap to the CDN community. The Real Time Streaming Protocol (RTSP) is a network control protocol designed for use in entertainment and communications systems to control streaming media servers. The protocol is used to establish and control media sessions between end points. Clients of media servers issue VCR-like commands, such as *play* and *pause*, to facilitate real-time control of playback of media files from the server. The transmission of streaming data itself is not a task of the RTSP protocol. Most RTSP servers use the Real-time Transport Protocol (RTP) for media stream delivery; however some vendors implement proprietary transport protocols. The RTSP server from Real Networks, for example, also features Real Networks' proprietary RDT stream transport. Web services can also be used to implement architecture according to Service-oriented architecture (SOA) concepts, where the basic unit of communication is a message, rather than an operation. This is often referred to as "message-oriented" services.SOA Web services are supported by most major software vendors and industry analysts. Unlike RPC Web services, loose coupling is more likely, because the focus is on the "contract" that WSDL provides, rather than the underlying implementation details. Middleware Analysts use Enterprise Service Buses which combine message-oriented processing and Web Services to create an Event-driven SOA.At the dawn of the third millennium a new breed of web application has risen: Web Services (WSs) These services are "self-contained, self-describing, modular applications that can be published, located, and invoked across the Web.Once a Web service is deployed, other applications (and other Web services) can discover and invoke the dep service." Since they first appeared, several research groups have worked on building efficient frameworks that enable the deployment of web services, exploiting technologies such as : XML (Extensible Markup Language), SOAP3 (Simple object Access Protocol), UDDI(Universal Discovery, Description and Integration),WSDL5 (Web Services Description Language), SOA (Service Oriented Architecture) , etc. Some of these groups have focused on developing ontologies that capture the WSs' main properties. Nevertheless, little work has been done to represent the non-functional features of WSs, the most critical part of which concerns their Quality of Service (QoS). Integrating QoS features in the profile of WSs is to the advantage of both users and providers. QoS profiles of WSs are crucial in determining which service best addresses the user desires and objectives. If the discovered WSs are accompanied with descriptions of their non-functional properties, then the automated WS selection and composition that takes place, considers the user's QoS preferences in order to optimize the user's WS-experience regarding features such as performance, reliability, security, integrity, and cost. On the other hand, QoS can give WS providers a significant competitive advantage in the e-business domain, as QoS aware services meet user needs better and thus attract more customers. Adopting a WS best effort policy that does not provide any guarantees on response time, security, throughput, or availability, may still be acceptable in light, non-time-critical and non privacy-sensitive WSs (e.g., static weather forecast report service); it is, however, totally unacceptable in more demanding cases, when for example dynamic composition of various heterogeneous WSs is required. Moreover, QoS awareness in WS provision, coupled with dynamic network resource allocation mechanisms, enables providers to maximize the utilization of their infrastructure, thus contributing to the increase of their profits. Lately, some research teams, having identified the importance of QoS featured WS profiles, have started to work on building QoS ontologies for web services, mainly focusing on developing ontology vocabularies, i.e., identifying the various QoS ontology parameters that are involved in web service provision However, as QoS parameters can be a lot more than type-value pairs, the need to develop a uniform way to efficiently represent the plethora of information concerning QoS parameters in a machine interpretable manner, while supporting enhanced reasoning functionalities, has appeared. A QoS ontology language that provides a standard model to formally describe arbitrary QoS parameters and exhibits properties such as completeness, flexibility, interoperability, reliability, scalability and accuracy. This language, combined with the proposed vocabulary, formulates a robust QoS semantic framework for WSs that can increase both the users' satisfaction and the providers' gains. The aim of this review focuses to the design of a framework to enable the QoS analysis of Web-service processes for real-time service provisioning (RTSP) based on service compositions. An integrated approach to quality of service for content delivery using Web services includes

- Quality definitions for the framework model
- User contracts
- Fault monitoring System
- Security measures and also
- QoS broker design that can be used in providing QoS Web services efficient;
- The end-to-end QoS issue for Web service composition;
- The Several complex service selection algorithms to be used by QoS brokers.

2. QoS Framework – A Review

2.1 Web services and streaming delivery

The framework deals with processes interacting with different actors and offering value added services that are able to satisfy user requests for complex objects, such as an e-learning object, a clinical health service, or an e-government service. The methods of quality analysis and the reference-tool architecture that combine the worlds of Web services and streaming by focusing on jointly provisioning complex services and their quality. We assume that the environment is composed of several nodes operating at two layers: Web services and their related protocols, and RTSP protocols. Figure 1 shows the reference scenario: a user requires, and eventually receives, a complex service obtained as a composition (possibly a choreography3) of different Web services; one of these (WS2 in the figure) provides streaming content. The main concerns of this review include addressing problems associated with the guarantee of QoS requirements in variable contexts and providing an active approach to solving or anticipating possible failures. Therefore, the focus of this article is not only on monitoring, but also on anticipating faults with prediction techniques.

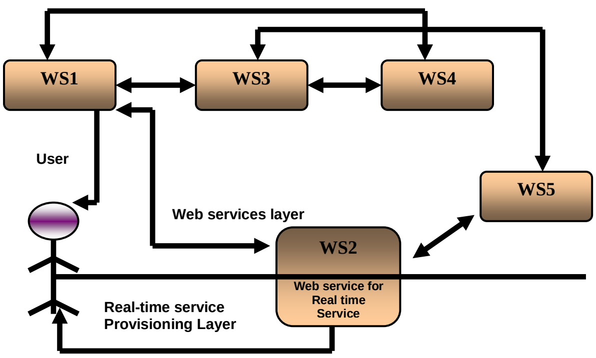


Fig 1 Reference Scenario

2.2 QoS definition

Defining a general QoS model is essential. Normally there are two QoS models, one for the Web service layer and one for the RTSP layer. The use two ontologies to represent quality parameters, with the semantics conforming to methodologies and techniques used in the Semantic Web community. The OWL Web Ontology Language to develop the QoS ontologies, and followed the conceptual structure proposed by Papaioannou. Web-servic QoS model relies on the following parameters describe the QoS related to a single synchronous operation provided by the server:

- *Response time*: Time elapsed between the instant a request is sent from the client and the instant the server computes the response.
- *Price*: Amount a client pays to the server for operation provisioning.
- *Availability*: Probability that a given operation is accessible at the moment of the request.
- *Reputation*: Ratio of the number of invocations with the requested QoS and the total number of invocations.
- *Data quality timeliness*: Freshness (up-to-date degree) of data.
- *Data quality accuracy*: Correspondence between given data and reference data considered as correct.
- Data quality completeness: Coverage of exchanged data with regard to total data representing the managed information.

2.3 Real-time service provisioning layer

The QoS of a multimedia stream is based on two classes of parameters, namely:

1. User related - These express the user's requirements and preferences in accessing multimedia services, and allow the evaluation of relevance to the user of each component (video, audio, and data) of the delivered multimedia flow.
2. Network related - These parameters support the assessment of the amount of available network resources (bandwidth, channel speed, and so on).

2.3.1 User Contracts

A QoS contract between a provider (server) and a consumer (user) regarding a set of parameters. User contract consists of two parts:

- ✓ Mandatory part - consisting of the seven levels of QoS one for each QoS parameter and of a rule used to determine whether a QoS violation occurs;
- ✓ Optional part - related to specific aspects, such as the QoS of real-time contents provided by the server.Table1 provides the entire set of parameters defining the QoS at the RTSP layer.

Table 1 RTSP Layer QoS Parameters

Parameter	***QoS Parameter***
User Related	Access count Video access count Audio access count Data access count Video degradation count Audio degradation count Data degradation count *x*-resolution *y*-resolution Chrominance Luminance Frame rate Audio channels Audio codex Audio frequency
Network Related	Video bandwidth Audio bandwidth Data bandwidth

2.4 QoS for web services

Future Web-based systems require a seamless integration of user processes, server applications, domain intelligence, and Web services over the Internet. Delivering QoS services for most multimedia and real-time applications is a critical and significant challenge because of the dynamic and unpredictable nature of user applications and Internet traffic. User applications with different profiles and requirements compete for the resources used to provide Web services. Without a careful management of Web service QoS, critical applications may suffer performance degradation, and resulting in customer dissatisfaction or media losses. The area of QoS management covers a wide range of techniques that match the needs of service requestors with those of the service provider's. QoS has been a major area of study in computer networking, real-time computing , and system middleware. For Web services, QoS guarantee and enhancement have started to receive some attention. The proposed work only consider the following quality attributes as part of Web service parameters.

- Response time (T): The amount of time to get a service request fulfilled at the client side. This includes service time Ts and transmission time Tt: T = Ts + Tt .
- Service time (Ts): The time a server needs to process a service request. The information is furnished by the service provider;
- Transmission time (Tt): The time needed to send a request to a server and get the result from the server (i.e., round trip communication time). It is decided by the network.
- Cost (C): Includes service cost Cs and transmission cost Ct: C = Cs + Ct.

Service cost (Cs): Service cost is the service charge for each unit of service. A Web service may be priced differently depending on the quality of the (media) service delivered. It is set by the service provider;

Transmission cost (Ct): The price that a service requestor has to pay for transmitting a request to a server and transmitting the result data from the server to the requester. The transmission cost is decided by the network operator;

- Service availability (A): The probability that a service is available at some interval of time. This only measures the server availability in terms of responding to a request, not the result quality. It can be computed from historical data: A = Ta / Tt where Ta: the amount of time that a service is available;Tt: total time interval that is measured

- Service reliability (R): The probability that a request is correctly fulfilled within the expected time. It can be computed from historical data: R = Ns / N where
Ns: number of requests successfully fulfilled;
N: total requests.

- Network bandwidth: The minimum network bandwidth required to receive the service. This is especially important for services with multimedia content such as video or large graphics. The bandwidth attribute will also be important for Web service brokers to decide if a service should be invoked if the client is using a low bandwidth network such as wireless connections. The above QoS attributes that the work is consider in the framework are both easy to understand and to measure. These attributes can be collected on a system without user intervention. For example, before and after each connection and invocation of a Web service, a software agent can automatically measure the response time, the service cost, the bandwidth used, and the number of connection attempts before the service is successfully delivered.

2.5 Prediction

To anticipate faults, the proposed work uses a prediction model and a support framework based on monitoring and machine learning. Because the global QoS varies in the runtime environment, to determine the global QoS by observing a set of parameters (the prediction global QoS is the tuple containing only the Web-service quality parameters).Some regularity can emerge from observation of the global QoS, depending on the values of specific parameters in different situations.

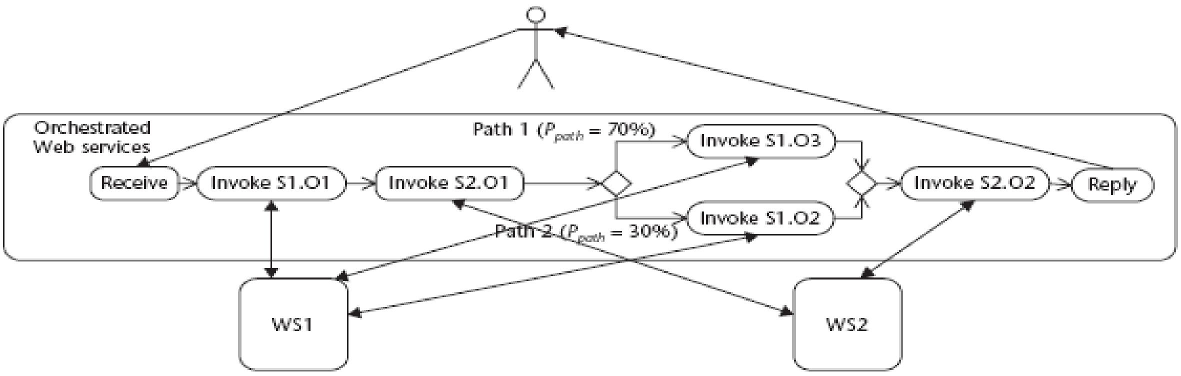


Figure 2. Sample Web services process for QoS prediction

For example, the global QoS on the same sequence of operations can change, and this regularity can be useful in determining the range variability. By observing these regularities, it can be define or predict the global QoS. In this scenario, a huge set of simulated data contained in the log files generated by the runtime environment is available, making it possible to analyze this data to define behavioral models. The model uses machine-learning techniques to build a system capable of providing suggestions on possible variations of the global QoS.Fig 2 shows a sample process useful in describing the learning problem. A process instance uses two services S1 and S2, and invokes several operations (O1, O2, . . ., On). To formalize the learning problem as follows: given an answer to a process operation Si.Oi and given the current global QoS level $Curr_{QoSLSi:Oi}$, let try to know with a certain probability $P_{globalQoS}$, the global QoS level global-QoS_j corresponding to $Curr_{QoSSSi:Oi}$, with j being an identifier of a future operation of S_i. The following features represent input instances for the classifier:

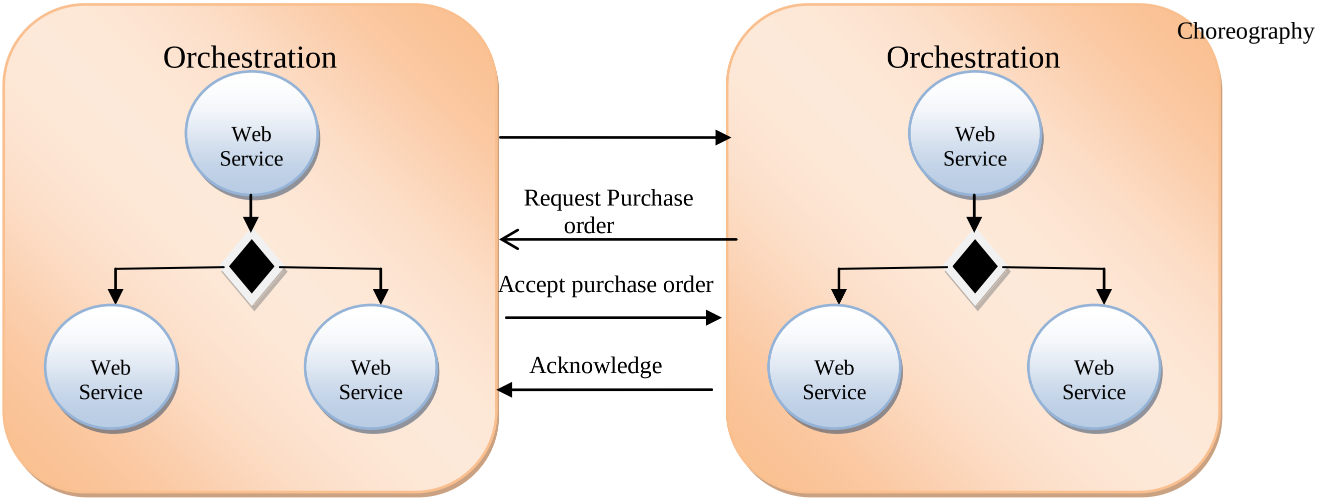


Fig 3 Orchestration versus choreography

- Service name S identifying the server providing the current operation;
- Operation name O identifying the current operation;
- QoS parameters T.Resp, P,Avail, Rep,DQ.Timel,DQ.Acc,and, DQ.Compl identifying the current values of each of the seven QoS parameters; and
- Target operation O_{Target} identifying the next operation upon which the QoS prediction will be performed.

2.6 QoS framework

The architecture monitors, detects, and predicts QoS faults, which the architectural components detect and manage by providing a self healing Web-service approach. The approach consists of making the services aware of possible faults and capable of repairing them. The business scenario consists of several coordinated services that provide real-time content. Process management occurs on the communication layers for Web-services interaction and for RTSP due to the technological differences between the two layers. However, it's necessary to manage the two layers uniformly from the user perspective, and hence the streaming server exposes a management interface to the Web-service layer. The separation is total for communication protocols, while for QoS, information exchange is enabled between the two layers to react to QoS faults.

2.6.1 Web-service Layer

In the Web-service layer, monitoring involves several purposes:

- Checking if the execution of the complex service correctly follows the interaction protocol defined by the global choreography
- Checking whether a QoS contract is respected;
- Estimating the QoS of following operations to prevent QoS faults.

Fig 3 shows Web-service choreography, none of the involved participants centrally executes the composite service. Each Web service is simply aware of its own status and doesn't have a global view encompassing all the cooperating services. However, the choreography definition represents a global perspectives on the composite service that a choreography-monitoring Web service can rely on to detect faults occurring at the choreography level. The choreography monitor, relying on the notification messages received by each Web service and on the global choreography description, can track the progress of the service execution.

2.6.2 Real-time service provisioning layer

Monitoring bandwidth available for RTSP plays a key role in the processes related to multimedia streaming. Specifically, knowledge of the bandwidth available on each network link enables the detection of bottleneck links. Monitoring bandwidth is therefore essential for regulating and improving the QoS associated with a streaming application. Table

2 refers the concept of global QoS as a tuple composed of the union of different QoS values or levels, due to the heterogeneity of the range domains.

3. Empirical Comparisons

The following table 2 shows the performance of the two frameworks

	Web Services and Streaming delivery framework			**QoS Ontology Framework Ambient QoS**		
	Input	Parameter	Values	Input	Parameter	Values
QoS Framework for CDN using Web Services	Current QoS	Response time	5 s	Current QoS Contract	Response Time	6s
		Price	7 Euros		Price	20 Euros
		Availability	0.9		Availability	0.88
		Reputation	0.75		Capacity	200
		Data quality timeliness	0.75		Scalability	0.80
		Data quality accuracy	0.75		(max) Jitter	1 (msec)
		Data quality completeness	0.75		(max) Error Rate	10-5
					(max) Latency	300(msec)
					(min) Throughput (Kbps)	384

Table 2 Empirical Comparisons – QoS Framework for CDN using Web Services

The following graph shows the performance of Ambient QoS Framework.

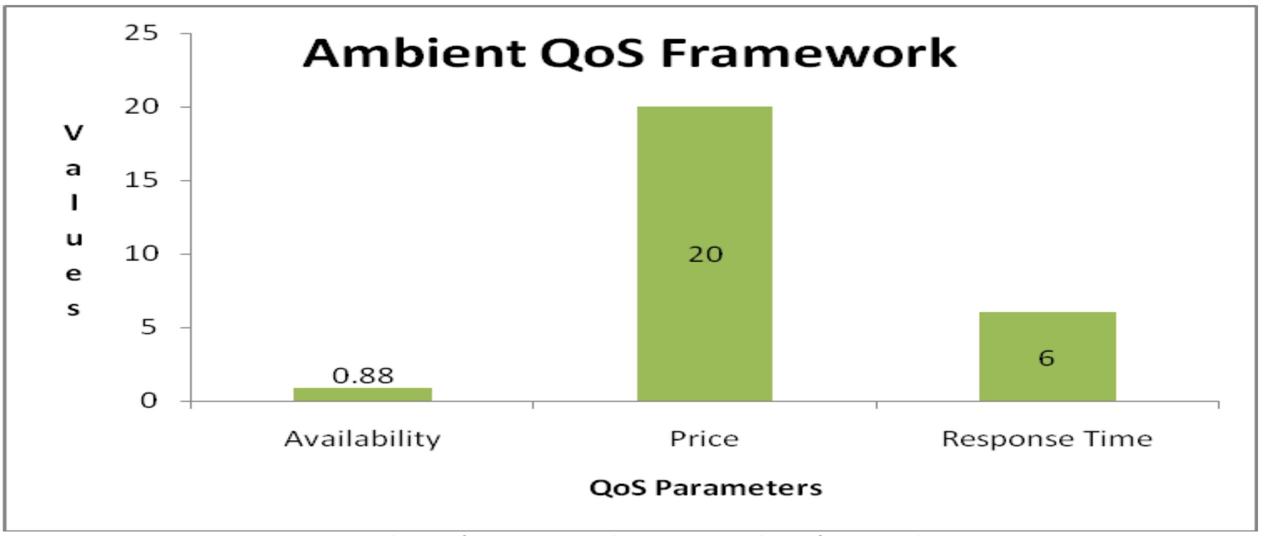


Fig 4. The performance graph – QoS Ontology framework

The following graph shows the performance of QoS Framework with streaming delivery.

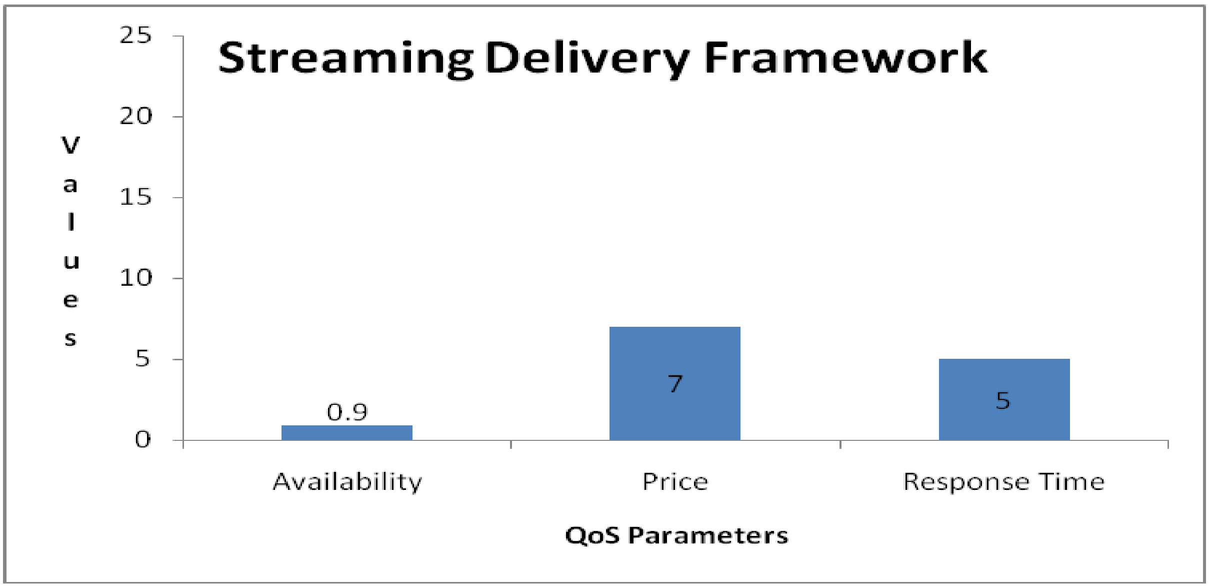


Fig 5 The performance graph – QoS framework with streaming delivery

4. Conclusion

The proposed model has End-to-End management infrastructure for applications, systems, and network. It gives them flexible control over business processes involving Web services. Content Delivery Networks (CDNs) address the problem of network congestion by storing and serving internet content from different distributed locations rather than from a few centralized origin points. The goal of QoS framework is to provide guarantees on the ability of a network to deliver predictable results. The network QoS refers to the ability of the network to handle the traffic such that it meets the service needs of certain applications.

5. References

[1] Francesco Buccafurri, Pasquale De Meo, Gianluca Lax, and Domenico Ursino university Mediterranea of Reggio Calabria, Mariagrazia Fugini, Stefano Modafferi, and Barbara Pernici Politecnico di Milano, "A Framework for Using Web Services to Enhance QoS for Content Delivery" IEEE Transactions on Intelligent and pervasive Multimedia systems, 2010.

[2] M. Pathan, R. Buyya, and A. Vakali, eds., Content delivery Networks, LNEE 9, Springer, 2008.

[3] F. Buccafurri et al., ''Analysis of QoS in Cooperative Services for Real Time Applications,'' Data and Knowledge Engineering, vol. 67, no. 3, 2008, pp. 463 484.

[4] D. Ardagna et al., ''PAWS: A Framework for Processes with Adaptive Web Services,'' IEEE Software, vol. 24, no. 6, 2007, pp. 39 46.

[5] J. Sokol and K.P. Eckert, ''MCDN: Multimedia Content Discovery and Delivery,'' Proc. Int'l Symp. Autonomous Decentralized Systems, (ISADS), IEEE Press, 2007, pp. 411-420.

[6] L. Ardissono et al., ''Monitoring Choreographed Services,'' Innovations and Advanced Techniques in Computer and Information Sciences and Engineering, T. Sobth, ed., 2007, Springer Verlag, pp. 283 288.

[7] T. Yu and K.J. Lin, ''QCWS: An Implementation of QoS-Capable Multimedia Web services,'' Multimedia Tools and Applications, vol. 30, no. 2, 2006, pp. 165-187.

[8] I.V. Papaioannou et al., ''A QoS Ontology Language for Web-Services,'' Proc. IEEE Int'l Conf. Advanced Information Networking and Applications,(AINA), IEEE CS Press, 2006, pp. 101-106.

[9] I.H. Witten and E. Frank, Data Mining: Practical Machine Learning Tools and Techniques, 2nd ed., Morgan Kaufmann Publishers, 2005.

[10] M. Little, E. Newcomer, and G. Pavlik, eds., Web Services Context Specification (WS-Context), OASIS committee draft, Nov. 2004; http://xml.coverpages. Org/WS-ContextCD-9904.pdf.

[11] D.L. McGuinness and F. van Harmelen, eds., OWL Web Ontology Language Overview, World Wide Web Consortium (W3C) recommendation, Feb. 2004;

[12] WS-Diamond team, ''WS-Diamond:WebServicesDIAgnosability,MONitoring, and Diagnosis,'' to be published in At Your Service: An Overview of Results of Projects in the Field of Service Engineering of the IST Program, Series on Information Systems, MIT Press.

[13] W.M.P. van der Aalst, A.H.M. ter Hofstede, and M. Weske, ''Business Process Management: A Survey,'' Proc. Int'l Conf. Business Process Management, (BPM), LNCS 2678, Springer, 2003, pp. 1-12.

[14] M. Jain and C. Dovrolis, ''End-to-End Available Bandwidth: Measurement Methodology, Dynamics, and Relation with TCP Throughput,'' IEEE/ACM Trans. Networking, vol. 11, no. 4, 2003, pp. 537 549.

[15] G. Araniti et al., ''Adaptively Controlling the QoS of Multimedia Wireless Applications through User Profiling Techniques,'' J. Selected Areas in Comm., vol. 21, no. 10, 2003, pp. 1546-1556.

K.Manivannan received B.E degree in Computer Science and Engineering from the Anna University, Chennai and M.E degree in Computer Science and Engineering from the same University. He is currently working toward the Ph.D degree in Computer Science and Engineering at the Anna University of Technology, Madurai.His current research interests include web services, distributed systems and network architecture.

C.G.Ravichandran received B.E degree in Electronics and Communication Engineering from Bharathiyar University in 1988 and M.E degree in Electronics Engineering from Anna University, Chennai in 1991.He received the Ph.D degree in Medical Image Processing from Anna University, Chennai in 2009.Currently he is a Principal of RVS College of Engineering and Technology Dindigul,Tamilnadu,India. His research interests include image segmentation, medical image processing, network architecture, distributed systems and web services.

Improved Exemplar Based Texture Synthesis Method for Natural Scene Image Completion

K. Sangeetha[1], Dr. P. Sengottuvelan[2], E. Balamurugan[3]

[1] Department of Computer Applications, Bannari Amman Institute of Technology,

Sathyamangalam, Erode District 638 401, Tamilnadu, India

[2] Department of Information Technology, Bannari Amman Institute of Technology,

Sathyamangalam, Erode District 638 401, Tamilnadu, India

[3]Department of Computer Applications, Bannari Amman Institute of Technology,

Sathyamangalam, Erode District 638 401, Tamilnadu, India

Abstract

Image inpainting or image completion refers to the process of filling in the missing data of an image in a visually plausible way. Given a natural scene input image with selected target regions, both the geometrical property and texture information should be propagated from the known parts of image into the target regions. Exemplar-based model fills the target region using the source exemplars in the known parts of image. Many works on this subject have been proposed these recent years. In this paper, we present an inpainting method based on the improved exemplar-based texture synthesis technique which includes image gradient information during the inpainting process. A comparison with some existing methods on different natural images shows the strength of the proposed approach.

Keywords: *Image Inpainting, exemplar based, Texture synthesis, image completion, Partial Differential Equation (PDE)*

1. Introduction

Image inpainting is an iterative method for repairing damaged pictures or removing unnecessary elements from pictures. This activity consists of filling in the missing areas or modifying the damaged images in a non-detectable way by an observer not familiar with the original images [1]. Applications of image inpainting range from restoration of photographs, films and paintings, to removal of occlusions, such as large unwanted regions, superimposed text, subtitles, stamps and publicity, from images. In addition, it is of significant importance in restoration of precious work of arts, calligraphies, and paintings in the digital museum with image inpainting technique [13].A number of algorithms address this image filling issue [2], [3], [7], [8], [12]; these image inpainting techniques fill holes in images by propagating linear structures (called isophotes in the inpainting literature) into the target region via diffusion. They are inspired by the partial differential equations (PDE) of physical heat flow, and work convincingly as restoration algorithms.

The PDE-based algorithm does not perform well for texture dominated pictures. For such cases the exemplar based algorithm is used instead [5], [10], [11]. Given a texture sample, the texture synthesis problem consists in synthesizing other samples from the texture. The usual assumption is that the sample is large enough to capture the stationarity of the texture. There have been many works extending texture synthesis to inpainting.

In the seminal paper [4], the authors have presented a simple yet effective nonparametric texture synthesis method based on local image patches. The texture is modeled as a Markov Random Field (MRF) by assuming that the probability distribution of brightness values for one pixel given the brightness values of its spatial neighborhood is independent from the rest of the image. The neighborhood is a square window around the pixel and its size is fixed by hand. The input of the algorithm is a set of model image patches and the task is to select an appropriate patch to predict the value of an unknown pixel. This is done by computing a distance measure between the known neighborhood of an unknown pixel and each of the input patches. The distance is a sum of squared differences (SSD) metric.

The authors of [5] proposed an extension of Efros and Leung's method, with two improvements. The first one concerns the order in which the pixels are synthesized. Indeed, a system for assigning priorities is used (assigning priorities to the pixels was also proposed in [6]. The priority order at a pixel is the product of a confidence term, which measures the amount of reliable information surrounding the pixel, and a data term, that encourages linear structures to be synthesized first. The second improvement is a speed-up process. Contrary to the original method in [4], when synthesizing a pixel, not only the value of this pixel is inpainted in the output image (using the patch that gives the smallest distance metric), but all the pixels in its neighborhood that have to be inpainted are filled in.

With these algorithms, the gap will be filled with non-blur textures, while at the same time preserve and extend the linear structure of the surrounding area. These algorithms, however, have some problems: firstly, it merely adopts a simple priority computing strategy without considering the accumulative matching errors; secondly, the matching algorithm for texture synthesis only uses the color information; thirdly, the filling scheme just depends on the priority disregarding the similarity. As a result of lacking robustness, their algorithm sometimes runs into difficulties and "grows garbage". To solve these problems, we propose an enhanced exemplar-based inpainting algorithm.

2. Exemplar-based Texture Synthesis Algorithm

The basic conception of exemplar based model is copying information from the source exemplar in the known parts of image into the target region [14]. The selection of the source exemplar is determined by similarity function. Similarity is computed by the difference between pixels in two exemplars. The input of the completion algorithm is an image which contains the masked target region D, and the output is a completed image with the region D filled.

In this paper, an improved exemplar-based image completion model is proposed. The improved model follows the basic conception in [5], and it does not need user's interaction. When the removing or restoring target region is selected, it processes image automatically. The whole image domain is Ω, the removing or restoring target region domain is D, and image information in Dc $(\Omega - D)$ is known; the basic unit of synthesis is the exemplar Ψ. Image completion generates image information in D according to the information in Dc, and the whole exemplar Ψ is copied every time. When there are different kinds of texture patterns in image, the linear structure is restored firstly [9], and then the proper exemplar could be found and copied. The difference between exemplars is used to weigh the exemplar similarity, and the gradient information is often adopted. When a textured image is processed, the geometrical difference between exemplars should be considered. So a composite similarity function which is determined by gradient information is used to compute the exemplar similarity. Then the proper textured exemplar could be found in the textured image completion.

3. The Proposed improved Exemplar based Texture Synthesis Algorithm

In [3], the texture image is inpainted with the algorithm from [4]. This algorithm can be using the method from [5]. Indeed, setting an inpainting order (depending on the edges and the confidence) to the pixels clearly leads to better results. Furthermore, copying an entire patch instead of only one pixel is faster. We here propose some improvements to this algorithm and the modified algorithm has the following properties:

• Properly completes the natural scene image when there are different kinds of texture patterns;
• Preserves the linear structure property;
• Finds the most similar exemplar in the textured image based on gradient information;

3.1 Improvement in Data Term

The first improvement concerns the data term in the priorities of [5]. The data term is defined as

$$D(p) = \frac{|\nabla I^{\perp}(p)|}{\alpha} \quad (1)$$

where $\nabla^{\perp}$ is the orthogonal gradient and a is a normalization factor (equal to 255 for grayscale images). It encourages the linear structures to be synthesized first and depends on the isophotes (contours) that eventually pass by *p*. If we compute this term on the texture image, we will only take into account the small contours and the noise contained in the texture, but not the important edges. Indeed, we can now use the tensors of texture *J*(*p*)

$$J(p) = \sum_{d=1}^{n} \nabla I_d(p) \nabla I_d^T(p), \quad (2)$$

to compute the data term where ∇I_d denotes the gradient image. The eigenvalues at pixel *p* are $\lambda^-(p)$ and $\lambda^+(p)$ and corresponding eigenvectors are $\theta^-(p)$ and $\theta^+(p)$.

Using these notations, the new data term is given by:

$$D(p) = \frac{\lambda^+(p) - \lambda^-(p)}{\alpha}. \quad (3)$$

3.2 Improvement in Search Direction

Another improvement concerns the directions in which the candidate patches are searched for. It can be applied directly to the algorithm from [5]. The idea is to remark that the best patch $\Psi_{\hat{p}}$ for the source patch Ψ_p is probably in the direction of the isophotes (the isophote direction is given by the eigenvector $\theta^-(p)$). We then propose to only look for the candidate patches Ψ_q that verify the following test:

$$\theta^-(p) \cdot \frac{p-q}{\|p-q\|} > 0.9. \quad (4)$$

In the original texture synthesis method, finding the candidate patch $\Psi_{\hat{p}}$ (centered at pixel $\hat{p}$) corresponds to solving

$$\hat{p} = \underset{q \in \bar{\Omega}}{\operatorname{argmin}} \sum_{r \in \Psi_p} M(r)\left(I(r) - I(q+r-p)\right)^2, \quad (5)$$

where *d* is the sum of square differences (SSD) function.

The best patches correspond to the patches that have the smallest associated distance measures. Restricting the direction of the search, eq. (5) becomes:

$$\hat{p} = \underset{q \in \bar{\Omega} | \theta^-(p) \cdot \frac{p-q}{\|p-q\|} > 0.90}{\operatorname{argmin}} \sum_{r \in \Psi_p} M(r)\left(I(r) - I(q+r-p)\right)^2. \quad (6)$$

Note that we compute the distance both on the texture and the structure images because the texture image can sometimes only contain non informative noise. We then finally get the following equation:

$$\hat{p} = \underset{q}{\operatorname{argmin}} \sum_{r \in \Psi_p} M(r)\left(\left(u(r) - u(r')\right)^2 + \left(v(r) - v(r')\right)^2\right), \quad (7)$$

where $r' = r - p + q$.

3.3 Steps in the Proposed Algorithm

1. For all the pixels p of the mask, compute the priorities P (p) =C (p)*D (p) (for the others, $p \in \bar{\Omega}$, P (p) =0). The pixels with higher values of P will be inpainted first.

2. Inpaint the texture image:
(a) Find the pixel $p \in \Omega$ having the highest priority value and that has not been inpainted yet.

(b) Deciding a pixel belongs to a strong Structure or Texture

• if $\lambda^+(p) - \lambda^-(p) < \beta$, then apply texture synthesis to the pixel p using equation (13) (in practice one patch is arbitrarily chosen between the best ones as in [4].

Copy image data from $\Psi_{\hat{p}}$ to Ψ_p for all the pixels of $\Psi_p \cap \Omega$.

• else do not change the pixel value.

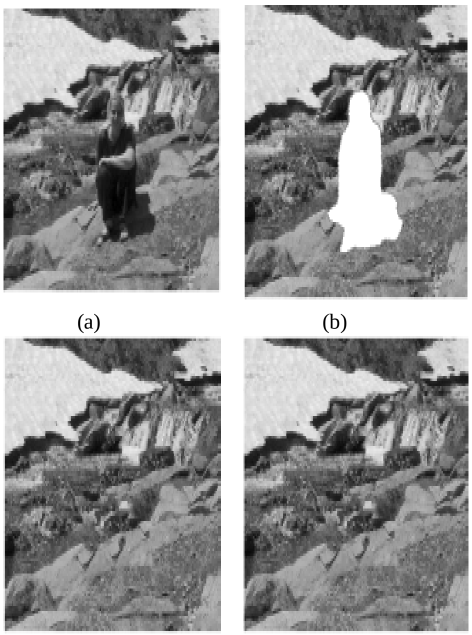
(c) Set $\Omega = \Omega \backslash p$.
(d) Return to (a).

4. Visual Comparison of Experiment Results

The improved exemplar-based algorithm is applied in different kinds of natural scene images to prove its validity in this section. The purpose of image completion is to restore the target region while satisfying the visual perception. The experimental results demonstrate that images generated by our algorithm have satisfactory results compared with Criminisi et al's algorithm [5].

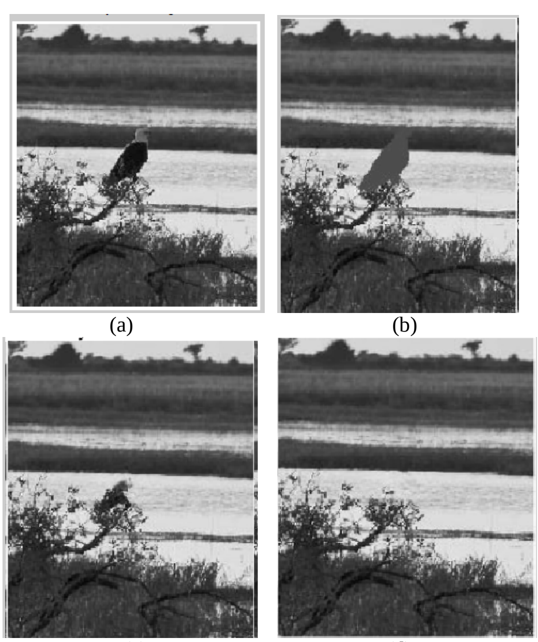
IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 2, September 2011
ISSN (Online): 1694-0814
www.IJCSI.org

Fig. 1 Reconstruction of lady occluded region of image. (a) Original Image. (b) The target region has been blanked out. (c) The final image in which the occluded area is reconstructed using Criminisi et al's algorithm. (d) Reconstructed image using proposed algorithm.



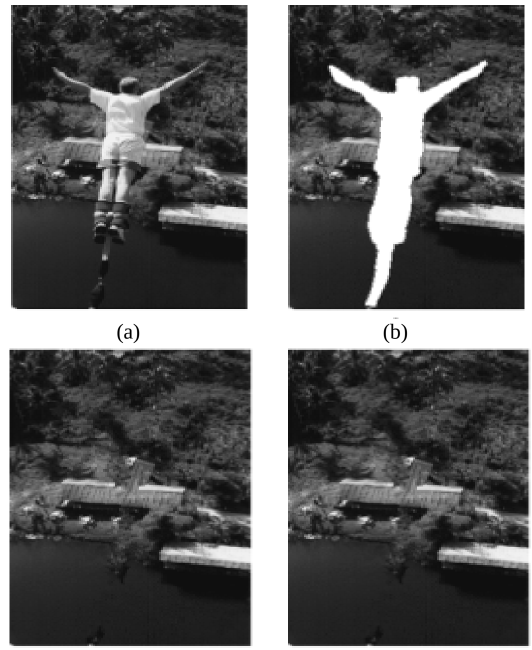
(a) (b) (c) (d)

Fig. 2 Reconstruction of bird occluded region of image. (a) Original Image. (b) The target region has been blanked out. (c) The final image in which the occluded area is reconstructed using Criminisi et al's algorithm. (d) Reconstructed image using proposed algorithm



(a) (b) (c) (d)

Fig. 3 Reconstruction of man occluded region of image. (a) Original Image. (b) The target region has been blanked out. (c) The final image in which the occluded area is reconstructed using Criminisi et al's algorithm. (d) Reconstructed image using proposed algorithm



(a) (b) (c) (d)

5. Conclusion

In this paper, we have proposed an image inpainting algorithm based on modifying the exemplar-based image inpainting method. The developed algorithm enhances the robustness and effectiveness by including image gradient information during the inpainting process. Several natural scene test images have been used and the results demonstrate that the developed algorithm can reproduce texture and as well as the structure of the surrounding area of the inpainted region. The method proved to be effective in completion of an image which has a hole after removing large objects from it, ensuring accurate propagation of textured pattern. The results obtained are preferable to those obtained by other similar methods. The examples presented demonstrate the effectiveness of the modified algorithm.

References

[1] M. Bertalmio, G. Sapiro, V. Caselles and C. Ballester, "Image Inpainting", Proceedings of SIGGRAPH 2000, New Orleans, USA, pp 417-424, July 2000.

[2] M. Bertalmio, A.L. Bertozzi, and G. Sapiro, "Navier stokes, fluid dynamics, and image and video inpainting",

Proceedings of Conf. Comp.Vision Pattern Rec., Hawai, pp.355–362, Dec 2001.

[3] M. Bertalmio, L. Vese, G. Sapiro, and S. Osher, "Simultaneous structure and texture image inpainting", Proceedings of IEEE conference on Computer Vision and Pattern recognition(CVPR'03), 2003.

[4] A. A. Efros and T. K. Leung, "Texture synthesis by nonparametric sampling," in Proceedings of International Conference on Computer Vision, 1999, pp. 1033-1038.

[5] Criminisi A., Perez P., and Toyama K., "Region filling and object removal by exemplar-based image inpainting", IEEE Trans. Image Processing 9, pp1200-1212, 2004.

[6] Harrison, P, " A non-hierarchical procedure for resynthesis of complex textures", In WSCG Conference Proceedings,2001.

[7] BianRu Li, Yue Qi, XuKun Shen, "An image inpainting method", Ninth International Conference on Computer Aided Design and Computer Graphics, IEEE, 2005.

[8] S. Roth and M. J. Black, "Fields of experts: A framework for learning image priors", Proc. IEEE Comp Vision and Pattern Rec., 2:860-867, 2005.

[9] Sun, J., Yuan, L., Jia, J., and Shum, H, "Image completion with structure propagation", In SIGGRAPH,ACM Special Interest Group on Comp. Graphics and Interactive Techniques, 2005.

[10] A. Criminisi, P. Perez, and K. Toyama, "Region filling and object removal by exemplar- based image inpainting," IEEE Transactions on Image Processing, Vol. 13, 2004, pp. 1200-1212.

[11] Sung Ha Kang, "Mathematical Approaches to Color Denoising and Image Inpainting Problems", Ph.D Thesis, University of California, Los Angeles 2002

[12] Manuel M. Oliveira Brian Bowen Richard McKenna Yu-Sung Chang, "Fast Digital Image Inpainting", Appeared in the Proceedings of the International Conference on Visualization, Imaging and Image Processing (VIIP 2001), Marbella, Spain, September 3-5, 2001

[13] Y. El Glaly, E.Atta, T. Nazmy, and B. Desouky, "Digital Inpainting for Cultural Heritage Preservation", International Conference on Computer Theory and Application, ICCTA, 2006

[14] Ji-Ying Wu And Qiu-Qi Ruan, "A Novel Exemplar-Based Image Completion Model", Journal Of Information Science And Engineering 25, 481-497 (2009).

Aspect-Oriented Requirements Engineering for Advanced Separation of Concerns: A Review

Narender Singh and Nasib Singh Gill

Department of Computer Science and Applications
Maharshi Dayanand University, Rohtak - 124001, Haryana, India.

Abstract

Software engineering was introduced to cope with software crisis with two fundamental principles: separation of concerns and modularity. Many programming paradigms have been proposed and used while considering the fundamental principles from the early days. Complex software systems were successfully modularized but complete separation of concerns is still impossible to achieve using today's most popular programming paradigms such as object-oriented programming. There are some concerns which could not be separated properly in a single class or module due to their highly coupling with other classes or modules' behaviors. We call such unmodularized concerns as crosscutting concerns and these are responsible for scattering and tangling.

Aspects are the natural evolution of the object-oriented paradigm. They provide a solution to some difficulties encountered with object-oriented programming, sometimes scattering and tangling. Hence, Aspect-Oriented Software Development (AOSD) is another step towards achieving improved modularity during software development. It gives emphasis to the separation of crosscutting concerns i.e. advanced separation of concerns and encapsulation of crosscutting concerns in separate modules, known as aspects. It later uses composition mechanism to weave them with other core modules at loading time, compilation time, or run-time. Aspect-Oriented Requirements Engineering (AORE) is an early phase in AOSD that supports separation of crosscutting concerns at requirements level. It does not replace but rather complements any of the existing requirements methodologies.

Over the last few years, several research efforts have been devoted to this area. Several techniques for crosscutting concern identification have already been proposed. In this paper, an attempt is made to review the existing approaches and understand their contribution to requirements engineering.
Keywords: *Separation of Concerns, Crosscutting Concerns, Aspect-Oriented Software Development, Aspect-Oriented Requirements Engineering.*

1. Introduction

The term "*Software Engineering*" was introduced in the NATO Working conference [1] on Software Engineering in 1968 to cope with *software crisis*. A number of approaches have been proposed to deal with the software crisis. Developing the complex software systems became easy. However, the progress in software engineering concepts did not keep track with increasing complexity of modern software systems. It is difficult to meet current and future needs in software development using today's most popular programming paradigm such as object-oriented programming.

According to the Standish Group in 1995 [2], only about 16% of software projects were successful, 53% were full with problems (cost or budget overruns, content deficiencies), and 31% were cancelled. The Standish Group's just-released report, "*CHAOS Summary 2009* [3]", only 32% of all projects succeeding which are delivered on time, on budget, with required features and functions" says Jim Johnson, chairman of The Standish Group, "44% were challenged which are late, over budget, and/or with less than the required features and functions and 24% failed which are cancelled prior to completion or delivered and never used". Evidence suggests that despite the improvement from 1995 to 2009 the current situation in software development is far from adequate.

Separation of concerns and *modularity* are the fundamental principles that drive the research in software engineering since the early days. The term "*separation of concerns*" was introduced by Edsger Dijkstra, to refer the ability of identifying, encapsulating and manipulating parts of software that are crucial to a particular goal or purpose in his book "*A Discipline of Programming*" [4]. The basic idea behind separation of concerns is to handle one property of a system at a time. In other words, a complex problem that is hard to understand should be divided into a series of smaller problems; those are less complex and easier to handle by the designer. These smaller problems may then be designed one at a time by different designers and finally integrated to solve the big problem. *Modularity* [5] [6] is the principle to structure software into modules where modules are self-contained, cohesive building blocks of software. A module is a device to implement a concern and modularity is a consequence of separation of concerns.

Many programming paradigm have been proposed with keeping the fundamental principles in mind. It becomes possible to modularize the complex software systems. But, it is still difficult to achieve complete separation of concerns using today's most popular programming paradigm such as object-oriented programming. There are some pieces of processing i.e. concerns that did not seem to fit in any particular single classes. This is because they are too tightly coupled to the behaviors in many other classes or modules. These unmodularized concerns are called as crosscutting concerns and are responsible for scattering: the implementation of a concern is spread over several program modules and tangling: a program module implements multiple concerns. Several empirical studies provide evidence that crosscutting concerns degrade code quality because they negatively impact internal quality metrics such as program size, coupling, and separation of concerns [7].

Aspects are the natural evolution of the object-oriented paradigm. They provide a solution to some difficulties encountered with object-oriented programming, sometimes scattering and tangling. AOSD [8] is another step towards achieving improved modularity during software development. It focuses on *crosscutting concerns* by providing means for their systematic identification, separation, representation and composition [9]. It encapsulates crosscutting concerns in separate modules, known as *aspects*. It later uses composition mechanism to weave them with other core modules at loading time, compilation time, or run-time [10].

AOSD was introduced first at programming level, with Aspect-Oriented Programming, where aspects are handled in code. A number of Aspect-Oriented Programming approaches have been proposed. Work has also been carried out incorporate aspects, and hence separation of crosscutting concerns, at the design level mainly through extensions to the UML metamodel [11] [12] [13]. However, crosscutting concerns are often present before the solution domain, such as in Requirements Engineering [14] [15] [16].

AORE [17] is still an emerging field with many open research issues. Research in the early phases of software development with aspect-oriented paradigm has been increasing. Handling crosscutting concerns in the early stages of software development is beneficial rather than handling them in later stages of software development because it not only makes the design simpler, but also helps to reduce the cost and defects that occur in the later stages of development. AORE focuses on identifying, analyzing, specifying, verifying, and managing the crosscutting concerns at the early stages of software development. It does not replace but rather complements any of the existing requirements methodologies.

In this section, an attempt is made to highlight the concept of AORE for advanced separation of concerns i.e. AORE for handling crosscutting concerns at the early stage of software development. Section 2 reviews the existing literature by many researchers. A roadmap to research is discussed in section 3. Finally, we conclude in section 4.

2. Literature Review

The success of a software system depends on how well it fits the needs of its users and its environment [18]. Software requirements comprise these needs, and *requirements engineering (RE)* is the process by which the requirements are determined [19]. Successful RE involves understanding the needs of users, customers, and other stakeholders; understanding the contexts in which the to-be-developed software will be used; modelling, analyzing, negotiating, and documenting the stakeholders' requirements; validating that the documented requirements match the negotiated requirements; and managing requirements evolution [20]. Existing requirements engineering approaches, such as use cases [21], viewpoints [22], and goals [23] provide good support for identification and treatment of some kinds of requirements. However, these requirements approaches do not explicitly support well broadly-scoped requirements, such as crosscutting concerns, and do not explicitly support their composition. Moreover, they all suffer from the "*tyranny of the dominant decomposition*" [24]. AORE, therefore, complements these approaches by providing systematic means for handling such crosscutting concerns.

Over the last few years, several research efforts have been devoted to developing AORE models that can help in extracting, identifying, and modeling aspects in the early phase of requirements analysis. In the following we discuss briefly these efforts.

Grundy's [25] proposed an approach called Aspect-Oriented Component Requirements Engineering (AOCRE) that focuses on identifying and specifying the functional and non-functional requirements relating to key aspects of a system each component provides or requires. In component-based systems, applications are built from discrete, inter-related software components, often dynamically plugged into running applications and reconfigured by end users or other components. Some components may have many aspects and others a few. Aspects may be decomposed into aspect details. Candidate components are found from OOA diagrams, by reverse engineering software components, or bottom-up consideration of individual, reusable components. The AOCRE process begins with identifying components' aspects, where for each component, we identify aspects for which the component provides services or requires services from other components using possible stakeholder requirements and object services. After identifying a component's aspects, we can reason about components and aspects. Further, we can infer inter-component relationships that allow engineers to reason about the validity relationships and aspects specified. Aspect-oriented component requirements also assist components design and implementation. They provide a focused set of functional and non-functional constraints for refining design, and provide a specification that an implementation can be tested against. AOCRE exhibit improved reusability and extensibility, and systems built with these components exhibit improved allocation of responsibility for data and behaviour among both reused and application-specific components. This approach is too specific for component development, not showing evidence of its use in software development in general. Besides, the identification of aspects for each component is not clearly defined and lacks tool support.

Rashid et al. [26] proposed a model for aspect-oriented requirements engineering. The model supports separation of crosscutting properties from early stages of the development and identification of their mapping and influence on later development stages. Now, it is possible to identify conflicts, establish possible tradeoffs, and promotes traceability throughout the system development and its evolution. The early separation of crosscutting concerns improves modularisation and hence, it is possible to build flexible and adaptable systems that meet the needs of volatile domains such as banking, telecommunications and e-commerce. The model supports separation of crosscutting properties from early stages of the development and identification of their mapping and influence on later development stages. But it lacks on validation of aspects, their composition with other requirements and resolution of possible conflicts resulting from the composition process. It also lacks a notation to describe aspects, their interactions and composition relationships at the requirements level.

Baniassad and Clarke [27] proposed the theme approach that provides support for aspect-oriented software development at two different stages. Theme/Doc, which is used for viewing and analyzing the requirements at requirements phase; and Theme/UML, which allows a developer to model features and aspects of a system, and specifies how they should be combined at design phase. In the theme approach, a theme is an element of design, which is a collection of structures and behaviours that represent one feature. Multiple themes can be combined or integrated to form a system. Themes are further classified as base themes, which may share some structure and behaviour with other base themes, and crosscutting themes which have behaviour that overlay the functionality of the base themes. Crosscutting themes are known as aspects. Action view is used to identify crosscutting behaviours. To create an action view, two inputs are needed: a list of key actions identified by the developer by looking at the requirements document and picking out sensible verbs, and the requirements as written in the original document. Theme/Doc then performs lexical analysis of the text and generates the action view. The theme approach involves three main activities as finding themes, modeling and composing themes, and checking Themes/UML. The process begins with finding themes. Here, we identify actions and generate an action view to examine their relationships. After analysis of the view, we determine that all of these actions will not be modelled as separate themes. Instead, we determine the relationships between the actions to decide how to group the actions into larger themes. Here, we also determine which themes are crosscutting and which are base. The next step in theme process is modelling and composing themes. Here, the theme view is used to drive the modelling and composition semantics for design using Theme/UML. To ensure that the developer carefully considers the order in which crosscutting themes are composed with base themes, Theme/UML allows only one crosscutting theme per composition. We therefore needed to inspect the crosscutting relationships to determine the order of binding. For this we used the clipped action view. In this view, the themes are positioned hierarchically, based on whether they crosscut one another. Finally, we check the validity of the design choices we made. This approach supports effective aspect identification, requirements coverage, traceability, and scalability of action views. However, this approach is only applicable for structured requirements document. As for the developers, they must possess the domain knowledge. Hence they must go through the whole requirements source document to identify the crosscutting concerns. They have to manually map the relationship between the themes and requirements. It is costly and time consuming to handle large amount of requirement sources.

Whittle et al. [28] proposed an approach to model scenario-based requirements using aspect-oriented paradigm. The main focus was on representing aspects during use case modelling. The approach provides a way to describe aspectual and non-aspectual scenarios independently and then merge them together to validate the complete set of scenarios. Aspectual scenarios, i.e., scenarios that crosscut other scenarios, are modeled as interaction pattern specifications (IPSs) and non-aspectual scenarios are modeled as UML sequence diagrams. Each aspectual and non-aspectual scenario is then translated into a set of state machines. The next stage of the process composes the aspectual and non-aspectual state machine for each entity. The result is an executable set of state machines that completely describe the requirements and in which aspectual and non-aspectual behaviour has been merged. Validation of these state machines can now take place using either a simulation harness or a code generator and the results can be feedback into the overall process. Composing (or weaving) aspectual and non-aspectual state machines helps the requirements engineer grasp the full picture. The approach supports better modularization and traceability but it lacks to address scalability. The developer must provide binding statements for each aspect and for each scenario that the aspect crosscuts. It also not proposed any systematic technique to aspectual scenarios identification.

Jacobson et al. [29] proposed an approach called aspect-oriented software development with use cases to handle crosscutting concerns. This approach is an extension to the traditional Use Case approach proposed by the same author and introduced new concepts like use case slices, extension use case, and pointcut. Here, a system is built use case by use case. The process begins with identifying use cases. Further, we need to specify each use case, to analyze it, and to design use cases in terms of use case slices and use case modules. Use case slices are used to employ aspects. Use case modules are used to contain the specifics of a use case over all system models. Extension use cases are the special kind of use cases that contain additional functionality of the use case. We implement and compose them using a composition mechanism to weave them at loading time, compilation time, or run-time to form a complete system. The approach includes processes like identifying, specifying, analyzing, designing, and implementing use cases. The approach strongly related to UML; but lacks in conflicts handling.

Moreira et al. [30] proposed an approach called concern-oriented requirements engineering (CORE), which treats each concerns uniformly. Here, a concern is any coherent collection of requirements. They also not classified concerns into viewpoints, use cases or aspects though their concerns encapsulate the coherent sets of functional and non-functional requirements. Concern space at the requirements level is represented as a hypercube. Each face of the hypercube represents a particular concern of interest. The process begins with identifying and specifying concerns using existing requirements elicitation mechanisms such as such as viewpoints, use cases, and goals. The identified concerns are specified using well-defined templates. The second step is to identify coarse-grained relationships among concerns by relating concerns to each other through a matrix. These relationships are identified using techniques such as domain analysis, ethnography, and natural language processing. The third step is to specify the possible projections of each concern on other concerns, which is achieved through composition rules. The fourth step is to identify and resolve conflicts (if any) among the concerns. This is achieved by building a contribution matrix, where each concern may contribute positively or negatively to the others. Prioritisation mechanism is used to solve conflicts and helping negotiation and decision-making. The last activity is to identify the dimensions of concerns. There are two dimensions of a concern at requirements level, which are mapping and influence. This approach supports multi-dimensional separation of concerns, which treats each concerns uniformly, hence, avoiding the dominant decomposing. It also establishes early trade-offs and solve conflicts that help negotiation and decision-making. But, it does not focus on the exact kind of relationships between two concerns, validation of proposed model with more case studies, and setting the concern specific actions and operators.

Araujo et al. [31] proposed an approach that incorporates aspect-oriented concepts into agile software development at requirements level. Agile software development aims at fast communication and incremental delivering of software artefacts. The aspect-oriented agile requirements approach focuses on defining and modelling initial crosscutting requirements as scenarios. Scenarios are descriptions of desired or existing system behaviour. Scenarios are commonly used in requirements engineering because they are easily understood by all stakeholders. Scenarios may crosscut other scenarios. Crosscutting scenarios are called aspectual scenarios. First, we begin with identifying main functionalities and refine those using scenarios. Secondly, we need to identify aspectual scenarios. This is achieved by analyzing all scenarios and observing some behaviour that crosscut several scenarios. Third, we need to compose aspectual and non-aspectual scenario definitions. Compositions are specified through simple rules. Finally, by analyzing the composition rules we may find ambiguities, errors omissions and conflicts in our scenarios. This analysis can be realized through inspections. Also, the participation of the stakeholders is crucial. Conflict identification can be accomplished by adapting the existing mechanism. The final set of aspectual and non aspectual scenarios plus composition rules are used to implement users' functionalities. The approach lacks on complete composition mechanism, conflict identification mechanism, validation, and tool support to guarantee that the approach will be used in an agile fashion.

Brito et al. [32] proposed an integrated approach for aspectual requirements to handle separation, modularization, representation and composition of concerns. The approach defines three main tasks, each one divided in several subtasks. The first task is identifying concerns, which aims at identifying the concerns of a system. A concern can be defined as a set of coherent requirements that the future system must have. This can be accomplished by analyzing the initial requirements, transcripts of stakeholders' interviews, etc. The second task is specifying concerns, which consists of many subtasks such as collecting information about concerns, specifying them using a template, and to design models like UML use case, interaction and class diagrams. The final task is composing concerns incrementally until the whole system is obtained. Each composition takes place in a match point in the form of a composition rule. A match point tells us which concerns (crosscutting or non-crosscutting) should be composed together. A composition rule shows how a set of concerns can be weaved together by means of some pre-defined operators. In order to accomplish this, we need to identify crosscutting concerns. The approach defined the main concepts as an extension of the UML metamodel, which allows a developer to better capture, analyze and understand the approach. The tool facilitates the specification of concerns, identification of crosscutting concerns, generation of the match point table and definition of composition rules. The approach does not define any method and tool with a reference model to support forward and backward traceability.

Z. Jingjun et al. [33] proposed aspect-oriented requirements modelling aiming to apply AOP paradigm at requirements engineering stages of software development. This approach supports separation of concerns both functional and non-functional, and modelling them in UML with class diagrams and state-chart diagrams respectively. The process includes five activities as follow: identifying and specifying concerns, analyzing concerns, composing concerns, weaving concerns, and simulating and validating requirements. First, identify both functional and non-functional concerns from system requirements, and then specify them in UML as OOP class and aspect class. Second, analyze the relation among concerns by detecting and removing the conflicts if any. Third, compose concerns by describing the static structure of the system. Next, during weaving concerns, the whole state-chart diagram of the system is given, and then finishes the weave process. Finally, simulate the system function with the whole state chart, and validate the function whether it meets the system requirements or not. If not, return to the first activity, identify and specify concerns again. Or, complete the model process. This model supports separation and modelling of concerns. It also supports an effective method to solve the mismatch among the aspects, which reduces the complexity of the system and increases software's reusability and maintainability. It uses terms functional and non-functional concerns as core and crosscutting concerns respectively. But, a crosscutting concern may be functional as well as non-functional. So, this method does not clearly identify and specify crosscutting concerns which are functional.

Chitchyan et al. [34] proposed Requirements Description Language (RDL), which is a symmetric AORE approach. It modularizes the requirements in symmetric fashion and represents them using the same abstraction, i.e., a Concern, to represent both crosscutting and non-crosscutting elements. A concern may be simple i.e. containing only requirements or composite i.e. containing requirements and other concerns. Both concern and requirement can be described as multi-sentence elements; where an element can be a subject or an object or relationship. A subject is described as an entity that undertakes actions and in RDL; it corresponds to the grammatical subject in the clause. An object is described as an entity that is affected by the actions undertaken by the subject of the sentence. In RDL, it corresponds to the grammatical object in the clause. A relationship is described as the action performed by the subject on or with regards to its object(s) and can be expressed by any the verbs or verb phrases in natural language. The main semantic load is carried by subject-relationship-object structure. The subject and object denote the entities of significance in it, whereas, the interactions between these entities are reflected by relationships. In this approach the relationship denotes the most central function, as it defines the functionality and/or properties that the subjects and objects provide. When specifying requirements stakeholders often qualify how important or significant a specific functionality or property is to them. In the RDL such qualifications are represented by the Degree element. Degree element depicts the strength of the relationship between the subject and object. The RDL elements discussed above are used for requirements description. The next activity is element composition. The assembling of separately defined requirements modules aiming to ensure their desired interactions or addressing undesired ones is termed as composition. Here, three sub-elements of a composition element are constraint, base, and outcome. A Constraint element specifies checks and restrictions applied on a set of requirements, and the action taken in imposing these constraints. Base element provides a query for selecting the set of requirements that are affected by some constraints; and the temporal or conditional dependency between these requirements and the constraints. The Outcome element defines how to treat the imposition of constraints upon the base sets of requirements. Composition specifications are written based on these semantics rather than requirements syntax, hence providing improved means for expressing the intentionality of the composition, in turn facilitating semantics-based reasoning about aspect influences and trade-offs. However, the approach requires a complete and precise requirements document, which can't be expected before requirements elicitation and analysis. Also, to validate this approach, it is still required to refine requirements, identify and resolve conflicts and trade-offs.

Jing Zhang et al. [35] proposed an aspect-oriented approach to supporting separation of crosscutting concerns in activity modelling. Aspect-specific constructs have been introduced as an extension to the activity models. Activity modeling describes the behavioural aspects of a system and is used to define a computational process as control flow and data flow among its constituent actions. It consists of many kinds of nodes and edges. The sequencing of actions is controlled by control flow and object flow edges. An activity node can be an action, an object node or a control node. An object node holds data that flow through the activity model. Control nodes are responsible for routing control and data flows in an activity. Activities can be divided into different partitions that represent different kinds of activity groups for identifying actions that have some characteristics in common. Activity specifications grow with increasing complexity of the system and require lifecycle maintenance for the concerns that crosscut different activity modules. Aspect-oriented paradigm provides a solution to above problem by encapsulating crosscutting concerns in specialized units called aspects. New concepts like joinpoints, pointcut, and advice are introduced here. A joinpoint specifies "where" the crosscutting concern emerges in the activity model. A pointcut is defined as a special construct containing a group of particular join points, which defines a pattern to identify matching join points. In activity modeling, the concern behaviour is implemented using an activity model referenced by a special action called advice, which specifies "what" (i.e., the behaviour) makes up the crosscutting concern. This paper applies an aspect-oriented approach to supporting separation of crosscutting concerns in activity modeling. Aspect-specific constructs have been introduced as an extension to the activity models. The current implementation of the pointcut specification only allows join point to be referred to action nodes. The approach lacks on covering other kinds of activity nodes and investigating more advanced pointcut selection patterns.

Budwell and Mitropoulos [36] proposed a methodology called Structured Lexicon for Aspectual Identification (SLAI) that is based upon the Language Extended Lexicon (LEL) for capturing requirements. LEL is used for vocabulary acquisition or understanding problem language. LEL consists of three elements: signs, notions, and behavioural responses. Signs represent any word or phrase that has a special meaning within the Universe of Discourse (UofD). Notions and behavioural responses help to define the meaning of each sign. Notions define the signs in the context of the UofD, whereas behavioural responses define how the sign is used. The principle of circularity and minimal vocabulary is used to define signs. The process begins with clearly identifying and defining aspects at the requirements phase of software development. An aspect is a crosscutting concern or matter of interest in a software system. A concern must be derived from either functional or non-functional requirements. The next step is to capture crosscutting requirements (functional or non-functional) as aspects. Functional requirements can be identified as aspects in the form of UML use cases. A functional aspect is defined as a crosscuts use case. This is a new type of use case in UML after modifying extends and includes use cases. Non-functional requirements are converted into operationalizations and then into use cases. This methodology is divided into two flows, functional requirements flow and non-functional requirements flow. In functional requirements flow, actors of the system are identified first. Next, use cases are identified and detailed them into use case steps. Later, the use case steps are recorded in the SLAI. In non-functional requirements flow, SIG graph is used where the roots of the SIG graph represent non-functional requirements and leaf nodes for each path represent operationalizations, which are either operations or design constraints. Since design constraints tend to be considerations that need to be addressed in the design phase, they are not included in the SLAI. The use case information is not enough to classify crosscuts use cases as aspects. Thus, for each crosscuts use case, a table is generated that lists each use case with which it interfaces. For each of these use cases listed, the condition of the extension, the composition rule operator, and the affected point are defined. The approach supports the identification of aspectual or crosscutting use cases from both functional and non-functional requirements as well as systematically identifying both aspects and potential aspects by the use of a limited set of vocabulary for the terms defining the requirements of the system. From the study, it is observed that the methodology lacks on aspect composition as well as conflict resolution.

Ali and Kashirun [37] developed a model to identify crosscutting concern and designed an automated tool based on the model. The model is based on the proposed approach which is adapted from Theme/Doc and Early Aspects Identification approaches. The model is fully automated and involves non-collaborative processes unlike conventional requirements engineering processes, which are generally collaborative and iterative in nature. The execution of processes in this model is sequential where each process requires output of the previous process as input for execution. The model consists of processes like structuring requirements, removing redundancy, part-of-speech (pos) analysis, semantic analysis, filter verbs identification, map relationship view, refining the relationship view, identifying dominating verb, and modelling crosscutting influences. Structuring requirements task involves numbering all the requirements agreed by the stakeholders to identify and manipulate each requirement uniquely in the next stages. Sometimes, same requirements are specified many times by different stakeholders. The redundant requirements are eliminated during removing redundancy process. Verbs are extracted from each requirement during POS analysis and they will be used for modelling the relation with the requirements and interdependency among other verbs. Semantic analysis task utilizes semantic tagger to analyse the context of the phrase in which the verb is used. This information is used to identify verbs used to describe similar requirements. Based on the semantic analysis performed, duplication of the verbs in terms of the context is discarded during filter verbs identification process. Next, we map the requirements using a matrix during map relationship view process. The requirements shared by more than one verb and the scattered verbs are identified based on the relationship view. Finally, identify the dominating verb in the requirement, which are the candidate aspect and model them to identify the crosscutting concern using Action View Model as used in Theme/Doc approach. This paper described a tool that provides automated support for crosscutting concern identification at the requirements level. The tool utilises natural language processing technique to reason about properties of the concerns and model their structure and relationship. But, this model lacks on conflict resolution and implementation and validation of the tool and tests it with case studies.

G. Mussbacher [38] proposed an approach Aspect-oriented User Requirements Notation (AoURN) that extends the User Requirements Notation (URN) with aspects. URN contains two complementary modelling languages Goal-oriented Requirement Language (GRL) and Use Case Maps (UCMs) for goals and scenarios respectively. GRL is a visual modelling notation and supports reasoning about goals and non-functional requirements (NFRs). UCM is a visual scenario notation that supports the definition of scenarios. A scenario describes a specific path through the UCM model where only one alternative at any choice point is taken. Given a scenario definition, a traversal mechanism can highlight the scenario path or transform the scenario into a message sequence chart (MSC). AoURN extends the URN by defining a joinpoint model for the GRL and UCMs. All nodes of GRL graphs or UCMs, which are optional to an actor or a component, are considered as joinpoints. Pointcut expressions are used for matching joinpoints in AoURN models to identify any URN node, and are defined on pointcut diagrams. Pointcut diagrams are standard URN diagrams that allow requirements engineer to increase matching power by using wildcards (“*”) and logical expressions (containing “and”, “or”, and “not”). This approach extends URN with aspects and thus unifies goal-oriented, scenario-based, and aspect-oriented concepts in one framework. Minimal changes to URN ensure that requirements engineers can continue working with goal and scenario models expressed in a familiar notation. At the same time, concerns in goal and scenario models, regardless of whether these concerns crosscut or not, can be managed across model types. But, it uses flexible composition rules that are only limited by the expressiveness of URN itself.

Iqbal and Allen [39] suggest a process modeling approach that represents aspect from the initialization of software to its implementation. It suggests the identification of aspects in the Use Case Model and Sequence Diagrams of the system. Use cases which involve multiple use cases like included or extended use cases may be considered as candidate aspects since they have the probability of crosscutting representation in design as well as in implementation. Similarly, the objects which have communication with multiple objects and which are represented in multiple sequence diagrams may also be regarded as candidate aspects. Proper specification of the candidate aspects can help identifying actual aspects. In this approach, it is not mentioned how to identify aspects. Also, it lacks on implementing the model and validation it with some case studies.

Hamza and Darwish [40] proposed a new approach to identify and model candidate aspects from functional and non-functional requirements of the system, and propagate them to the design phase. The approach needs problem definition as input and produces EBT-NFR analysis model as output. The EBT-NFR model identifies crosscutting NFR and visually shows how they are scattered across the various modules in the system. The process begins with performing requirements and domain analysis. The outcomes of this step are lists of: functional requirements (FRs): non-functional requirements (NFRs), Enduring Business Themes (EBTs), Business Objects (BOs), and Industrial Objects (IOs). The next step is to develop and document main use cases in the system using identified FRs, EBTs, BOs, and IOs. After identifying the main use cases, NFR Matching is performed where each identified NFR in the system is matched with a set of use cases. The next step is Concept Analysis step where, concepts within the problem are discovered using use cases, EBTs, BOs, and IOs. Here, classify the EBTs as concept EBTs (i.e form a formal concept) and none-concept EBT. Next, we need to identify candidate aspects by establishing and understanding the relationships between different NFRs and EBTs in the system. This is achieved by classifying NFRs, and then developing an EBT-NFR model that shows the relationships between EBTs and NFRs. An NFR can be classified into one of three types: Localized-NFR (L-NFR), Distributed-NFR (D-NFR), or an Aspect-EBT (A-EBT). To classify a NFR, calculate Coupling Factor (CF), which measures the coupling between a NFR and each EBT in the system. Finally, develop the EBT-NFR analysis model that visualizes the interaction between the different NFRs and EBT modules in the system and can be used as a link to the design phase. The approach uses Formal Concept Analysis and EBTs that can be identified using Software Stability Model to understand the interaction between NFRs and FRs, and to identify possible aspects in early stages of the development. Further, the proposed approach is the only approach that identifies crosscutting NFRs with respect to the structural nature of the system. The approach lacks a tool to partially automate the proposed approach and applicability of the approach to several case-studies to validate the results.

Liu et al. [41] proposed a use case and non-functional scenario template-based approach to identify aspects. This approach consists of a sequence of activities. The process begins with identifying and defining actors and use cases and building an initial use case model. The next activity is to refine use case model by identifying extensions and inclusions of the use cases. After this, describe NFRs at key association points in the use case model. Association points are of many types as NFRs Association Points, which are specific points of use cases where NFRs can be associated; use case association points associate NFRs to the described functionality, actor association points specify NFRs related to external entities, actor-use case association points represent NFRs related to interaction between external entities and a functionality, and system boundary association points define NFRs that are global in nature. A concern can be identified with a set of architectural policies, and each of these can be described using specific dimensions that specify with more details the NFRs in architectural terms in each use case. Hence, describe architectural policies at platform-independent level through Architectural Policy Scenarios. Finally, identify aspects, where an aspect is a function that influences more functions or more use cases. This approach is based on use cases and described and map non-functional requirements into function and architectures through non-functional scenario template. It not only improves modularity in the requirements which make it possible to begin tackling the problem of tangling, scattering of the requirement as early as in requirement analysis phrase, but also improves traceability from requirement analysis level to implement level, so it achieves a smooth transition between the system analysis and the design. It lacks on supporting the approach with formal method and applying it in more case studies and real systems.

3. A Roadmap to Research

Based on the above discussion, we present a roadmap to our research. The focus of our research work will be exactly on handling crosscutting concerns during requirements phase and to propose a new requirements engineering model for advanced separation of concerns using aspect-oriented concepts having the following challenges:

- Prevent the *tyranny of dominant decomposition* symptom.
- Improve the ability to identify, specify and compose both crosscutting and non-crosscutting concerns.
- Handle conflicts that could emerge when two or more crosscutting concerns affect each other.

4. Conclusion

Crosscutting concerns are the concerns which affect other concerns. These concerns often cannot be clearly decomposed from the rest of the system in both design and implementation and hence result in scattering, tangling, or both; that are difficult to understand and maintain. AOSD is used to identify and specify such crosscutting concerns in separate modules, known as aspects. This results in better support for modularization hence reducing development, maintenance and evolution costs.

AORE is a process that focuses on identifying, analyzing, specifying, verifying, and managing the crosscutting concerns at the early stages of software development. In this literature review, we have discussed many AORE approaches to deal with crosscutting concerns at early stages of software development. As compared with traditional approaches like use cases, viewpoints, and goals; AORE approaches are too young and still need to validate them with more case studies and avoid the tyranny of dominant decomposition, improve the ability for identifying both functional and non-functional crosscutting concerns, offer an automatic mechanism for specifying concern compositions, and to handle conflicts (if any). Some work has been done in this area, but the development of a complete methodology is needed. Without the formulation of this methodology, the full benefits of the aspect-oriented programming paradigm cannot be realized.

References

[1] P. Naur and B. Randell, “*Software Engineering: Report of the Working Conference on Software Engineering*”, Garmisch, Germany, October 1968. NATO Science Committee, 1969.
[2] The Standish Group. Chaos Report. Technical report, Standish Group International, 1995, http://www.it-

cortex.com/Stat_Failure_Rate.htm#The%20Chaos%20Report %20(1995).

[3] The Standish Group, "CHAOS Summary 2009", Technical report, Standish Group International, Boston, Massachusetts, April 23, 2009, http://www1.standishgroup.com/newsroom/chaos_2009.php

[4] Dijkstra, E., "A Discipline of Programming", 0-13-215871-X, Prentice-Hall, 1976.

[5] D. L. Parnas, "*A Technique for Software Module Specification with Examples*", Communications of the ACM (CACM), 15(5):330–336, 1972.

[6] D. L. Parnas, "*On the Criteria to be Used in Decomposing Systems into Modules*", Communications of the ACM (CACM), 15(12):1053–1058, 1972.

[7] Mark.E. et al., "*Do Crosscutting Concerns Cause Defects?*", IEEE Transactions On Software Engineering, Vol. 34, No. 4, July/August 2008.

[8] ACM, "*Special Issue on Aspect-Oriented Programming*", Communications of the ACM, 44 (10), 2001.

[9] Rashid, A., Moreira, A., Araújo, J., "*Modularization and Composition of Aspectual Requirements*", In 2nd Aspect-Oriented Software Development Conference (AOSD'03), Boston, USA, ACM Press. 11-20, 2003.

[10] Baniassad, E., Clements, P., Araújo, J., Moreira, A., Rashid, A., Tekinerdogan, B., "*Discovering Early Aspects*", IEEE Software Special Issue on Aspect-Oriented Programming. 23(1): 61-70, 2006.

[11] Suzuki, J., Yamamoto, Y., "*Extending UML with Aspects: Aspect Support in the Design Phase*", In Object-Oriented Technology Workshop at 13th European Conference on Object-Oriented Programming (ECOOP'99), Lisbon, Portugal, 1999.

[12] Clarke, S., Walker, R., "*Composition Patterns: An Approach to Designing Reusable Aspects*", In 23rd International Conference on Software Engineering, (ICSE'01), Ontario, Canada, ACM Press, 5-14, 2001.

[13] France, R., Ghosh, S., "*A UML-Based Pattern Specification Technique*", IEEE Transactions on Software Engineering, IEEE Computer Society, **30**(3): 193-207, 2004.

[14] Moreira, A., Araújo, J., Brito, I., "*Crosscutting Quality Attributes for Requirements Engineering*", In 14th Software Engineering and Knowledge Engineering Conference (SEKE'02), Ischia, Italy, ACM Press. 167 – 174, 2002.

[15] Sutton Jr, S., Rouvellou, I., "*Modeling of Software Concerns in Cosmos*", In 1st Aspect-Oriented Software Development Conference (AOSD'02), Enschede, Netherlands, ACM, 127-133, 2002.

[16] Rashid, A., Moreira, A., Araújo, J., "*Modularization and Composition of Aspectual Requirements*", In 2nd Aspect-Oriented Software Development Conference (AOSD'03), Boston, USA, ACM Press, 11-20, 2003.

[17] Rashid A. et al., "*Modularisation and Composition of Aspectual Requirements*", AOSD 2003, ACM, pp. 11-20, 2003.

[18] Nuseibeh, B., Easterbrook, S.: Requirements engineering: a roadmap. In: Proc. of the IEEE Int. Conf. on Soft. Eng. (ICSE), pp. 35–46 (2000)

[19] Parnas, D.L.: Software engineering programmes are not computer science programmes. Ann. Soft. Eng. 6(1), 19–37 (1999).

[20] Betty H.C. Cheng and Joanne M. Atlee, "*Current and Future Research Directions in Requirements Engineering*", Design Requirements Workshop, LNBIP 14, pp. 11–43, Springer-Verlag Berlin Heidelberg 2009.

[21] Jacobson, I., Chirsterson, M., Jonsson, P., Overgaard, G., "*Object-Oriented Software Engineering - a Use Case Driven Approach*". 978-0201544350. Addison-Wesley, 1992.

[22] Sommerville, I., Sawyer, P. (1997b), "*Requirements Engineering - A Good Practice Guide*", 978-0471974444. John Wiley, 1997.

[23] VanLamsweerde, A. "*Goal-Oriented Requirements Engineering: A Guided Tour"*, In 5th Requirements Engineering Conference (RE'01), Toronto, Canada, IEEE Computer Society. 249 – 262, 2001.

[24] M. Bruntink, A.V. Derusen, R.V. Engelen, T. Tourwe, "*On the Use of Clone Detection for Identifying Crosscutting Concern Code"*, IEEE Transactions on Software Engineering, Vol. 31, No. 10, October 2005.

[25] J. Grundy, "*Aspect-Oriented Requirements Engineering for Component-based Software Systems*", IEEE International Symposium on Requirements Engineering, IEEE CS, pp. 84-91, 1999.

[26] Rashid, A., Sawyer, P., Moreira, A., and Araújo, J. "*Early Aspects: a Model for Aspect-Oriented Requirements Engineering*", Proc. of Int. Conference on Requirements Engineering (RE'02), 2002.

[27] E. Baniassad, S. Clarke, "*Theme: An Approach for Aspect-Oriented Analysis and Design*", In Proceedings of the 26th Int. Conf. on Software Engineering (ICSE04), 2004.

[28] Araújo, J. Whittle, and D-K. Kim, "*Modeling And Composing Scenario-Based Requirements With Aspects*" In Proc. of the 12th IEEE International Requirements Engineering Conference (RE 04), 2004.

[29] Jacobson, I.," Aspect-Oriented Software Development with Use Cases", 978-0-321-26888-4, Addison-Wesley, 2004.

[30] A. Moreira, J. Araújo, A. Rashid, "*A Concern-Oriented Requirements Engineering Model*", *Proc. Conference on Advanced Information Systems Engineering*, Portugal, LNCS 3520, pp. 293 – 308, Springer-Verlag Berlin Heidelberg 2005.

[31] J. Araujo and J. C. Ribeiro "*Towards an Aspect-Oriented Agile Requirements Approach*", Proceedings of the 2005 Eighth International Workshop on Principles of Software Evolution (IWPSE'05), IEEE 2005.

[32] Isabel Sofia Brito and Ana Moreira, "*Towards an Integrated Approach for Aspectual Requirements*", 14th IEEE International Requirements Engineering Conference (RE'06), IEEE 2006.

[33] Zhang Jingjun, Li Furong, and Zhang Yang, "*Aspect-Oriented Requirements Modeling*", Proceeding of the 31st IEEE Software Engineering Workshop SEW-31 (SEW'07), Baltimore, MD, USA, 2007.

[34] Chitchyan, R., Rashid, A., Rayson, P.,Waters, R., "*Semantics-Based Composition for Aspect-Oriented Requirements Engineering*", In 6th Aspect-Oriented Software Development Conference (AOSD'07), Vancouver, Canada, ACM Press. 36-48, 2007.

[35] Jing Zhang, Yan Liu, Michael Jiang, and John Strassner, "*An Aspect-Oriented Approach to Handling Crosscutting Concerns in Activity Modeling*", Proceedings of the International MultiConference of Engineers and Computer Scientists 2008 Vol I, IMECS 2008, Hong Kong, 19-21 March, 2008.

[36] Caroline C. Budwell and Frank J. Mitropoulos, "*The SLAI Methodology: An Aspect-Oriented Requirement Identification Process*", 2008 International Conference on Computer Science and Software Engineering, pp. 296-301, 978-0-7695-3336-0/08, IEEE 2008.
[37] Busyairah Syd Ali and Zarinah Mohd. Kasirun,"*Developing Tool for Crosscutting Concern Identification using NLP*", IEEE 2008.
[38] G. Mussbacher, "*Aspect-Oriented User Requirements Notation: Aspects in Goal and Scenario Models*", MoDELS 2007 Workshops, LNCS 5002, pp. 305–316, 2008, Springer-Verlag Berlin Heidelberg 2008.
[39] S. Iqbal, and G. Allen, "*Representing Aspects in Design*", presented at 2009 Third IEEE International Symposium on Theoretical Aspects of Software Engineering, Tianjin, China , 2009.
[40] S. Hamza and D. Darwish, "*On the Discovery of Candidate Aspects in Software Requirements*", Proc. Of Sixth International Conference on Information Technology: New Generations, 2009.
[41] Xiaojuan Zheng, Xiaomei Liu, and shulin Liu, "*Use case And Non-functional Scenario Template-Based Approach to Identify Aspects*", Second International Conference on Computer Engineering and Applications, 2010.

Narender Singh is an Assistant Professor in Maharishi Markandeshwar University, Mullana-Ambala, India and pursuing his PhD degree from Department of Computer Science & Applications, Maharshi Dayanand University, Rohtak, India. He has earned his MCA and M.Phil degrees from Department of Computer Science & Applications, Kurukshetra University, Kurukshetra, India in 2006 and 2008 respectively. He has published more than 5 research papers in national & international journals & conference proceedings. His research interest includes Aspect-Oriented Requirements Engineering and Software Product Lines.

Nasib Singh Gill is Professor and Head, Department of Computer Science & Applications, Maharshi Dayanand University, Rohtak, India. He has earned his Doctorate in Computer Science in the year 1996 under the supervision of a renowned academician and researcher – Prof. P.S. Grover of Delhi University and carried out his Post-Doctoral research at Brunel University, West London during 2001-2002. He has received *Commonwealth Fellowship Award* of British Government for the Year 2001. He has published more than 145 research papers in national & international journals, conference proceedings, bulletins, books, and newspapers. He has authored three popular books, namely, *'Software Engineering', 'Digital Design and Computer Organisation' and 'Essentials of Computer and Network Technology'*. He is a fellow of several professional bodies including IETE (The Institution of Electronics and Telecommunication Engineers). He has been awarded with *'Best Paper Award'* by Computer Society of India in the year 1994 for contributing the best paper *"A New Program Complexity Measure"* in their Journal. He is presently guiding researchers in the areas - *Measurement of Component-based Systems, Complexity of Software Systems, Component-based Testing, Data mining & Data warehousing, Aspect-Oriented Software Development, and NLP.*

A Cryptographic Algorithm based on Bilinear Transformation

Phani krishna kishore M [1] and Venugopal IVS [2]

[1] **Professor, GVP College of Engineering, Department of IT**
Visakhapatnam, Andhra Pradesh, 530048, India

[2] **Asst. Professor, GVP College of Engineering, Department of IT**
Visakhapatnam, Andhra Pradesh, 530048, India

Abstract

The technique of producing ciphers via transformations from one domain to another domain has been widely studied. In this paper a new cryptographic algorithm is presented based on complex bilinear transformations, the strength and performance of the algorithm is analyzed.

Keywords: Cryptography, Complex Bilinear Transformations, Encryption, Decryption.

1. Introduction

Security has become an important aspect in the digital world with the advent of technological advancements. Though the art of designing new methods and algorithms for secure transmission of data dates back to centuries it emerged as a separate branch of technology during the last few years. Several researchers devised methods for encryption and decryption based on mathematical functions.

The Conventional cryptography uses symmetric key algorithms in which a common key is shared between the sender, receiver and operates block wise. Most of the modern symmetric ciphers operate at bit level and largely designed around fiestel structure. DES, AES, BLOWFISH, IDEA, TWO FISH are some of the popular algorithms. The paradigm shift brought by the RSA algorithm towards the public key cryptosystem has changed the scenario of security systems.

For several years the rich theory of numbers has dominated this domain. For the past few years several algorithms were developed using algebraic theory and other fields of mathematics.

However not much work is reported on methods using complex arithmetic. Recently Dimitrov et al, [4] in their work discussed some algorithms for multi exponentiations based on complex arithmetic.

Elsayed Mohammad et al described the elliptic curve cryptography over Gaussian integers [5].

Mohammad Ahmad Alia et al presented a new digital signature scheme based on Mandebrot and Julia fracal sets [8]. In this paper a new symmetric key algorithm that operates block wise based on complex bilinear transformation is proposed.

2. Bilinear Transformation

A mapping f: C → C, (C denotes complex plane) of the form w= f (z) = $\frac{az+b}{cz+d}$ (ad-bc≠0) where a, b, c and d ϵ C, are constants and z=x+iy, w=u+iv be complex variables in different planes, is called a bilinear transformation from C to C. In fact several transformations can be considered from C to C, however this transformation is a one-to-one mapping from C to C. A bilinear mapping is called conformal if it preserves the angles and orientations between the curves under the transformation. A bilinear transformation is actually a combination of translation, dilation, rotation and inversion.

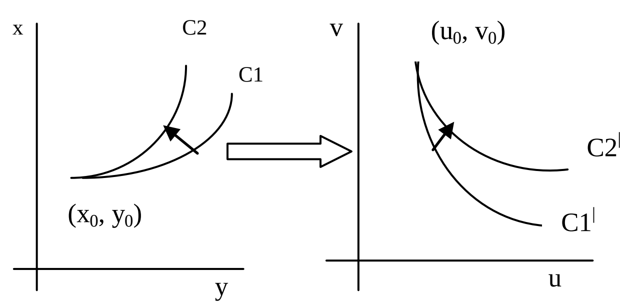


Fig. 1 Bilinear Transformation from one curve to another.

These transformations are widely studied in the theory of complex variables and got many applications in several branches of engineering.

3. Proposed System

In this proposed system data is encoded onto a complex plane and then using a bilinear transformation the data points are transformed into a another set of points in another complex plane producing the cipher text. The transformation being reversible the original data can be obtained using inverse transformation. For the application of the transformation a, b, c and d (constants) act as key. Several methods can be adopted to exchange these keys. However a new system is developed to identify the a, b, c and d based on a preshared information, which actually prevents from transmission of a, b, c and d. Initially sender and receiver will agree on a common text message. A hash code is developed from the text message from a randomly chosen part of the text which is used in the construction of the key.

A common hash table is also shared that consists of information to retrieve information regarding computation of key.

The encryption algorithm makes use of pseudorandom numbers to select the index for the starting point of shared text and to select indicator from hash table to generate the key. The encryption algorithm generates a random number within a specified range that acts as an index for the selection of position of the shared text. The selected text is transformed into 128-bit hash code by MD5.Again another pseudorandom number is generated to identify a splitting criteria from the hash table, the 128 bit hash code is separated into four parts by using the split values obtained from the hash table, that are converted to integers to form a, b, c and d.

The text to be encrypted is processed in blocks of seven characters each. Each block is converted into equal parts of binary and then into decimal x, y to form x+iy(=z) (zeros are appended if necessary).

By using a, b, c and d w= $\frac{az+b}{cz+d}$(=u+iv) is computed.

Then w together with the pseudorandom numbers used for the block are converted into text form that forms the cipher text. A predefined protocol for embedding the pseudorandom numbers can be used. (Eg: 128 bits after first 10 bits or at the end of w etc;)

On receipt of cipher text, receiver separates pseudorandom numbers and u+iv. Receiver computes a, b, c and d by using the pre shared information and pseudo random numbers.

By using z= $\frac{b-wd}{wc-a}$, the receiver gains the text message from x+iy.

4. Algorithms

4.1 Generation of keys(on sender side)

Input: Common text message, hash table.
Output: values of a, b, c and d.
Step 1: Generate pseudorandom number(r_1) ($0<r_1<$length(shared text))
Step 2: Starting at r_1^{th} character the shared text is converted into binary and 128 bit binary code is generated using MD5.
Step 3: Generate pseudorandom number(r_2) ($0<r_2<N$) (N denote the maximum number of splits stored in hash table).
Step 4: Select a particular split values to generate a, b, c and d from hash table using r_2. (Number of bits to compute a, b, c and d respectively).
Step 5: Divide 128 bits obtained in step 2 into four parts and convert the bits into decimal form as per values obtained in step 4.

4.2 Algorithm for encryption

Input: Plain text
Output: Cipher text
Step 1: Split the plain text into block of seven characters each and convert into binary form (zeros are appended to binary form to make it a multiple of 2).
Step 2: Compute x, y by splitting binary form obtained in step 1 into two equal to from z=x+iy.
Step 3: compute w= $\frac{az+b}{cz+d}$ (=u+iv)
Step 4: Convert w, r_1, r_2 into text to from cipher text.
Step 5: Repeat steps 1 to 4 until entire text is converted

4.3 Algorithm to generate keys (on receiver side)

Input: common shared message, hash table, cipher text.
Output : a, b, c and d.
Step 1: From the block of cipher text retrieve r_1, r_2.
Step 2: use r_1 to generate 128 bit binary code from common text.
Step 3: using r_2 find the split for 128 bit code from hash table.
Step 4: compute a, b, c and d.

4. 4 Algorithm for decryption

Input: cipher text
Output: original text
Step 1: Convert each cipher text block into decimal equivalent to from w=u+iv.

Step 2: compute z= $\frac{b-wd}{wc-a}$ (=x+iy)

Step 3: Convert x, y into binary form and hence into text form.
Step 4: Repeat the steps 1 to 3 until entire message is decrypted.

5. Example

Consider the plain text: (Empty space is also considered as a character)
"Cryptography is used to encrypt and decrypt data"
The plain text is separated into blocks of seven characters each and corresponding value of z(=x+iy) are given by
Cryptog = 8899535.0 + 1914855.0i
raphy i= 1.5042438E7 + 1.8763881E7i
s used = 1.5106991E7 + 7959072.0i
to encr= 1.5318278E7 + 1603428.0i
ypt and= 1.597533E7+ 16034328i
decryp= 4257518.0 +8174832.0i
t data =1.5237926E7 + 4010144.0i
The pseudo random number generated from common message is 4. Hence plain text is converted into 128 bit binary from "ptography is used to encrypt and decrypt data"
The 128-bit hash code is
00001110001001011011001001001111111110000000111
1110001111101100100010101000011001101011000100
0110111111010100010001001100010001100
The split for a, b, c and d for first 21 characters is 32, 28, 24, 44. The decimal equivalents are
a(237351503),b(260111485),c(9523405),d(6.7494952499 9E+12)
The split for b, c, a and d for next 21 characters is 28, 24, 32, 44. The decimal equivalents are
b(14834468),c(16744700),a(2106675405),d(6.749495249 99E+12)
The split for c, a, b and d for remaining characters is 36,40,24,28. The decimal equivalents are
c(3797624063),a(553991901520),b(13460023),d(2225797 82).
W= $\frac{az+b}{cz+d}$ for each block is
Cryptog=0.005514710810825561 -1.060427362652612E-9i
raphy i=0.005514707075943306 -1.4888801636146341E-9i
s used=0.005514708260135056 -1.2527272333429005E-9i
to encr=0.33331102880924857 +1.790851616570252E-5i
ypt and=0.33329850919102816 +3.4948303171133865E-6i
decryp= 0.3333050105270643 +5.385138305935736E-5i
t data = 0.3333050105270643 +5.385138305935736E-5i
w,r_1,r_2 for each block is converted into text form which is given by
As{qvnescqjx"hq!wrge"um!goas{qv!cof!fdas{qv!f`v`
and is sent to receiver.
The receiver receives cipher text as:
As{qvnescqjx"hq!wrge"um!goas{qv!cof!fdas{qv!f`v`
The receiver converts them into equivalent float equivalents.
Receiver has information regarding ciphers(w) as
0.005514710810825561-1.060427362652612E-9i
0.005514707075943306-.4888801636146341E-9i
0.005514708260135056 -1.2527272333429005E-9i
0.33331102880924857 +1.790851616570252E-5i
0.33329850919102816 +3.4948303171133865E-6i
0.3333050105270643 +5.385138305935736E-5i
0.3333050105270643 +5.385138305935736E-5i
The receiver computes a,b, c and d values from algorithm used to generate the key by decryption algorithm.
Hence receiver computes plain text z= $\frac{b-wd}{wc-a}$ and converts it into string equivalent.
The receiver then gets plain text z as
8899535.0 + 1914855.0i 1.5042438E7 + 1.8763881E7i 1.5106991E7 + 7959072.0i 1.5318278E7 + 1603428.0i 1.597533E7+ 16034328i 4257518.0 +8174832.0i 1.5237926E7 + 4010144.0i
The receiver transfers above into string equivalent which forms the plain text:
"Cryptography is used to encrypt and decrypt data"

6. Cryptanalysis

The strength of algorithm is analyzed in cryptanalysis. Several techniques are proposed in literature namely, linear cryptanalysis, differential cryptanalysis etc. The present method produces cipher text using a mathematical function and operates on blocks of text and hence the following analysis is made, given the algorithm is known.

6.1 Cipher text only

In this case if the attacker knows only the cipher text since there are infinitely many combinations of a, b, c and d it is infeasible to guess a, b, c and d for a text block. Even if a, b, c and d are guessed they are limited to 21 characters.

6.2 Cipher text and corresponding plain text

The knowledge of shared message and cipher text may reveal the 128 bit key, if four blocks of plain text and the corresponding cipher text (z_1, w_1), (z_2, w_2), (z_3, w_3), (z_4, w_4) are known. Then by solving four equations a, b, c and d can be found. However since in this method a, b, c and d

are chosen for every three pairs and hence even if (z_1, w_1), (z_2, w_2), (z_3, w_3) are known then three equations for four unknowns are obtained, its difficult to guess a, b, c and d.
While transmission of cipher text the indices r_1, r_2 are embedded. They can be embedded in variety of ways with a mutual agreement. Even if they are revealed, since the common message and hash table are not known they are not useful to attacker.

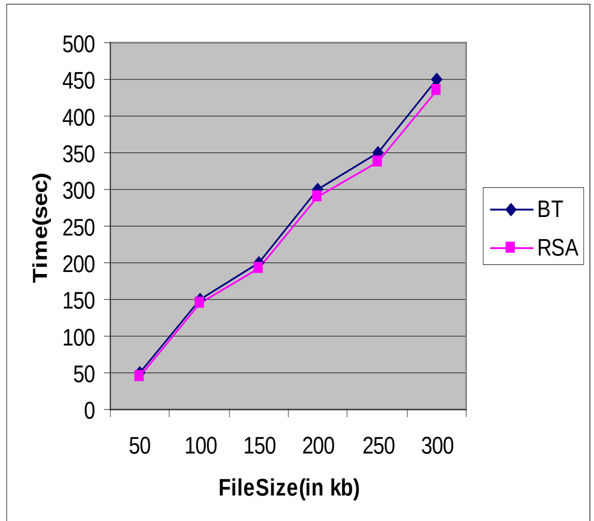
6.3 Cipher text, corresponding plain text and key

While decrypting the cipher text even if the attacker grabs the 128 bit key used to generate a, b, c and d there are ${}^{127}c_4$ possible ways of splitting the 128 bits into blocks and since again a, b, c and d can be arranged in 4! Ways altogether ${}^{127}c_4 * 4! = 1488186000$ possible ways to split 128 bits to generate a, b, c and d and this has to be carried out for every 21 characters of text.
If a, b, c and d values are taken as complex numbers then if the attacker grabs the 128 bit key used to generate a, b, c, and d there are ${}^{127}c_8$ possible ways of splitting the 128 bits into blocks and since again complex numbers a, b, c and d(4 real parts and 4 imaginary parts) can be arranged in 8! Ways altogether
${}^{127}c_8 * 8! = 2728665444597882048000$ ($>2^{71}$) possible ways to split 128 bits to generate a, b, c and d and this has to be carried out for every 21 characters of text, which further increases the complexity and security of the system.
Again, 128 bits in hash table can be rearranged in 2^{128} ways, which correspondingly changes values of a, b, c and d. Hence total number of combinations now become $2^{128} * 2^{71} = 2^{199}$.
The time that is required for computation of $w = \frac{az+b}{cz+d}$ is 0.12 milliseconds. Hence 2^{199} combinations would require $2^{199} * 0.12 = 9.642e+58$ milliseconds = 3.056e+48 years.
If the attacker is in possession of hash table and cipher text then has to try 2^{199} combinations by brute force which takes approximately 3.056e + 48 years run on an Intel Pentium-IV 2.4 Ghz Processor.
Even if an malicious user requests to communicate with a valid user exchanges the secret message and hash table, by using different texts and different hash tables for different users this kind of attack can be avoided.
Most of block ciphers with symmetric keys operate at bit level and if once the keys are broken entire message is revealed. In this case since each pair of parameters that are used to generate keys are useful for only 21 characters and hence this method is much stronger.

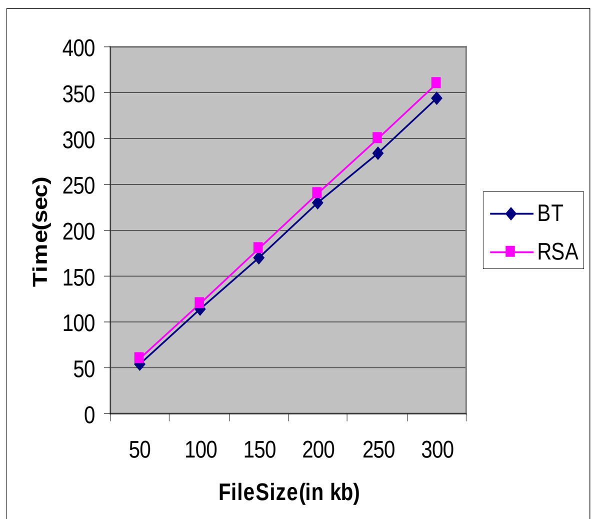
7. Experimental Results

This method based on mathematical operations, the performance of the method is compared with the much popular public key algorithm RSA.

The time taken for encryption and decryption are observed for different file sizes.



Encryption



Decryption

Fig. 2 Comparison between Bilinear Transformation and RSA Algorithms

It is observed that the proposed method takes almost same time to encrypt or decrypt as file size increases. Since the two parameters are embedded into cipher text for every 21 characters the size of the encrypted file will increase in size marginally, however by using efficient data structures like linked lists, each block may contain much larger plain text there by making the increase of encrypted file negligible. The data type 'double' is used to store the values of x, y which restricts the block size to be seven characters. However block size can be increased by using more efficient data types.

8. Usage of the method

This method can be used in different ways. This method can be used in network security applications. Apart from the direct application, the theme can be used in conjunction with any method that involves encoding the text onto a curve or a two dimensional plane. First in case of Elliptic curve cryptography the plain text is encoded onto an elliptic curve. The key used in elliptic curve cryptography can be used to generate the a, b, c and d in bilinear transformation and thereby bilinear transformation can be applied on the encoded points of the elliptic curve.
The images under bilinear transformation form the cipher text thereby providing a double encryption with the same key and some additional preshared information regarding the breakup of key into a, b, c and d.
Secondly, once the text is encoded onto the complex plane, bilinear transformations can be repeatedly applied on the points possibly with different sets of values of a, b, c and d in each round there by increasing the security.

9. Conclusion

In this paper a new cryptographic algorithm has been proposed based on the concept of bilinear transformation from one complex plane to another plane, with additional features the new system is observed to provide more security when compared to existing block ciphers and performance is compared with popular public key algorithm RSA.
The work can be extended to several other complex transformations which are great cryptographic potential.

References

[1] Bruce Schneier, "cryptography engineering", John Wiley & Sons,March15,2010 ISBN: 9780470474242

[2] Christophe De Canniere, Alex Biryukov, and Bart Preneel, "*An Introduction to Block Cipher Cryptanalysis*", proceedings of the IEEE, VOL. 94, NO. 2, December 2008,pp.1274-1281

[3] Chris Karlof David Wagner, "*Hidden Markov Model Cryptanalysis*", Report No. UCB//CSD-03-1244 Computer Science Division (EECS) University of California Berkeley, California 94720,VOL 155,pp.243-249,2009

[4] Dimitrov VS, G.A.Jullien and W.C.Miller VLSI Research Group, University of Windsor, Ontario Canada N9B 3B4,"*Algorithms for multi exponentiations based on complex arithmetic*",VOL.211. pp.208-215.,2009

[5] Elsayed Mohammada and Hassan Elkamchounchi, Alexandria university, Egypt, "Elliptic curve cryptography over Gaussian integers ", International Journal of Computer Science and Network Security, VOL.9 No.1, January 2009

[6] Majithia sachin and Dinesh kumar, "Implementation and Analysis of AES, DES and Triple DES", International Journal of Computer Science and Network Security, VOL.10 No.1, January 2010,pp.298-303

[7]Mircea Andrasiu university Wales – Romania, "StatisticalEvaluationofcryptographic algorithms",IEEE,vol.6,NO.3,January-2009,pp.255-261

[8] Mohammad Ahmad Alia and Azman Bin Samsudin, School of Computer Sciences, Universiti Sains Malaysia, "A New Digital Signature Scheme Based on Mandelbrot and Julia Fractal Sets", American Journal of Applied Sciences 4 (11): 848-856, 2007

[9]RSA Laboratories, http://www.rsa.com/rsalabs/

[10] Song Y Yan, "*Computability, Learnability and Breakability in Cryptanalysis*", volume 45,CIMCA 2008, IAWTIC 2008, and ISE 2008.

Dr.M.Phani Krishna Kishore, is working as a Professor in the department of Information Technology, Gayatri Vidya Parishad College of Engineering,Visakhapatnam. He obtained his Ph.D in 2006, M.Phil in 1998, M.Sc(Mathematics) in 1995, from Andhra University.

Mr.I.V.S.VENUGOPAL, is working as Asst. Professor in the department of Information Technology, Gayatri Vidya Parishad College of Engineering,Visakhapatnam. He obtained his M.Tech in 2010, B.Tech in 2006.

GPRS and Bluetooth Based Devices/Mobile Connectivity Shifting From Manual To Automation For Performance Optimization

Nazia Bibi[1], Tahira Mahboob[2] and Malik Sikander Hayat Khiyal[3]

[1] **Software Engineering, Fatima Jinnah Women University**
Rawalpindi, Pakistan

[2] **Software Engineering, Fatima Jinnah Women University**
Rawalpindi, Pakistan

[3] **Software Engineering, Fatima Jinnah Women University**
Rawalpindi, Pakistan

Abstract

Many companies/organizations are trying to move towards automation and provide their workers with the internet facility on their mobile in order to carry out their routine tasks to save time and resources. The proposed system is based on GPRS technology aims to provide a solution to problem faced in carryout routine tasks considering mobility. The system is designed in a way that facilitates Workers/field staff get updates on their mobile phone regarding tasks at hand. This System is beneficial in a sense that it saves resources in term of time, human resources and cuts down the paper work. The proposed system has been developed in view of research study conducted in the software development and telecom industry and provides a high end solution to the customers/fieldworkers that use GPRS technology for transactions updates of databases.

Keywords: *GPRS (General Packet Radio Service), GSM (Global System for Mobile communication), UMTS (universal Mobile Telecommunications System), HACS (SMS Based Wireless Home Appliance Control System), HTTP (hypertext transfer protocol), FTP (File Transfer Protocol), TCP/IP (Transmission Control Protocol / Internet Protocol)*

1. Introduction

Organizations are moving towards automation. With the growing need of gadget, mobility becomes the integral part of people's life [1] and now they are becoming more dependent on the mobility. They need solutions that are low in cost, high in speed, high in performances, saves human resources, increase the organization revenue and net profit, facilitate the customers, cut down the paper work etc. After analysis a system is proposed that provide the solution of the problem faced by the organizations, organization employees and customer. Following the results of research survey conducted proposed system is divided into two parts; "Server Side" [2] and "Client Side" [3]. Both Applications are connected to central database. Client side application is basically designed for mobile devices/phones and is connected to central database via GPRS [4]. "Client Side" application is used by field staff/workers, first they login then they are able to view the customer information and invoices in the database via GPRS connectivity, and are able to save the status of the customer in the database after delivering the product. "Server Side" application is especially for the administration side (the organization providing mobile data services) and is used by the employees who first login and then view, add, update, and delete the records.

The remaining part of this paper is divided into sections. In section II the reviewed literature is described, in sections III methodology is described in detail with frame work and mathematical model elaborated, in section VI implementation is described, in section V research analysis results and experimental results are illustrated and in last section conclusion and future work is summarized.

2. Literature Review

GPRS Mobile Phones - An Overview for Test Professionals [5], this paper focuses on the difference between GPRS and GSM, how GPRS operates and how GPRS is tested. GPRS (General Packet Radio Service) is widely growing technology that facilitates mobile users to transfer information in packets.

SMS Based Wireless Home Appliance Control System (HACS) for Automating Appliances and Security [6] remotely control the home appliances and also to provide the security to user when user is away from home and wireless technology has its impact on the living standards. This system provides solution for those problems which is faced by the home owner regarding security in their daily life. This system uses the GSM technology to access the appliance control and system for security.

The motivation behind design the system is to facilitate the user to automate their homes with a low cost system. Data Transfer Over GPRS: Palm OS [7], Transfer of files from a device to a pc (server) can be done in different way: like HTTP, FTP and TCP/IP. In these entire cases one thing that remains same is the channel that is the "Cellular Network". Things are bit different while sending data files across to the server, when we are using the cellular network. Software architecture should is considered very important when designing application depending on the GPRS.

Growth in the cellular telephony has increased demand for the wireless data services. There are many standards like 3G (3rd generation) UMTS (universal Mobile Telecommunications System) mobile networks but these are still away from the wide-scale development. In the meantime, the extension of GPRS (General Packet Radio Services) to current 2G GSM networks provides a widely deployed solution for data access. In the wired internet TCP is the dominant protocol and is very important for the mobile users too.

"Implementation of Telecontrol Applications over GPRS Networks" [8], depicts how GPRS technology has reached sufficient levels of development and reliability. GPRS can be applied in technological applications, either as a backup for private communications channels or as a main communications technology. It also tells how GPRS Applications work over network**.**

3. Methodology

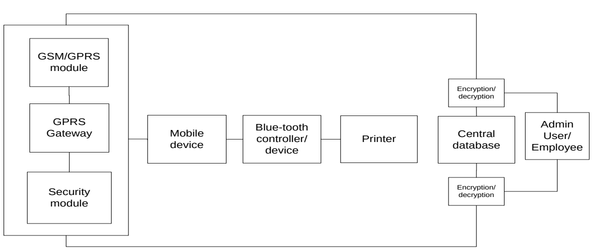


Figure 1: Frame work

The working of the system model given in fig1 is as follow:

Mobile Device (Bluetooth enabled): the mobile user wants to see the invoices of the customer. User first connects to the central database via GPRS and then user provides the login information. After successful login user will be able to view the invoices.

GPRS module: GPRS module provides the connectivity to the central database.

Central database: mobile user connects to the central database via GPRS module and Admin user/Employee connects to central database without using GPRS module.

Encryption/Decryption module: Encryption technique Caesar Cipher is used in order to provide the database security [9] [10].

Admin user/Employee: employee wants to insert, update and delete the records. Employee first connects to the central database and provides login information. After successful login employee can perform insertion, updating and deletion of records.

Bluetooth connectivity: mobile user print invoice by Bluetooth printer using Bluetooth connectivity.

Table 1: List of Factors

Symbols	Description
MU	Mobile user/Staff member/Worker
G	GPRS
Ec/Dc	Encryption/Decryption
S	Server
DB	Database
B	Bluetooth
P	Printer
R	Record
I	Insert
U	Update
D	Delete
PO	Place Order
C	Customer

E	Employee
Sv	Save
Ci	Customer Information
V	View
Aut	Authenticate
Linfo	Login Information
OI	Order Information
Ent	Enter
St	Save status in database
Pinv	Print invoice
Pdev	Deliver product

A. Client Side

Var1=C(PO) ———————— **(i)**

Var2=OI ———————— **(ii)**

Var3=E(Ent(Linfo)) ———————— **(iii)**

Var4= Aut(E(Ent(Linfo))) ———————— **(iv)**

Var5= MU(Ent(Linfo)) ———————— **(v)**

Var6=Aut(MU(Ent(Linfo)))

Var7=MU(V(Dc(Sv(Ci)))) ———————— **(vi)**

Var8=MU(V(Dc(E(Sv(Ec(OI)))))) ———— **(vii)**

Var9=MU(Pdev) ———————— **(viii)**

Var10= MU(Sv(Ec(St))) ———————— **(ix)**

Var11=MU(Pinv) ———————— **(x)**

B. Server Side

Var1=E(Ent(Linfo)) ———————— **(i)**

Var2=Aut(E(Ent(Linfo))) ———————— **(ii)**

Var3=DB(Sv(Ec(E(I(R))))) ———————— **(iii)**

Var4=DB(Sv(Ec(E(U(R))))) —————— **(iv)**

Var5=E(D(R)) ———————— **(v)**

4. Implementation

4.1 Characteristics of System

The system allow mobile user to remotely connect to the database in order to retrieve and enter the information, system security is provided by using the encryption/decryption at both ends (client side and server side).

4.2 Strengths of System

The proposed system has many advantages:

- Remotely access the database
- GPRS technology is used to provide connectivity to the central database
- The system is of low cost; provide high speed, cut down the paper works, saves on human resources.
- Easiness to the workers to carry out their tasks quickly and efficiently.
- Works gets tasks on their mobiles devices via GPRS, no need to go to offices in order to get their tasks.
- Workers get connected to central database via GPRS and can retrieve, update, save information directly in to database
- Data transactions are secured because encryption/decryption is applied at both ends (client side/server sides).

5. Results

Research is divided into two phases

5.1 Research Analysis Results
5.2 Experimental Results

5.1 Research Analysis Results

The research survey is divided into three sections i) Mobile applications in Organization ii) GPRS iii) Databases Sample size is 30 and the results of research are shown in graphs represented by Figure 2, Figure 3 and Figure 4.

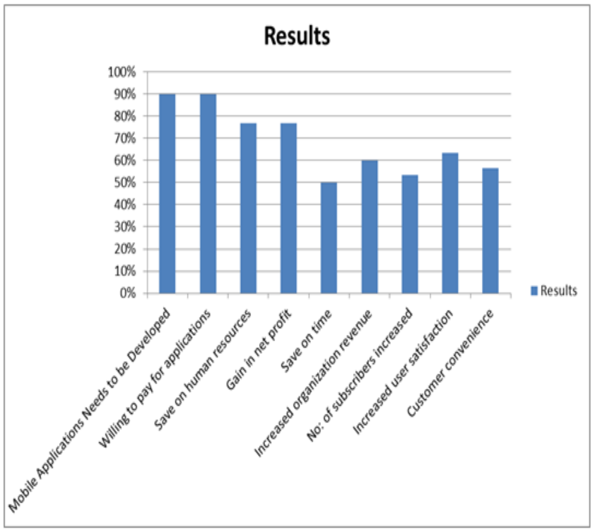


Figure 2: Importance of mobile applications

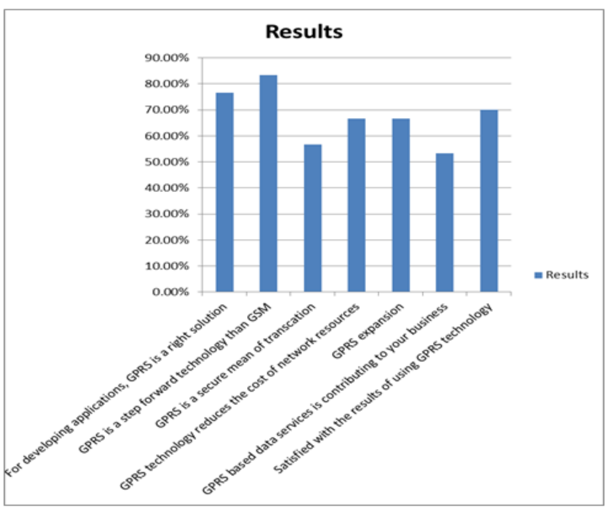


Figure 3: GPRS implementation

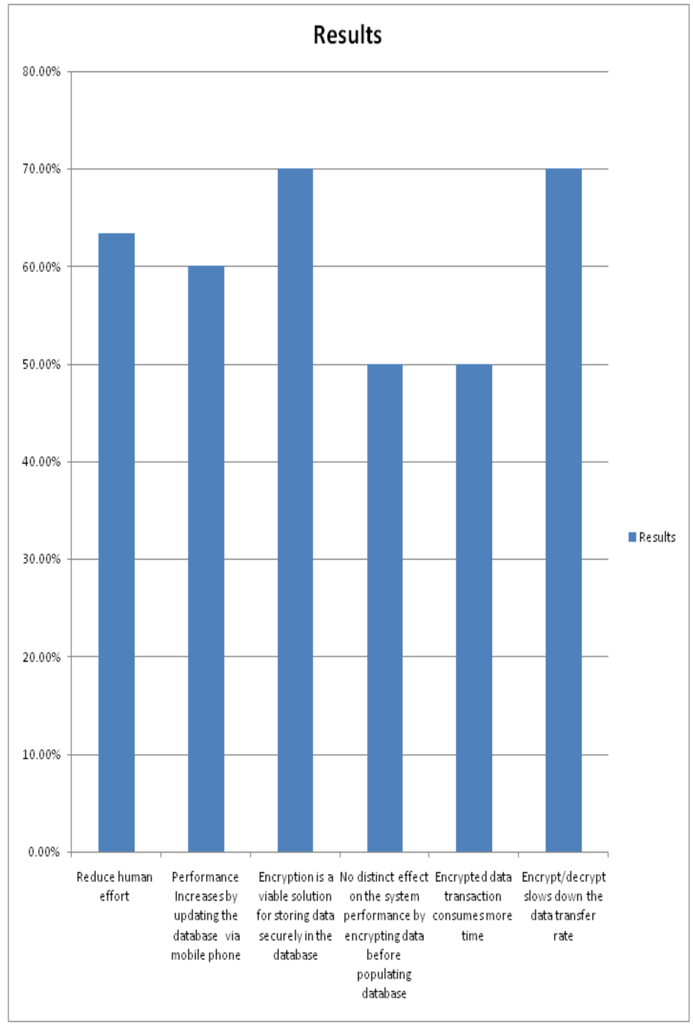


Figure 4: Databases

5.2 Experimental Results

5.1 Memory usage without encryption

\\PERSONAL-PC

Memory

% Committed Bytes In Use	32.118
Available MBytes	1,059.000
Cache Faults/sec	0.000

5.2 Memory usage with encryption

\\PERSONAL-PC

Memory

% Committed Bytes In Use	32.537
Available MBytes	1,040.000
Cache Faults/sec	0.000

5.3 Processor time without encryption

The fig 5 shows the processor time without encryption.

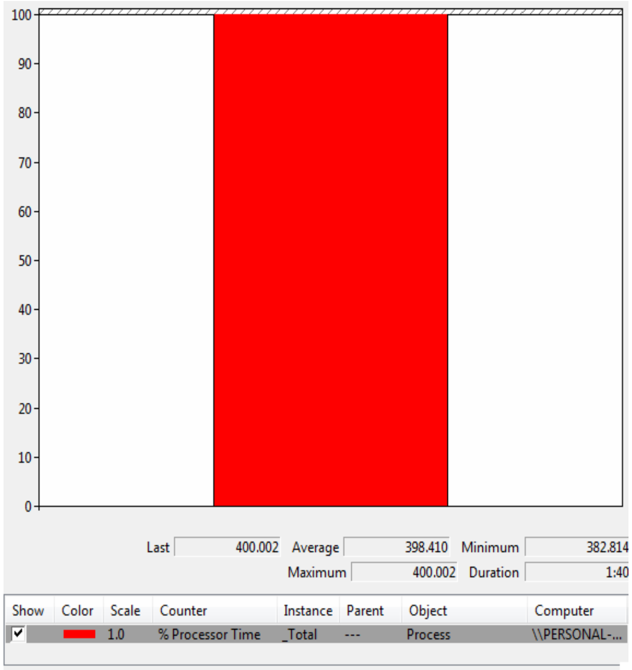


Figure 5: Processor time without encryption

5.4 Processor time with encryption

The fig 6 shows the processor time with encryption.

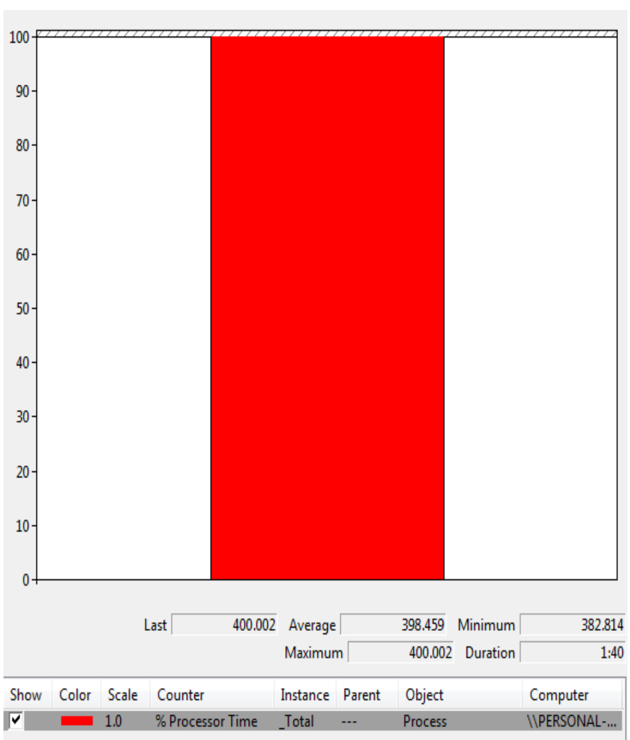


Figure 6: Processor time with encryption

5.5 Constraints of System

- **Mobile application(Client-Side) failed**
  Mobile user must have backup of client-side application in any case of failure of application that is installed on his mobile device.
- **GPRS not available in some areas**
  If worker is in that area where GPRS services are not available then he can save or update the data in the database that is locally available in his mobile device.
- **Mobile devices corrupted/not properly working**
  In any case if mobile devices get corrupted/not working properly then worker should have an extra mobile device with the feature of GPRS and Bluetooth

6. Conclusion and future work

Organizations can easily move towards automations,saves on human resources, increase net profit,saves on extra effort by using this proposed system and survey results shows that organizations are willing to pay for such proposed system/applications.For developing applications GPRS is the right solution and survey results also shows that people are satisfied with the results of this technology as it is a step forward technology then GSM and it provides secure mean for data transaction and also saves on network resources.As our proposed system is divided in to two application one is Client side application and other is server side application and both are connected with the central database.Survey results shows that use of databases cut down the paper work, increases performance by updating the databases via mobile phones rather than manually entering the data, encryption data before storing in the database is a viable solution for sceuring data.
Experimental results shows that there is a negligeable effect on the memory when using encryption/decryption and processor time with and without encryption/decryption is almost same.

Anyone can use different technology to implement the system other than GPRS. Secondly, anyone can use the same technology but can use different method to connect to the central database as here IP address in the URL is given to connect to the central database. Third, anyone can add the additional functionality that after delivering product can print the invoice through the Bluetooth mobile/devices by using the Bluetooth printer. Fourth, If GPRS service is not available in an area then the mobile user is provided with some mechanism that he save records locally in the database that is available in his mobile and when GPRS facility is available then he sinks the local database with the central database. Last, anyone can embed the feature of the GPS (Global Positing system) so that to keep track of the current position of the employee.

References

[1] http://download.pwc.com/ie/pubs/convergence_monito_personal_mobility.pdf,15th Feb,2011.
[2] http://www.roseindia.net/servlets/introductiontoserverside.shtml, 20th Feb,2011.
[3] http://www.maffeis.com/articles/research/client_server.pdf, 15th March,2011.
[4] http://www.item.ntnu.no/fag/tm8100/Pensumstoff2004/GPRS_Tutorial.pdf,23 March,2011.
[5] http://aeroflex.com/ats/products/prodfiles/appnotes/96/gprs-extra.pdf, 25 March,2011.
[6] Malik Sikandar Hayat Khiyal, Aihab Khan, and Erum Shehzadi. SMS Based Wireless Home Appliance Control System (HACS) for Automating Appliances and Security. The journal of issues in informing sciences and information technology, Vol 6, PP 877-886, 2009.
[7] http://www.techrepublic.com/whitepapers/data-transfer-over-gprs-palm-os/2273283, 23rd March,2011.
[8] http://www.techrepublic.com/whitepapers/implementation-of-telecontrol-applications-over-gprs-networks/2253065, 25th April,2011.
[9] http://www.nku.edu/~christensen/section%202%20caesar%20ciphers.pdf
[10] Dennis Luciano, Gordon Pritchett. Cryptology: From Caesar Ciphers to Public-Key Cryptosystems. The College Mathematics Journal, January 1987, Volume 18, Number 1, pp. 2–17.

Nazia Bibi
Nazia Bibi is enrolled in CASE University for MSSE, Islamabad. She received her Bachelors degree from Fatima Jinnah women university, Rawalpindi. She got first position in ICS and awarded with shield. Her areas of interest are databases, networking and development

Tahira Mahboob
She is a registered Engineer with the Pakistan Engineering Council. She received her bachelors degree from University of Engineering & Technology, Lahore in 2007. Currently enrolled in the MS/MPhil Computer Engineering program at Center for Advanced Studies in Engineering CASE(2010), UET Taxila. She has industry experience in telecom sector (Intelligent networks & VAS) Currently serving as a lecturer at Fatima Jinnah Women University, Rawalpindi. She has supervised thesis/projects at Bachelors and Masters Degree Programs in mobile/computer communications, voice recognition, mobile automation and cloud computing. Her area of interests are computer/mobile communications & networks, information security, mobile automation and adhoc/ sensor networks.

Malik Sikandar Hayat Khiyal
Dr. **Malik Sikandar Hayat Khiyal** is Chairperson Department of Computer Sciences and Software Engineering at Fatima Jinnah Women University, Pakistan. He received his M.Sc degree from Quaid-e-Azam University, Islamabad. He got first position in the faculty of Natural Science of the University. He was awarded the merit scholarship for Ph.D. He received his Ph.D. degree from UMIST, Manchester, U.K. He developed software of underground flow and advanced fluid dynamic techniques. His areas of interest are Numerical Analysis, Modeling and Simulation, Discrete structure, Data structure, Analysis of Algorithm, Theory of Automata and Theory of Computation. He has more than hundred research publications in National and International Journals and Conference proceedings.

A New Routing Algorithm for a Three-Stage Clos Interconnection Networks

Zahra sadat Ghandriz[1] and Esmaeil Zeinali Kh.[2]

[1] Department of Computer Engineering, Islamic Azad University, Qazvin Branch, QIAU, Iran

[2] Department of Computer Engineering, Islamic Azad University, Qazvin Branch, QIAU, Iran

Abstract

Clos Interconnection network is one of the known connection networks in processing systems and distributed systems, which is used extensively in many fields such as Telecommunication networks, ATM switches and Data transmission. In order to eliminate the blocking in such networks, various routing algorithm have been proposed, each imposing extra costs due to hardware use and re-routing algorithm. This study offers a routing algorithm which takes a blocking-avoidance approach hence avoiding related costs. There is no blocking while the primary routing is performed from the input to output. This method has the complexity of $\boldsymbol{O(N\sqrt{N})}$. The results show that this algorithm is simpler than the algorithms previously proposed.

Keywords: *Clos Interconnection Networks, Routing Algorithm, Blocking Avoidance, Three Stage Interconnection Networks.*

1. Introduction

As one of the most important parts of parallel processing systems, interconnection networks make connection between switches [1, 2, 15]. The topologies typically found in processing systems are often regular. They are classified into two groups: statistic and dynamic. The networks using dynamic topology fall in the following categories of structures: bus networks, crossbar switching networks, multi-stage connection networks. Multistage interconnection networks consist of blocking networks, rearrangeable non-blocking networks, wide-sense non-blocking networks and strictly non-blocking networks [3, 4, 5, 6].

Routing algorithm can be divided into two groups: 1- blocking avoidance routing, 2- routing irrespective of blocking. In the first method, routing in the algorithm is performed in such a way that there is no blocking but in the second method, initially the routing is performed from input set to output set. In case of any blocking, attempts are made to eliminate it through changing the arrangements in network switching, using routing algorithm.

Routing algorithm in multi-stage connection can be performed by 2 methods: 1- Graph coloring algorithm, 2- decomposition algorithm using matrixes. Generally, graph coloring algorithm enjoys a better time complexity than decomposition algorithm, using matrix. However, as the network size is too large, it will be inefficient. Decomposition algorithm using matrixes has the following advantages: Directly locating the problems and enjoying simplicity in switching settings .Unfortunately, many proposed decomposition algorithms using matrix are incomplete. Neiman algorithm has been implemented, using decomposition matrix and time complexity $O(N\sqrt{N})$ [13]. GS algorithm with complexity $O(N\sqrt{N})$ uses 2 matrixes, namely: Specification and Count. But Siu, Chiu, Lee and Carpinelli believes GS algorithm to be incomplete [7, 13]. Later, a new method called the modified methods of GS was proposed [7]. This algorithm was proposed at time $O(N\sqrt{N}Log\sqrt{N})$, by adding 3 steps to the main algorithm and deleting 2 steps for purpose of eliminating indefinite loops. Later, an algorithm on the basis of Heuristic Routing Algorithm using minimum distribution priority scheme was introduced for routing Clos networks. This method had the capability of accessing all non-blocking routings, reducing its time complexity to $O(N\sqrt{N})$ in the worst situation [14].

This study presents the routing mechanism in Clos interconnection networks, taking blocking-avoidance approach. This method can be applied extensively in distributed and parallel networks for purpose of processing. Section 2 introduces the Clos networks and blocking-avoidance routing algorithm. Section 3 compares this new method with the previously proposed methods and section 4 concludes this study.

2. Blocking Avoidance Routing Algorithm in Clos Networks

2.1 Clos Network

Using small crossbar switches, Charles Clos introduced a type of interconnection network which is extensively studied and applied as a framework for ATM switches because it is economical, regular, scalable, fault-tolerant and highly efficient. Special attention has been paid to Clos three-stage networks as they are rearrangeable for developing multi-stage networks. These three stage-networks are intended to be used for data communication and parallel computing system.

A switching network is composed of one or more switch stages that can create various paths through creating various connections between their inputs and outputs. Clos three-stage network is an example of multi-stage switching networks (Fig.1).

Clos three-stage networks $N \times M$ are represented as $c(n_1, n_2, m, r_1, r_2)$ where N represents the overall inputs of network and M as overall outputs of network. n_1, n_2, m, r_1, r_2, represent the number of inputs of each switch of input stage, the number of outputs of each switch of output stage, number of switches of middle stage ,number of switches of input stage and number of switches of output stage, respectively. If $N = M$ then $r_1 = r_2$, $n_1 = n_2$. In this case, Clos symmetrical networks are shown as $c(n, r, m)$ [15, 16]. The first stage of a three-stage network is called input stage which includes $r\,(n \times m)$ switches. The second stage is called middle stage which includes $m\,(r \times r)$ switches. The third stage is called output stage which includes r $(m \times n)$ switches. The network is capable of connecting one to one and one to many with $N = nr$ and $m \geq n$. There is a link between 2 switches in 2 continuous stages [7, 8]. As $c(n, r, m)$ knows all possible permutations between inputs and outputs. A link can be accessed between stages, provided it is usable and not engaged.

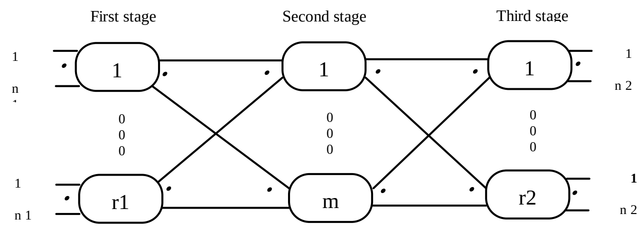


Fig. 1. Clos Three-stage network [15].

The Clos networks have the following advantages over other interconnection networks [8, 9, 10, 11, 12]:

1. This network is more efficient in terms of switching data.
2. The lower hardware costs of this network have drawn the attention of many researchers.
3. Clos networks have the capability to add switching packages in their configuration, increasing fault-tolerance compared to one-stage interconnection networks.

2.2 Vector Modeling of Clos Networks

The connections between each input and output can be shown as follows:

$$P = \begin{bmatrix} 0 & 1 & \cdots & i & \cdots & N-1 \\ \pi(0) & \pi(1) & \cdots & \pi(i) & \cdots & \pi(N-1) \end{bmatrix} \quad (1)$$

Where input i is connected to output $\pi(i)$ and $0 \leq i \leq N - 1$, $N = nr$. The switch between input and output is assumed to be of non-blocking type. P may be changed between input switches and output switches [13].

In this algorithm, the matrix is used in such a way that network initially has its connections. Then, the new inputs are added into the network in such a way that there will be not any blocking. This is made possible through 4 main matrixes (A, B, C, D) and an mono dimensional array (e). The matrixes consist of elements 0, 1. 0 represents the free link and 1 represents the engaged link in the network.

Now the matrixes and array (e) are explained briefly:

1. Matrix A represents a connection between input switches and inputs of each input switch. The number of rows (i) represents input switches and the number of columns (j) represents the inputs of each input switch.
2. Matrix B represents the connection between middle switches and inputs of each middle switch. The number of rows (i1) represents the middle switches and the number of columns (j1) represents inputs of each middle switch.

3. Matrix C shows the connections between output switches and inputs of each output switch. The number of rows (i3) encompasses output switches and the number of columns (j3) represents the input of each output switch.
4. Matrix D indicates the connection between the outputs of each output switch and output switches. Here, the row (i4) represents output switches and columns (j4) represent the outputs of each output switch.
5. Array (e): each member of this array represents the number of the engaged input of each middle switch. The length of the array equals the number of middle stages.

Definition: All steps are complete when the elements of matrix A are transferred to matrix D. That is, all contents of input switches are added in the network and the input stage links are free for receiving new inputs.

2.3 Blocking-Avoidance Routing Algorithm

1. Initially, 0 is assigned to matrixes B, C, D and array (e).
2. Matrix A is derived from the inputs.
3. In matrixes B and A, a movement is made in parallel in such a way that in matrix A, the movement is made from row to row and in matrix B form column to column. When 1 is seen in matrix A, a free link is needed in middle switches to which the inputs are transferred. Then we turn to matrix B, searching for 0. On finding the first 0 in matrix B, we should replace it by 1 in matrix A.
4. Having transferred all engaged links of matrix A (input stage) to matrix B (middle stage) in such a way that there is no blocking in the input switches, now the output of middle stages should be arranged. To this purpose, the number of engaged input links into each middle switch is calculated, then added to the array (e). Having calculated all engaged links, we should assign the passive output links in middle switches (matrix B) to the engaged input links in middle switches (in array (e)).
5. We should arrange matrix B (middle stage) to matrix C (output stage) in parallel so that matrix B is examined in terms of columns and matrix C in terms of rows. On seeing 1 in matrix B, we should replace it by 0 in matrix C.
6. Finally, in order to transfer the engaged links to the output, we should make use of matrix D. That is, the movement in matrix C should be made from row to row and in metric D from the row to row in parallel. On seeing 1 in matrix C, it is replaced by 0 in matrix D.

Now, pseudo code of blocking-avoidance routing algorithm is presented as follows:

Step1) Initialize by Setting Middle Matrix (B), $Output_1$ Matrix(C), $Output_2$ Matrix(D) and Middle array (e) = False.

Step 2) Read Input Matrix (A).

Step 3) For each of row elements of Matrix A, that is "True", if $(A[i,j] == 1)$ then {Find The First column elements of Matrix B, that is "False", if $(B[i_1,j_1] == 0)$ {Swap $(A[i,j], B[i_1,j_1])$; }}

Step 4) For each of row elements of Matrix B, that is "True", if $(B[i_1,j_1] == 1)$ then {S++;}e [k++]=S;
if $(e[k > 0])$ then { $B[k][j_{11}] = 1$, k-- ;}

Step 5) For each of column elements of Matrix(B), that is "True" , if $(B[i_1,j_1] == 1)$ then {Find The First row elements of Matrix C, that is "False", if $(C[i_3,j_3] == 0)$ {Swap $(B[i_1,j_1], C[i_3,j_3])$; }}

Step 6) For each of row elements of Matrix C , that is "True", if $C[i_3,j_3] == 1$ then { Find The First row elements of Matrix D, that is "False", if $(D[i_4,j_4] == 0)$ { Swap$(C[i_3,j_3], D[i_4,j_4])$; }}

As you can see, all elements in matrixes A, B and C can be transferred to the next stage matrix in a parallel manner. This speaks for the parallel routing in the proposed manner.

3. Comparing blocking- avoidance routing algorithm with the previous methods

To compare the algorithm proposed by this study with the previous ones, we can examine them from two perspectives:

- The complexity of routing algorithm
- Time

3.1 The complexity of routing algorithm

The following should be taken into account to compute the complexity of routing algorithm:

- m = n, this indicates that the number of middle stage switches (m) equals the number of input ports of the switches of input stage (n).
- N= r×n, this indicates that overall network inputs are obtained by multiplying the number of switches of input stage (r) by the input ports of each switch of input stage (n).
- $n = \sqrt{N}$.

Each matrix is read by time $O(N)$ and the main body of algorithm is formed at $O(Nn)$. thus, the overall complexity of proposed algorithm is :

$$O(Nn) + 3O(N) = O(Nn) \xrightarrow{n=\sqrt{N}} O\left(N\sqrt{N}\right)$$

Table1 shows the complexity of blocking-avoidance algorithm, drawing on the previously proposed algorithms.

Table 1: shows the complexity of blocking-avoidance algorithm in comparison with the previously proposed algorithms.

Algorithm name	Complexity of Routing Algorithm
GS Algorithm	$N\sqrt{N}$
GS Modification Algorithm	$N\sqrt{N}Log\sqrt{N}$
Heuristic Algorithm	$N\sqrt{N}$
Blocking-Avoidance Algorithm	$N\sqrt{N}$

As table 1 shows, the proposed routing algorithm outperforms the GS modified routing algorithm, with the former maintaining the network hardware. The proposed routing algorithm also equals Heuristic routing algorithm and GS routing algorithm in terms of algorithm complexity.

In GS routing Algorithm, there are indefinite loops which don't appear in the proposed routing algorithm in this study, resulting in higher efficiency of the proposed method.

In heuristic routing method, rearrangement is required. This leads to additional costs. While, in blocking avoidance approach, there is no such cost. This is because the blocking is eliminated while routing is being done.

Hence the better performance of this method over other methods.

3.2 Time

Regarding time, the algorithm proposed by this study outperforms the previously proposed methods. The previous methods start routing without taking account of blocking. That is, when routing is over, these methods make an attempt to eliminate the blocking through using some new algorithm, taking some time. This method proposed by this study starts routing so as to prevent the occurrence of any blocking, saving a lot of time.

4. Discussion and Conclusion

This study presents a routing mechanism in Clos interconnection networks, taking a blocking-avoidance approach so that the routing is performed properly from input to output without blocking the links in network. While in the previously proposed methods, the strictly non-blocking network was defined by simple routing which results in high hardware costs. Or alternatively, the routing was performed irrespective of blocking. Then, a new algorithm was used to eliminate these blocking, leading to some costs. The method proposed by this study has solved this problem. On the other hand, this algorithm completes its cycle at time $O(N\sqrt{N})$. This type of routing can be used in communications switching as well as in data transmission networks for purpose of reducing the delay in transference time and for controlling the network traffic. The future studies can examine the ways in which new algorithms can be used to reduce routing time and the use of memory and to reduce the complexity of algorithm.

References

[1] H. Changa, J.Y. Guoa, F.K. Hwanga and C.K. Linb, “Wide-sense nonblocking for symmetric or asymmetric 3-stage Clos networks under various routing strategies”, Theoretical Computer Science, No. 314, 2004, pp. 375–386.
[2] L.M. Ni and P.K. Mckinley, “A survey of wormhole routing techniques in direct network”, IEEE Computer Society, VOL. 26, NO. 2, February 1993.
[3] P. Haynos and Y. Yang, “An Analytical Model on the Blocking Probability of a Fault Tolerant Network”, IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 10, NO. 10, October 1999.
[4] Q. Ngo, ”A new routing algorithm for multirate rearrangeable Clos networks”, Theoretical Computer Science, No. 290, 2003, pp. 2157–2167.
[5] Holmberg, “Optimization Models for Routing in Switching Networks of Clos Type with Many Stages“, An Electronic International Journal - AMO Advanced Modeling and Optimization, Volume 10, Number 1, 2008, pp. 1841-4311.

[6] S. Sharma, P.K. Bansal and K.S. Kahlon, “On a class of multistage interconnection network in parallel processing “, IJCSNS International Journal of Computer Science and Network Security, VOL.8, No.5, May 2008.

[7] F.K. Hwang, "A Modification to a Decomposition Algorithm of Gordon and Srikanthan", IEEE TRANSACTIONS ON COMPUTERS, VOL. 46, August 1997.

[8] X. Duan, D. Zhang and X. Sun, "Topology and Routing Schemes for Fault-tolerant Clos Network", International Conference on Networks Security, Wireless Communications and Trusted Computing, 2009.

[9] Y.Yang and J.Wang, “A New Design for Wide-Sense Nonblocking Multicast Switching Networks“, IEEE TRANSACTIONS ON COMMUNICATIONS, VOL. 53, NO. 3, March 2005.

[10] Y.Yang and J.Wang, "A Fault-Tolerant Rearrangeable Permutation Network", IEEE TRANSACTIONS ON COMPUTERS, VOL. 53, NO. 4, April 2004.

[11] Yuanyuan Yang, "A New Conference Network for Group Communication", IEEE TRANSACTIONS ON COMPUTERS, VOL. 51, NO. 9, September 2002.

[12] uanyuan Yang, "A Class of Multistage Conference Switching Networks for Group Communication", IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 15, NO. 3, March 2004.

[13] Y. Lee, F.k. Hwang and J.D. Carpinelli, "A New Decomposition Algorithm for Rearrangeable Clos Interconnection Networks", IEEE TRANSACTIONS ON COMMUNICATIONS,VOL. 44, NO. 11, November 1996.

[14] Duan and S. Liu, "A Heuristic Routing Algorithm for Clos Network", IEEE World Congress on Proceedings of the 7th, Intelligent Control and Automation, Chongqing, China, June 2008, pp. 25-27.

[15] Esmaeil Zeinali Kh. and Mehdi N. Fesharaki, "NMR: A New Approach for Optimal Desgin of Strictly Non-Blocking Multistage Interconnection Networks", Journal of Applied Sciences, Vol. 8, No. 19, 2008, pp. 3453-3459.

[16] Esmaeil Zeinali Kh., Mehdi N. Fesharaki and Mohammad Teshnehlab, "Optimal Design of Switch Sizes in Strictly Non-Blocking Clos Three-Stage Interconnection Networks", Journal of Applied Sciences, Vol. 8, No. 22, 2008, pp. 4234-4237.

Zahra Sadat Ghandriz was born in Yazd, Iran. She is currently a Master student in Hardware Engineering at Azad University of Qazvin, Iran. She received her B.Sc. degree in Software Engineering from Azad University, Maybod, Iran. Her research interests include Parallel Processing in Interconnection Networks and Parallel Algorithms.

Esmaeil Zeinali Khasraghi earned his B.Sc. degree in Computer Engineering (Software) from Qazvin Islamic Azad University (QIAU), and his M.Sc. and PhD degrees in Computer Engineering (Computer Architecture and Hardware) from Science and Research University, Tehran, Iran. He is an Assistant Professor of computer engineering and the QIAU Vice-President for Education since 2004. Working as a full-time faculty member of QIAU, he also directs the High Performance Computing and Networking Research Labs at the university. His research interests include Parallel and Distributed Systems, Optical and Wireless Networks and High Performance Computer Architecture.

Application of Neural Networks for Noise and Filter Classification to enhance the Image Quality

Dr. T. Santhanam[1], S. Radhika[2]

**[1]Assoc. Prof.& Head, PG & Research Department of Computer Science
D. G. Vaishnav College, Chennai, India**

**[2]Assoc. Prof., Department of Computer Applications
R. M. K. Engineering College, Tiruvallur Dt., TN, India**

Abstract

Image processing operations can be categorized into three major categories: image enhancement, image compression and image restoration. The objective of image enhancement is to improve the insight of the information in images for human viewing or to be used as a preprocessing tool for other image processing techniques. Filtering techniques play a crucial role in enhancing the quality of an image. This article explores the possibility of using an Artificial Neural Network for image noise classification followed by the suitable filter classification. Probabilistic Neural Network strikes a superior performance in identifying the noise as well as the suitable filter for the removal of a specific type of noise.

Keywords: *Image enhancement, Noise and Filter classification, neural networks.*

1. Introduction

Image enhancement aims in improving the quality of a digital image needed for visual inspection or for automated image processing operations. The main objective of image enhancement is to process an image so that the result is more suitable than the original image for a specific application. Image enhancement approaches fall in to two broader categories namely spatial domain methods and frequency domain methods. The approaches in the spatial domain are based on direct manipulation of pixels in an image and the approaches in the frequency domain refer to the modification of the Fourier transform of an image. When image enhancement techniques are used as pre-processing tools for other image processing techniques, then quantitative measures can determine which techniques are most appropriate.

1.1. Spatial Domain methods

The value of a pixel with coordinates (x,y) in the enhanced image F is the result of performing some operation on the pixels in the neighborhood of (x,y) in the input image, F. Neighborhoods can be any shape, but usually they are rectangular. Gray level transformations, Histogram processing, Enhancement using Arithmetic/Logic operations, Smoothing filters and Sharpening filters are some of the methods used in the spatial domain for image enhancement.

1.2. Frequency domain methods

Image enhancement in the frequency domain is straightforward. The Fourier transform of the image to be enhanced is computed, multiply the result by a filter and take the inverse transform to produce the enhanced image. Ideal filters, Butterworth filters, Gaussian filters, Laplacian filters and Homomorphic filters are used in the frequency domain for image enhancement.

In Information Technology, biometric authentication refers to technologies that measure and analyzes human physical and behavioural characteristics. Physical characteristics such as fingerprints, irises and facial patterns when captured, as raw images requires some image processing techniques to get the exact image so that the individual's identity is established. Filters are widely employed in the field of biometrics for improving the quality of images before extracting the features for identification/recognition.

2. Artificial Neural Network

An Artificial Neural Network (ANN), usually called "Neural Network" (NN), is a mathematical model or computational model that tries to simulate the structure and/or functional aspects of biological neural networks. It consists of an interconnected group of artificial neurons and processes information using a connectionist approach to computation. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase. Neural networks can be used to model complex relationships between inputs and outputs or to find patterns in data.

Classification is one of the major research areas of neural networks and many neural networks have emerged as an important tool for classification. The recent research has established that neural networks are a promising alternative to various conventional classification methods. The advantage of neural networks is that it makes use of self-adaptive methods to adjust to the data without any explicit specification. Fabio Roli and G N Marcialis used a single layer perceptron with a class-separation loss function for classifying individuals based on their fingerprints. [2]

Image denoising is an important image processing task, both as a process itself, and as a component in other processes. Very many ways to denoise an image or a set of data exists. The main properties of a good image denoising model are that it will remove noise while preserving the fine details of the image. Many image denoising algorithms were prevalent in the past, but they weren't successful when it comes to automatic image restoration. [3][4][5][6] Identification of the noise is crucial for an image denoising model to be accurate.

The use of a Multi Layer Perceptron (MLP) [7], Back Propagation Network (BPN) [7] and Probabilistic Neural Network (PNN) [8] to classify the image noise, based on the statistical features like skewness and kurtosis is available in the literature. Noise identification is vital for determining the denoising procedure for an image, which leads to image enhancement for further processing.

2.1 Back Propagation Network (BPN)

Back Propagation Network is a multilayer feedforward network employing back propagation algorithm. As the name implies, the errors propagate backwards from the output nodes to the inner nodes.

The steps in the BPN algorithm are [9]

1. Select a pattern X_k from the training set T, and present it to the network.
2. Compute activations and signals of input, hidden and output neurons in that sequence.
3. Find the error over the output neurons by comparing the generated outputs with the desired outputs.
4. Use the error calculated in step 3 to compute the change in the hidden to output layer weights, and the change in input to hidden layer weights (including all bias weights), such that a global error measure gets reduced.
5. Update all weights of the network in accordance with the changes computed in step 4.

   Hidden to output layer weights

   $$W_{hj}^{k+1} = W_{hj}^{k} + \Delta W_{hj}^{k} \quad (1)$$

   Input to hidden layer weights

   $$W_{ih}^{k+1} = W_{ih}^{k} + \Delta W_{ih}^{k} \quad (2)$$

   where ΔW_{hj}^{k} and ΔW_{ih}^{k} are weight changes computed in step 4.
6. Repeat steps 1 through 5 until the global error falls below a predefined threshold.

T. Kalpalatha Reddy and N. Kumaravel have used BPN for classification of bone samples at different locations of the jawbone region [12] and Graham Kendall *et al.* have used it to classify document zone content in technical document images. [13]

2.2 Multilayer Perceptron (MLP)

The most common neural network model is the multilayer perceptron (MLP). This type of neural network is known as a supervised network because it requires a desired output in order to learn. The goal of this type of network is to create a model that correctly maps the input to the output using historical data so that the model can then be used to produce the output when the desired output is unknown.

The inputs are fed into the input layer and get multiplied by interconnection weights as they are passed from the input layer to the first hidden layer. Within the first hidden layer, they get summed then processed by a nonlinear function. As the processed data leaves the first hidden layer, again it gets multiplied by interconnection weights, then summed and processed by the second hidden layer. Finally the data is multiplied by interconnection weights then processed one last time within the output layer to produce the neural network output.

Real-world task like the classification of hand-written numerals may be efficiently and economically accomplished by means of a general-purpose Multi Layer Perceptron. [1]

2.3 Probabilistic Neural Network (PNN)

Probabilistic neural networks are forward feed networks built with three layers. They are derived from Bayes Decision Networks. They train quickly since the training is done in one pass of each training vector, rather than several. Probabilistic neural networks estimate the probability density function for each class based on the training samples.

There is an input unit for each dimension in the vector. The input layer is fully connected to the hidden layer. The hidden layer has a node for each classification. Each hidden node calculates the dot product of the input vector with a test vector subtracts 1 from it and divides the result by the standard deviation squared. The output layer has a node for each pattern classification. The sum for each hidden node is sent to the output layer and the highest values wins.

The Probabilistic neural network trains immediately and is used for classifying data. Probabilistic neural networks handle data that has spikes and points outside the norm better than other neural nets. PNN is used in classifying cancer data. [10][11]

3.Methodology

Step 1: The noises in an image are classified as non-gaussian white, gaussian white and salt and pepper noise by the PNN as given in [8] and the performance of the network is above 90%.

Step 2: The classified noises are given as input to BPN, MLP and PNN networks which identifies the suitable filters for noise removal. The filters that are classified by the neural network are Wiener filter (Non gaussian white) [14], Infinte Impulse Response filter (Gaussian white) [15] and Median filter (Salt and pepper noise) [16] . The performance of BPN, MLP and PNN networks are given in Table-1.

A k-fold cross-validation is used. i.e. the images are partitioned in to k (say 10) partitions at random. Training is carried out with k-1 partitions and testing is carried out with the left out partition. The cross validation process is then repeated k times, with each of the k partition being used exactly once for testing. The k results are averaged to determine the resulting accuracy.

Table 1: Performance analysis of PNN, MLP and BPN network

	Wiener filter	***IIR filter***	***Median filter***
PNN	93.33%	96.67%	93.27%
MLP	96.67%	86.67%	90%
BPN	93.33%	93.33%	86.67%

The entries in the table give the correct % of filters classified for a specific type of noise. The performance of PNN in classifying the filter for noise removal is greater than 90% and is superior to MLP and BPN network.

4.Conclusion

The use of Neural Network for classification of noise followed by classification of filter is explored in this article. CASIA-Irisv3 database have been used to test the performance of the network and the experiments have been carried out in MATLAB. The results show that PNN proves to be a better network in classifying the noises as well as filters than the MLP and BPN models. The results obtained can be used in employing the suitable filter for noise removal, thus enhancing the image for further processing.

References

[1] M. Costa, E. Filippi and E. Pasero, “A Modular cyclic Neural Network for character recognition”, Proceedings of the INNS World Congress on Neural Networks (WCNN '94), S. Diego (CA), Vol 3, June 5-9 1994, pp 204-210.
[2] Marcialis, G., Roli, F., “Fusion of multiple fingerprint matchers by singlelayer perceptron with class-separation loss function”, Pattern Recognition, Letters 26, 2005, pp.1830–1839.
[3] Wenbin Luo, “Efficient removal of impulse noise from digital images”, IEEE Transactions on Consumer Electronics, Vol 52, Issue 2, 2006, pp. 523 – 527.
[4] M. A. T. Figueiredo, R. D. Nowak, “An EM algorithm for wavelelt based image restoration”, IEEE Transactions on Image Processing, vol. 12, 2003, pp 906-916.
[5] Buades, A.; Coll, B.; Morel, J.-M., “A non-local algorithm for image denoising”, CVPR 2005. IEEE Computer Society Conference, vol. 2, 2005, pp. 60 – 65.
[6] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, “Image denoising using scale mixtures of

Gaussians in the wavelet domain", IEEE Transactions on Image Processing, vol. 12, 2003, pp. 1338–1351.

[7] T. Santhanam and S. Radhika, "A Novel Approach to Classify Noises in Images Using Artificial Neural Network", Journal of Computer Science 6 (5): pp. 541-545, 2010

[8] T. Santhanam and S. Radhika, "Probabilistic Neural Network – A better solution for noise classification", Journal of Theoretical and Applied Information Technology, Vol 27 No. 1, 2011, pp. 39-42.

[9] Satish kumar, "Neural Networks: A Classroom Approach", Tata Mc-Graw Hill Publications, pp. 169, 2004.

[10] Chenn-Jung Huang and Wei-Chen Liao, "Application of Probabilistic Neural Networks to the Class Prediction of Leukemia and Embryonal Tumor of Central Nervous System", Neural Processing Letters, Vol. 19, 2004, pp. 211-226.

[11] Yichu Shan, Ruihuan Zhao, Guowang Xu, H. M. Liebich and Yukui Zhang, "Application of probabilistic neural network in the clinical diagnosis of cancers based on clinical chemistry data", Analytica Chimica Acta, Volume 471, Issue 1, 2002, pp. 77-86.

[12] T. Kalpalatha Reddy and N. Kumaravel, "Texture Analysis of Bone CT Images for Classification and Characterization of Bone Quality", International Journal of Soft Computing 4(5), 2009, pp. 223-228. [2] Marcialis, G., Roli, F., "Fusion of multiple fingerprint matchers by singlelayer perceptron with class-separation loss function", Pattern Recognition, Letters 26, 2005, pp.1830–1839

[13] Ibrahim Z, Isa D, Rajkumar R, Kendall Graham, "Document Zone Content Classification for Technical Document Images using Artificial Neural Networks and Support Vector Machines", Proceedings of ICADIWT '09, pp. 345-350, 2009

[14]J. S. Lim, "Two Dimensional Signal and Image Processing", Prentice Hall, 1990

[15]Sorensen J.S., Johannesen L, Grove U.S.L, Lundhus K, Couderc J.P, Graff C, "A comparison of IIR and wavelet filtering for noise reduction of the ECG", Computers in Cardiology, 2010, pp. 489-492.

[16]http://homepages.inf.ed.ac.uk/rbf/HIPR2/median.htm

Dr. T. Santhanam is working as Associate Professor and Head in D. G. Vaishnav College, Chennai. He is having over 20 years of experience in teaching and has many research publications in international journals and conferences. His areas of interest are Biometrics, Image Processing, Neural Networks and Data mining.

S. Radhika is working as Associate Professor in R. M. K. Engineering College and is pursuing her research in D. G. Vaishnav College affiliated to University of Madras, under the supervision of Dr. T. Santhanam. Her research interests are Image Processing and Artificial Neural Networks.

Development Strategy using Cognitive Domain in e-Requirement Engineering Learning System

Nor Azliana Akmal Jamaludin[1], Shamsul Sahibuddin[2]

[1] Faculty of Computer Science and Information System, Universiti Teknologi Malaysia, UTM Skudai, 81310 Johor, Malaysia

[2] Advance of Informatics System, Universiti Teknologi Malaysia, UTM Kuala Lumpur, 54100 Wilayah Persekutuan, Malaysia

Abstract

Current trend of e-learning promote continuous learning environment. Unfortunately, it fails to optimize the learning between student and learner. Some factors are discussed to encounter the current problem. It should have suitable contents representations that will be remembered and applied in practice by the students. The e-learning discussed in context of Requirement Engineering domain. Maximize the usage of e-Requirement Engineering Learning System, the cognitive domain is suggested to cooperate with the e-learning system. It will help Higher Learning Education (HLE) deliver students with employability skill with critical thinking strategy. In addition, it gives a big impact to software development project. Future work will be discussed on the quantitative analysis to measure the effectiveness of e-Requirement Engineering Learning System.

Keywords: *Requirement Engineering, cognitive Domain, Higher Learning Education, software development project, employability skill, problem-solving, bloom's taxonomy.*

1. Introduction

Developed country is boost with learning in virtual environment. Then, developing country with high effort make real a virtual learning in Higher Learning Education (HLE) to become a same level with developed country is undeniable fact [13, 24]. Most of the HLE [32] probably has the e-learning in their website. Unfortunately, the e-learning content different among others HLE in Malaysia. Sometimes, the e-learning website is not fully utilizing the needs for students learning [1, 2]. It make the e-learning is less effective. The improvement of e-learning should be growth fast [14] to deliver knowledgeable students with employability skill [35]. It should be concurrent with the industrial needs [22].

The concept of designing an e-learning [7, 25] should be similar with new car engine. The new engine is design with high performance to make a car move smoothly. At the same time, the car is very useful to the driver. To that reason, developing e-learning with cognitive domain is a way to improve the virtual learning between student and lecturer [3, 31].

2. Current Trend of e-Learning

Traditional learning promotes more on face-to-face environment [4, 26]. The constraint of face-to-face is based on limited of time and place for students to have guidance from the lecturer. They should come to the class as per schedule and only have the consultation hour by appointment. The learning is growing slowly. Especially, in Requirement Engineering subject, the technological tool is rapidly changes and the students unable to have continuously learning environment.

E-learning is taking place [15] to offer continuously learning environment with virtual medium. The internet is utilized by the students and lecturer to commit into e-learning environment. However, if e-learning does not have suitable contents and approaches, then e-learning will not enhance the learning entirely [17]. The success of e-learning depends on it being 'brain friendly' [8]. It defines that the students can totally optimize learning by e-learning contents representations that will be remembered and applied in practice [18].

3. Development of e-Requirement Engineering Learning System

Requirement Engineering [34] is a first step in controlling the whole software development project [20, 21] run successfully and meets stakeholder needs. Understanding the process and practice skill will help student to be a demanding employee. Somehow, the problems occur in traditional class which communication with the lecturer and stakeholder become a barrier to the student in finishing their task [33].

To that reason, some researchers have predicted that the traditional classroom will disappear [6, 9]. E-learning has entered the education [12] as well as the corporate world in a major way and it also complements the traditional delivery methods. It has facilitated the traditionally difficult educational paradigms such as adult learning or distance learning.

e-Requirement Engineering Learning System come across to provide a versatile learning environment, which many companies realize that educational institutions should be a kick start in deliver a student with employability skill [27, 38]. It is a combination of e-learning and Requirement Engineering. There are several factors are identified in implementing successful collaboration between cognitive domain [16] that cooperate with e-Requirement Engineering Learning system.

Development of web technology over the last two decades has led to a productive blooming of e-Learning methodology [18]. E-Learning materials and environments have several identifying characteristics [36, 37]: 1) availability – the learning materials are available on the net and students can easily download them; 2) multiple representations – the learning materials combine text, graphics, animation, sound and video; 3) multiple communication tools – several specially-developed social tools support e-Learning such as: discussion groups, e-mail, video conference, blogs and social networks. E-Learning instructors at the beginning of twenty-first century rely on this environment to implement various pedagogies.

Currently, e-Learning had been implemented but not favorably used [18]. Haverila and Barkhi (2011) identified that more experienced students felt e-learning is less effective than the less experienced students in using e-learning. When the students lack the experience in e-learning, it is important that the procedures, software tools, and materials are well organized and expectations are explained in detail before the course starts.

3.1 Infrastructure Factor

Lack of infrastructure support will lead to e-learning failure. In organization, e-learning can help in employees development in term of consistency of message, flexibility of learning, availability of learning and monitoring of progress and performance. Technology in e-learning [28] known as web-based learning is another channel of learning, which provides 1) Real time communication (Synchronous technology) such as instant message, audio/video conferencing and 2) Anywhere and anytime (Asynchronous technology) to access the courses over the Internet or intranet [19, 29].

Table 1. The Influence of Experience, Ability and Interest on e-learning Effectiveness (Haverila, M. & Barkhi, R., 2011).

		Place	
		Same	Different
Time	Synchronous	Traditional Method	Distance Learning
	Asynchronous	Recorded	e-Learning

3.2 Students / Learner Factor

The other factors contribute in e-learning failure is individual perspective. It includes self-motivation among students. The e-learning level decrease if the student gets difficult to use e-learning if they don't have technological skill [22, 35].

Responsible (self-directed leaning) on finishing the task given by the lecturer is another crucial factor [14]. Better learner attitudes contribute to increase learner's control [17].

4. Element in Learning Environment

There are several arguments and similarity from three well-known researchers about the element in successful learning environment. According to Bloom's taxonomy (1956) of learning domains, there are three domains of educational activities 1) Cognitive Domain, which involves knowledge and the development of intellectual and mental skills, 2) The Affective Domain, which describes the way we face things emotionally and 3) The Psychomotor Domain, which involves physical movement, coordination, and use of the motor-skills.

In addition, Parnas (1999) encouraged Software Engineering curriculum to involve with five complementary elements in learning environment. It includes 1) principles (lasting concepts that underlie the whole field), 2) practices (problem-solving techniques which good professionals will apply consciously and regularly), 3) applications (areas of expertise where the principles and practices find their best expression), 4) tools (state-of-the-art products that facilitate the application of these principles and practices) and 5) mathematics (the formal basis that makes it possible to understand everything else).

Furthermore, Jazayeri (2004) states that learning environment should integrate with projects. This is often recognized as a very critical issue in Software Engineering education. Replaying the complexity of real-life projects in an educational environment can be impossible. Thus we need to find innovative ways of integrating project [20, 21] work in curricula.

Table 2. Similarity element of successful learning.

Author	Year	Element (s)	Similarity on
Bloom's taxonomy	1956	1. Cognitive Domain 2. Affective Domain 3. Psychomotor Domain	• Cognitive Domain
Parnas	1999	1. Principles 2. Practices 3. Application 4. Tool 5. Mathematics	• Practices • Application • Tool
Jayazeri	2004	1. Project	• Project

5. Design e-Learning with Cognitive Domain Perspectives

Zamfir (2007) indicated that "cognitive domain, focused on learning about learning, both as individual and as organization, imply information technology and two effects of it: diversity and globalization".

Figure 1 shows that some factors influence in designing [7, 10, 11] the software towards cognitive domain. It is particularly based on 1) individual (refer to student, lecturer and stakeholder (who contribute in learning environment), 2) team (is among student's team members) and 3) project (refer to students team, lecturer (monitor student's project progress), and stakeholder (who will approve the requirement deliverables) factors.

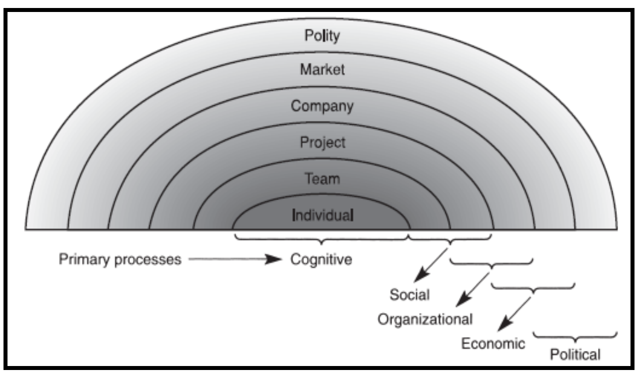


Figure 1. Factors influencing the software design process (Curtis *et al*., 1988)

Cognitive Strategy used as an internal process by which the student [37] controls their own ways of thinking and learning. Example: Engaging in self-testing to decide how much study is needed; knowing what sorts of questions to ask to best define a domain of knowledge; ability to form a mental model of the problem [10]. Problem solving is combining lower level rules to solve problems in a situation never encountered by the person solving the problem. It may involve generating new rules which receive trial and error use until the one that solves the problem is found. Bloom (1956) described six sub-categories in the cognitive domain, which are measured by degrees and levels of difficulties so that an individual cannot master one of these levels if the student has not first mastered the preceding sub-category.

6. Conclusion

As a conclusion, the e-Requirement Engineering Learning System should integrate with Cognitive Domain to increase problem-solving skill among student. The students should have a responsibility in order to finish their task according to the milestone given by the lecturer.

Table 3 discussed the suggestion of assessment method regarding the Cognitive Domain (Blooms, 1956) and Bloom's Revised Taxanomy (Pohl's, 2000). Several top universities such as Harvard [40] and Cambridge [41] University were used quiz as a medium for students to memorize learned information.

Other assessment [23] methods (case study, small group assignment and project) have collaboration between students, lecturer and stakeholder [20]. It should be an important element in an interactive and process-oriented course. It will facilitate a collaborative learning context to improve critical thinking skills. Otherwise, projects [21] are used to tailor with the needs of employability skill [33]. The skill of presentation can be improved by using peer-to-peer assessment of assignments and conferencing in a team environment.

In addition, the expectations should be clear with proper milestones and set it in specific time frames. Students should be informed that the deadlines and schedules should be taken seriously [18]. Ada (2009) highlight a positive correlation between the quality of the group's engagement in a collaborative process and the quality of cognitive skills fostered. Future work will be discussed on the challenges in developing e-Requirement Engineering Learning System.

Table 3: Suggestion of e-Requirement Engineering Learning System based on Cognitive Domain (Blooms, 1956) and Bloom's Revised Taxanomy (Pohl's, 2000).

Cognitive Domain	**Bloom's Revised Taxonomy**	**Suggestion assessment method e-Requirement Engineering Learning System**	**Participant**
Knowledge: Recall data or information	**Remembering**: Recall previous learned information.	**Quiz / Exercise / Short Essay**	**Student**
Comprehension: Understand the meaning, translation, interpolation, and interpretation of instructions and problems. State a problem in one's own words	**Understanding**: Comprehending the meaning, translation, interpolation, and interpretation of instructions and problems. State a problem in one's own words.	**Exercise / Case Study**	**Student / Lecturer**
Application: Use a concept in a new situation or unprompted use of an abstraction. Applies what was learned in the classroom into novel situations in the work place.	**Applying**: Use a concept in a new situation or unprompted use of an abstraction. Applies what was learned in the classroom into novel situations in the work place.	**Case Study / Assignment**	**Student / Lecturer / Industry**
Analysis: Separates material or concepts into component parts so that its organizational structure may be understood. Distinguishes between facts and inferences.	**Analyzing**: Separates material or concepts into component parts so that its organizational structure may be understood. Distinguishes between facts and inferences.	**Small Groups Assignment / Project**	**Student / Lecturer / Industry**
Synthesis: Builds a structure or pattern from diverse elements. Put parts together to form a whole, with emphasis on creating a new meaning or structure.	**Evaluating**: Make judgments about the value of ideas or materials.	**Project / Presentation**	**Student / Lecturer / Industry**
Evaluation: Make judgments about the value of ideas or materials.	**Creating**: Builds a structure or pattern from diverse elements. Put parts together to form a whole, with emphasis on creating a new meaning or structure.	**Project / Presentation**	**Student / Lecturer / Industry**

References

[1] Ali, E., G., Magalha, & R. (2008). Barrier to implementing e-learning: a Kuwaiti case study. International Journal of Training and Development , pp. 36-52.

[2] Annansingh, F., & Bright, A. (2010). Exploring barriers to effective e-learning: case study of DNPA. IADIS e-learning Conference, pp. 55-65. Algarve, Portugal: Interactive Technology and Smart Education.

[3] B. S. Bloom and D. R. Krathwohl. (2003). Taxonomy of Educational Objectives: The Classification of Educational Goals. INFORMATICA, Vol. 2, no. 1.

[4] Bliuc, AM. (2011). A blended learning Approach to teaching foreign policy: Student experiences of learning through face-to-face and online discussion and their relationship to academic performance. Computer & Education. Vol.56,Iss.3;p.856-864. ISI Proceeding.

[5] Bloom B. S. (1956). Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain. New York: David McKay Co Inc.

[6] Blustain, H., Goldstein, P. and Lozier, G. (1999). "Assessing the New Competitive Landscape" in Dancing with the Devil, Editors: Richard N. Katz. Associates, Jossey-Bass Publishers, San Francisco.

[7] Budgen, D. (2003). Software Design. Second edition. Pearson Addsion Wesley.

[8] Dror, I. E. (2011). Brain friendly technology: What is it? And why we need it? In I. E. Dror, Technology Enhanced Learning and Cognition. John Benjamins, Amsterdam. Vol. 33, No. 4 , Pages 291-296.

[9] Drucker, P. (1997). "An Interview with Peter Drucker" Forbes Magazine, March 10.

[10] Gagné, R.M. and Briggs, L.J. (1974). Principles of instructional design (2nd ed.). Holt, Rinehart, and Winston.

[11] Gea, J. M. C. d., J. Nicol, Alemán, J. L. F., Toval, A., Ebert, C, and Vizcaíno, A. (2011). Requirements Engineering Tools. IEEE Softw. Vol. 28, No. 4. pp: 86-91.

[12] Gilbert, J. (2007). E-learning: The student experience. British Journal of Educational Technology, 38(4), 560–573..

[13] Grigoriadou, M., Papanikolaou, K., Tsaganou, G., Gouli, E. & Gogoulou, A. (2010). Introducing innovative e-learning environments in higher education. International Journal of Continuing Engineering Education and Life-Long Learning (1560-4624). Vol.20,Iss.3-5; p.337-355

[14] Guglielmino, P. J., & Guglielmino, L. M. (2001). Moving towards a distributed learning modelbased on self-managed learning. SAM Advanced Management Journal, Summer . pp.36–43.

[15] Hanafizadeh, P., Khodabakhshi, M & Hanafizadeh, M.R. (2011). Recommendations for promoting e-learning in higher education institutions: A case study of iran. Higher Education Policy. Vol.24,Iss.1;p.103-126. Scopus.

[16] Handbook I: Cognitive Domain. Longmans, New York, 1956.

[17] Hannafin, M. J., & Land, S. M. (1997). 'The foundations and assumptions of technology enhanced student-centered learning environment'. Instructional Science, Vol. 25 , pp.167–202.

[18] Haverila, M. & Barkhi, R. (2011). The Influence of Experience, Ability and Interest on e-learning Effectiveness, European Journal of Open, Distance and e-Learning.

[19] Huffaker, D. (2003) Reconnecting the classroom: E-learning pedagogy in US public high schools. Australian Journal of Educational Technology, 19(3), 356-370.

[20] Jamaludin, N. A. A., Sahibuddin, S. & Hidayat, N. H. (2011). Challenges of Project-Based Learning Towards Requirement Engineering. The 10th WSEAS International Conference on Software Engineering, Parallel and Distributed Systems (SEPADS '11)".

[21] Jamaludin, N. A. A., Sahibuddin, S., Jusoff, K., Hidayat, N. H. (2010). Development of a Project-Based Learning

Approach in Requirement Engineering. International Journal of Computer Science and Information Security, IJCSIS, Vol. 8, No. 9, pp. 6.

[22] Kalling. (1999). 'Gaining competitive advantage through information technology', A resource-based approach to the creation and employment of strategic it resources. Doctoral Disseratation. Lund University. Lund Business Press.Lund.

[23] Kim, S. W., & Lee, M. G. (2007). Validation of an evaluation model for learning management systems. Journal of Computer Assisted Learning, 24(4), 284–294.

[24] Kocur,D. & Kosc, P. (2009).e-Learning Implementation in Higher Education. Acta Electrotechnica et Informatica Vol. 9, No. 1, 20-26

[25] Li, Y.; Li, T.; Liu, C. (2010). Designing the e-learning systems of China's higher education: A case study of statistics. Proceedings - 2010 International Conference on Artificial Intelligence and Education, ICAIE 2010. p.640-643. Scopus.

[26] Loftus, M. (2011). Face-to-Face Learning. Computer Graphic World. Vol.34;p.44-46. ISI Proceedings

[27] Marks, A. and Scholarios, D. (2008). Choreographing a system: Skill and employability in software work. Economic and Industrial Democracy Vol. 29, no. 1, pp: 96-124.

[28] Martinez-Torres, M. R., Toral Marı´nb, S. L., Garciab, F. B., Vaizquezb, S. G., Oliva, M. A., & Torres, T. (2008). A technological acceptance of e-learning tools used in practical and laboratory teaching, according to the European higher education area. Behaviour Information Technology, 27(6), 495–505.

[29] Okiki, C. O. (2011). Information communication technology support for an e-learning enviroment at the University of Lagos, Nigeria. Library Philosophy and Practice Academic OneFile. Web. 14 Aug. 2011.

[30] Parnas, D. L. 1999. Software Engineering Programs Are Not Computer Science Programs. IEEE, 16 (6), 19-30.

[31] Pohl, M. (2000). Learning to Think, Thinking to Learn: Models and Strategies to Develop a Classroom Culture of Thinking. Cheltenham, Vic.: Hawker Brownlow.

[32] Saekow, A. & Samson, D. (2011). A Study of E-learning Readiness of Thailand's Higher Education Comparing to the United States of America (USA)'s Case. Computer Research and Development (ICCRD), 2011 3rd International Conference on 2011. Vol. 2, pp: 287-291.

[33] Sikora, E., Tenbergen, B. and Pohl, K. (2011). Requirements engineering for embedded systems: an investigation of industry needs. In *Proceedings of the 17th international working conference on Requirements engineering: foundation for software quality* (REFSQ'11), Daniel Berry and Xavier Franch (Eds.). Springer-Verlag, Berlin, Heidelberg, pp: 151-165.

[34] Silhavy, R., Silhavy, P. and Prokopova, Z. (2011). Requirements gathering methods in system engineering. In *Proceedings of the 13th WSEAS international conference on Automatic control, modelling & simulation* (ACMOS'11), World Scientific and Engineering Academy and Society (WSEAS), USA. pp: 106-110.

[35] Sridharan, B., Deng, H. & Corbitt, B. (2010). The perceptions of learners on the effectiveness of e-learning in higher education. ECIS 2010 Proceeding.

[36] Tzouveli, P., Mylonas, P. & Kollias, S. (2007). An intelligent e-learning system based on learner profiling and learning resources *adaptation.*. Computers & Education. vol. 51, p: 224–238.

[37] Ulf-D. E. (2004). Quality in e-Learning from a Learner's Perspective. Best Paper Award at the Third EDEN Research Workshop 2004, Oldenburg, Germany. European Journal of Open, Distance and e-Learning. the 2010 International Conference on Education Technology and Computer, New Jersey, USA, 22-24 June, pp. 167-171.

[38] Y. M. Yusoff, M. Z. Omar, A. Zaharim, A. Mohamed, N. Muhamad, & R. Mustapha. (2010). Enhancing employability skills through industrial training programme. In *Proceedings of the 7th WSEAS international conference on Engineering education* (EDUCATION'10), World Scientific and Engineering Academy and Society (WSEAS), pp: 398-403.

[39] Zamfir, G. (2007). Information Technology and The cognitive domain. Journal of Applied Quantitative Methods, Vol. 2, No. 4 .

[40] Harvard University Website. Retrieved from http://www.harvard.edu/.

[41] Cambridge University. Retrieved from http://www.cam.ac.uk/

[42] Ada, W.W. (2009). Computer supported collaborative learning and higher order thinking skills: A case study of textile studies. *Interdisciplinary Journal of E-Learning and Learning Objects, 5*, 145-167.

Nor Azliana Akmal Jamaludin is a Lecturer for Bachelor of Science Software Engineering and Head of Developer for Master Degree Software Engineering program at Universiti Industri Selangor. She received the Master Degree in Computer Science (Real-Time Software Engineering) from Advanced Informatics School (formerly known as Centre for Advanced Software Engineering (CASE), Universiti Teknologi Malaysia, in 2004. Currently, she is doing her Doctorate in Computer Science, specialize in Software Engineering at UTM and supervised by Prof. Dr. Shamsul Sahibuddin. She is a member of Malaysian Software Engineering Interest Group, Malaysia. Nor Azliana Akmal has been active in scholarly conference proceedings-journals writing and publishing. She has published more than 3 citation index/impact factor conference proceeding-journal papers. Her field of expertise is in software requirement, analysis, system integration, and e-learning. Her current research interest is on techniques that can enhance skill among Software Engineering undergraduate of higher institutions in Malaysia using machine learning system.

Shamsul Sahibuddin received the PhD in Computer Science at Aston University, Birmingham, United Kingdom in 1998 and Master Science (Computer Science) at Central Michigan University, Mt. Pleasant, Michigan in 1988. He is currently a Dean / Professor at Advanced Informatics School, Universiti Teknologi Malaysia (UTM) and has been a Member of the Program Committee for Asia-Pacific Conference on Software Engineering. His field of expertise is in software requirement, software modelling technique, software development and software process improvement.

An Image Compression Approach using Wavelet Transform and Modified Self Organizing Map

G Boopathi[1],
Assistant Professor Of Computer Applications,
SNR Sons College(Autonomous),
Coimbatore 641006, India

Dr.S.Arockiasamy[2],
Asst. Dean for Training
Head, Information Systems
University of Nizwa,
Sultanate of Oman

Abstract

Image compression helps in storing the transmitted data in proficient way by decreasing its redundancy. This technique helps in transferring more digital or multimedia data over internet as it increases the storage space. It is important to maintain the image quality even if it is compressed to certain extent. Depends upon this the image compression is classified into two categories: lossy and lossless image compression. There are many lossy digital image compression techniques exists. Among this Wavelet Transform based image compression is the most familiar one. The good picture quality can be retrieved if Wavelet-based image compression technique is used for compression and also provides better compression ratio. In the past few years Artificial Neural Network becomes popular in the field of image compression. This paper proposes a technique for image compression using modified Self-Organizing Map (SOM) based vector quantization. Self-Organizing Feature Map (SOFM) algorithm is a type of neural network model which consists of one input and one output layer. Each input node is connected with output node by adaptive weights. By modifying the weights between input nodes and output nodes, SOFM generate codebook for vector quantization. If the compression is performed using Vector Quantization (VQ), then it results in enhanced performance in compression than any other existing algorithms. Vector Quantization is based on the encoding of scalar quantities. The experimental result shows that the proposed technique obtained better PSNR value end also reduces Mean Square Error.

***Keywords*—**Data Compression, Image Compression, Neural Networks, Self-Organizing Feature Map, Vector Quantization, Wavelet Transform

1. Introduction

Image compression is a result of applying data compression to the digital image. The main objective of image compression is to decrease the redundancy of the image data which helps in increasing the capacity of storage and efficient transmission. Image compression aids in decreasing the size in bytes of a digital image without degrading the quality of the image to an undesirable level. There are two classifications in image compression: lossless and lossy compression. The reduction in file size allows more images to be stored in a given amount of disk or memory space. This supports in decreasing the time required for the image to send or download from internet. Consequently compression methods are being hastily developed to compress large data files such as images, where data compression in multimedia applications has recently become very important [1].

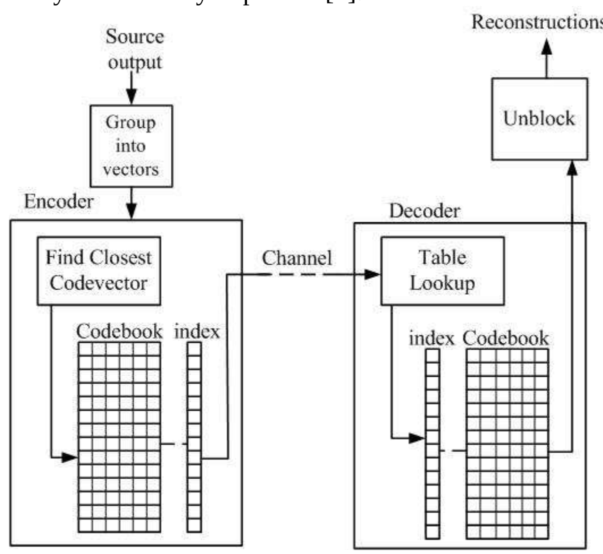


Figure 1.Vector Quantization

In general wavelets are a mathematical tool for hierarchically decomposing functions. The huge numbers of lossy compression techniques are proposed in the past. Among this Wavelet Transform based image compression is the most familiar one. Wavelet-based image compression provides better enhancements in picture quality even at higher compression ratios. It is an established transform used for a number of image compression standards in lossy compression methods. Divergentto the discrete cosine transforms, the wavelet transform is not Fourier-based and therefore wavelets do a superior job of handling discontinuities in data. Wavelet Transforms (WT) based image compression is a prevailing method that is favored by most of the researchers to get the compressed images at higher compression ratios with higher PSNR values [2].

The usage of Artificial neural network (ANN) in image processing applications has been increased in recent years. Due to the advantages over the existing methods in terms of handling the noisy or partial data, the Artificial Neural Networks can be used in image compression technique. An Artificial Neural network is appropriate technique for image compression as it has the ability to reproduce the original data with the help of available fewer components. Different types of Artificial Neural Networks have been trained to perform Image Compression. Some of them are Feed-Forward Neural Networks, Self-Organizing Feature Maps and Learning Vector Quantizer Network

Artificial neural networks are well-resembled in function approximation, owing to their capability to fairly accurate complicated nonlinear functions. Several techniques have been projected previously for image compression using neural networks and wavelet transform. In the past wavelet transform and a neural network are suggested for image compression [4].Similarly, variety of image compression techniques were combined with neural network classifier for various applications [5] [6].Some recent papers show that the combination of neural network based approach and classical wavelet based approach leads to better compression ratio [7]. Combining the Wavelet Transform and Artificial Neural Networks utilizes the advantages of the two techniques thereby improving the compression ratio. They may also ensure the quality of the compressed image.

The modified Self-Organizing Feature Map (SOFM) based vector quantization for image compression is proposed in this paper. Self-Organizing Feature Map (SOFM) algorithm is a type of neural network model which consists of one input and one output layer. Each input node is connected with output node by adaptive weights. By modifying the weights between input nodes and output nodes SOFM will generate codebook for vector quantization. If the compression is performed using Vector Quantization (VQ), then it results in enhanced performance in compression than any other existing algorithms. Vector Quantization is based on the encoding of scalar quantities. The experimental shows that the proposed technique will provide better PSNR value and also reduces Mean Square Error.

The remainder of this paper is organized as follows. Section 2 of the paper discusses the earlier proposed techniques related to image compression using wavelet transform and neural networks. Section 3 explains the proposed approach for image compression. Section 4 illustrates the experimental results with relevant explanations and Section 5 concludes the paper with fewer discussions for future work.

2. Related Work

A lot of works were found in literature related to the wavelet based image compression technique using the neural network technique. This section of the paper discusses some of the earlier work proposed on image compression using neural networks and wavelet transform.

Debnath et al., proposed an image compression method combining discrete wavelet transform (DWT) and vector quantization (VQ). First, a three-level DWT is carried out on the original image resulting in ten separate subbands (ten codebooks are generated using the Self Organizing Feature Map algorithm, which are then used in Vector Quantization, of the wavelet transformed subband images, i.e. one codebook for one subband). These subbands are then vector quantized. VQ indices are Huffman coded to raise the compression ratio. A new iterative error correction scheme is presented to continuously check the image quality after sending the Huffman coded bit stream of the error codebook indices through the channel so as to improve the peak signal to noise ratio (PSNR) of the reconstructed image. Ten errors are also generated for the error correction method by means of the difference between the original and the reconstructed images in the wavelet domain. This technique shows better image quality in terms of PSNR at the same compression ratio as compared to other DWT and VQ based image compression techniques found in the literature.

A method of still image compression was put forth by Wilford Gillespie in [11]. The fundamental approach to image compression consists of a number of key steps. They are wavelet packet decomposition, quantization, organization of vectors, neural networks approximation or storage, and lossless encoding and reduction. As an initial stage of image compression, the image is put through several layers of wavelet packet decomposition. The results of the decomposition are then divide or processed in some way, depending on the method. Integer quantization is performed on all of the decomposed wavelet sections. The quantization value is the determining factor of quality. A quantization value of 1 is near lossless quality, although little to no compression is achieved. This is accomplished by taking each section and dividing it by a set value and rounding to the nearest integer. There are many ways to systematize a tree of decomposition sections. Three methods were tried with this compression scheme. The type of neural network used in their approach was a two-layer feed-forward network with a standard back propagation learning function. At last, the entire data stream is taken and is processed by a run length encoded (RLE) method and saved in a lossless state using the ZIP file format.

A Neuro-Wavelet based approach for image compression was put forth by Singh et al. in [12]. Images have large data quantity. For storage and transmission of images, high efficiency image compression methods are under wide attention. They proposed a neuro- wavelet based model for image compression which combines the advantage of wavelet transform and neural network. Images are decomposed using wavelet filters into a set of sub bands with different resolution corresponding to different frequency bands. Different quantization and coding schemes are used for different sub bands based on their statistical properties. The coefficients in low frequency band are compressed by differential pulse code modulation (DPCM) and the coefficients in higher frequency bands are compressed using neural network. Using their proposed scheme one can accomplish satisfactory reconstructed images with large compression ratios. Their experimental results revealed that their proposed technique of image compression out performed some of the conventional image compression approaches.

Barbalho et al., [13] presented a novel approach involving vector quantization (VQ) that relies on the design of a finite set of codes which will substitute the original signal during transmission with a minimal of distortion, taking advantage of the spatial redundancy of image to compress them. Algorithms for instance LBG and SOM work in an unsupervised way toward finding a good codebook for a given training data. However, the number of code vectors (N) required for VQ.

increases with the vector dimension, and full-search algorithms such as LBG and SOM can lead to large training and coding times. An alternative for reducing the computational difficulty is the use of a tree-structured vector quantization algorithm. This approach presents an application of a hierarchical SOM for image compression which reduces the search complexity from O(N) to O(log N), enabling a faster training and image coding. Results when compared with conventional SOM, LBG and HSOM, shows the better image compression result.

Amar et al. in [14] proposed a wavelet networks approach for image compression. Wavelet networks are a combination of radial basis function (RBF) networks and wavelet decomposition, where radial basis functions were replaced by wavelets. The wavelet network is a combination of wavelets and neural networks. The network can be considered composed of three layers: a layer with N_i inputs, a hidden layer with N_w wavelets and an output linear neuron receiving the weighted outputs of wavelets. Both input and output layers are fully connected to the hidden layer. Moreover they used a feed forward propagation algorithm from input neurons to output neurons. The main similarity between the proposed wavelet network and the neural network is that both networks calculate a linear combination of nonlinear functions to adjust parameters. These nonlinear functions depend on adjustable parameters (dilations and translations). During training stage the weights, dilations and translations parameters, are iteratively adjusted to minimize the network error. They used a quadratic cost function to evaluate this error. In order to test the robustness of their approach, they have implemented and compared the results with some other approaches based on neural networks (MLP).

Kwang-Baek et al., [15] puts forth a novel vector quantization approach for image compression using wavelet transform and enhanced SOM algorithm for medical image compression. The enhanced self-organizing algorithm is presented to improve the defects of SOM algorithm, which, at first, reflects the error between the winner node and the input vector to the weight adaptation by using the frequency of the winner node. Secondly, it adjusts the weight in proportion to the weight change at hand and the previous weight change as well. To decrease the blocking effect and Improve the resolution, by using wavelet transform the vectors are constructed and applied the enhanced SOM algorithm to them.

Khashman et al. in [16] proposed a technique for compressing the digital image using neural networks and Haar Wavelet transform. The aim of the work presented within the paper was to develop an optimum image compression system using Haar wavelet transform and a neural network. With Wavelet transform based compression, the quality of compressed images is typically high, and the option of a perfect compression ratio is complicated to formulate as it varies depending on the content of the image. They proposed that neural networks can be trained to ascertain the non-linear relationship between the image intensity and its compression ratios in search for an optimum ratio. Moreover their paper suggested that a neural network could be trained to be familiar with an optimum ratio for Haar wavelet compression of an image upon presenting the image to the network. The method utilized Haar compression with nine compression ratios and a supervised neural network that learns to correlate the grey image intensity (pixel values) with a single optimum compression ratio. Two neural networks receiving different input image sizes are developed in their work and a comparison between their performances in finding optimum Haar-based compression was presented.

3. Proposed approach

The proposed methodology deals with the combination of wavelet and vector quantization for image compression. The image compression technique proposed here is applicable to those areas of digital images where high precision reconstructed image is required like criminal investigations, medical imaging, etc., The image of certain quality is need to be transmitted by user in order to retrieve the original image without any loss in quality. This method is tested on gray scale images, but it can be easily extended to color images by processing the three color matrices separately.

A. Self-Organizing Map

Each data from data set recognizes themselves by competing for representation. The weight vectors initialization is the starting process of SOM mapping. Then the sample vector is randomly selected and the map of weight vectors is searched to find which weight best represents that sample. Each weight vector has neighboring weights that are close to it. The weight that is chosen is rewarded by being able to become more like that randomly selected sample vector. The neighbors of that weight are also rewarded by being able to become more like the chosen sample vector. From this step the number of neighbors and how much each weight can learn decreases over time. This whole process is repeated a large number of times, usually more than 1000 times.

In sum, learning occurs in several steps and over much iteration:

1. Each node's weights are initialized.
2. A vector is chosen at random from the set of training data.
3. Every node is examined to calculate which one's weights are most like the input vector. The winning node is commonly known as the Best Matching Unit (BMU).
4. Then the neighborhood of the BMU is calculated. The amount of neighbors decreases over time.
5. The winning weight is rewarded with becoming more like the sample vector. The neighbors also become more like the sample vector. The closer a node is to the BMU, the more its weights get altered and the farther away the neighbor is from the BMU, the less it learns.
6. Repeat step 2 for N iterations.

B. Modified Self Organizing Feature Map

The basic operation of the Kohonen's network is to classify the input patterns with a set of m× n weight matrix where m is the number of nodes in input layer and n is the grid size. Existing learning system considers the previously learned patterns while adopting the weight matrix for the current input pattern that is avoided by the proposed subsystem. The modification in existing learning system is highlighted below –

1. *Adoption of Weights*: Existing learning systems deals with the previously stored patterns which are already been learnt. It increases the learning time exhaustively. Learning time for each pattern is a factor of the number of previously learned patterns. But the modified system only tries to operate on the recently given pattern sample. It avoids the previously learned patterns for the swiftness of learning process.

2. *Regular learning system* using KSOM offers the modification of weights for all the connections among the two layers. It indicates the static size of neighbors. Due to the rapid change of neighborhood size, number of weight adoption easily decreased with the time. The modified MKSOM system proposes a function for changing the neighborhood size along with the change of the distance of winner node

$\nu(t+1) = \nu(t) - 0(t)(d(t) - d(t-1))$.

A 3-level 2-D DWT is firstly applied to the test image in the proposed method (i.e. the image to be compressed) and then VQ is used to different subbands for compression. Ten subbands are created after the application of 3-level 2-D DWT using SOFM, and thus all these codebooks are used for this all subbands individually. 3-level 2-D DWT is applied to images because the low frequency subband, which contains the maximum energy content of the original image, becomes of smaller size so that in case of vector quantization this subband is treated with a codebook size of 7-bits only. These vector indices are subjected to Huffman coding [6] for improving the compression ratio of the transmitted data. Whole compression process of this work is divided into three steps, i) Codebook generation, ii) Encoding of the original image and iii) Decoding of the image. All of these steps will now be discussed. The proposed method uses a total of twenty codebooks, ten codebooks for original image reconstruction and other ten are used to reconstruct the error images.

Algorithm

Step 1:

Initial image= Input image;
Input image→10 sub-images using 3-Level DWT;
For (each of 10 image)
    Vector quantization using separate codebook foreach subband;
    Codebook indices are now Huffman coded and then transmitted to the decoder;
End;
If (PSNR < Threshold T_h*)*
    Move to step 2;
Else
    Stop;
End;

Step 2:
At the encoder end
Initial image (I.I) = obtained subbands;
For (each image)
Subbands are reconstructed→ this is the reconstructed image (R.I);
End;
goto step 3;
Step 3:
Image Error (I.E) = Difference between I.I and R.I;
These I.E's (Ten error subbands)→ Vector Quantized using the error codebooks (Ten different error codebooks used);
Error Codebook →Huffman coding →transmitted to the decoder;
Again at the encoder end
Using these error codebooks I.E's are reconstructed = R.I.E;
R.Inew=R.I + R.I.E; \\ Recalculate the new Reconstructed image
If (PSNR<Th)
If No. of Iteration < 3
Goto step 3;
Else
Stop;
End;
Stop;
End;

C. Codebook Generation

In the codebook generation step (i.e. the training stage) four different standard images (namely Lena, Couple, Frog, and Baboon) are used to generate ten original codebooks and also ten error codebooks are generated in this step.3-level 2-D DWT is applied to each of these original training images in all ten codebook generation step. These generate ten wavelet sub bands for each of the original images. Similar sub bands of each image are then combined to form a single frame and this frame is then considered as a new image. Therefore there are ten separate images available at this stage. Using these ten separate images, ten separate codebooks are generated using SOFM. Then in the error codebook generation step, using these generated ten codebooks ten sub band images are vector quantized and then these sub bands are reconstructed. These ten reconstructed images are then compared with the original ten images in the wavelet domain; the error of this comparison was taken to generate the error codebooks. In this case SOFM or Modified SOFM is used.

D. Encoding and Decoding

In this step, 3-level 2-D DWT is applied to the test image (i.e. the image to be compressed). Then each of these available ten subbands is vector quantized using the original codebooks, so that separate codebook is used for different sub bands. The codebook indices of this VQ process are transmitted to the decoder after Huffman coding. At the encoder end image is reconstructed using the transmitted image indices and peak signal to noise ratio (PSNR) of this transmitted image is calculated to test the image quality. If the calculated PSNR is higher or equal to the desired PSNR then the process ends, otherwise the iterative error correction method is executed. In this iterative error correction method, error between the original image and the reconstructed image (I.E), is calculated in the wavelet domain. Vector quantization using the available error codebooks is then applied to these subband errors between the original and reconstructed image (R.I.). Error codebook indices are also transmitted to the decoder after Huffman coding. The transmitted error image is reconstructed from the transmitted error codebook indices (at the encoder or transmission end). Then the reconstructed image errors (R.I.E) are added (algebraically) to the previously reconstructed image, and thus R.I. is modified. This iterative error correction process continues until the PSNR of the modified reconstructed image is larger than or equal to the desired PSNR or the maximum number of iteration (considering the case of infinite loop, the iteration process stopped by force at the third iteration) is reached. In the decoding phase the decoder first receives the Huffman coded bit-stream of the VQ indices corresponding to the original wavelet coefficients from the channel. It then reconstructs the codebook indices of the different wavelet sub bands. In the initial stage the receiver receives the reconstructed image and successively in the later steps the receiver receives image errors (actually it receives Huffman coded image errors, and reconstructs the image error coefficients from these Huffman coded indices). The receiver adds (algebraically) the received errors of each sub band. In the final step the image is reconstructed using 3-level inverse 2-D DWT.

4. Experimental Results

In order to evaluate the performance of the proposed approach of image compression using modified SOM algorithm based vector quantization two standard images are considered. The work is implemented using MATLAB. Lena and Cameraman are the two standard images used to explore the performance of the proposed approach of image compression. The experiments are carried out with the number of clusters of 4, 8, 16, 32 and 64. The evaluation of the proposed approach in image compression was performed using the following measures,

$$bpp = \frac{\text{Encoded number of bits}}{\text{Number of Pixels}}$$

$$PSNR = 10\log_{10}\left[\frac{255^2}{MSE}\right](dB)$$

$$MSE = \frac{1}{3MXN}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}\sum_{k=1}^{3}\left\{X_k(j,i)-\overline{X}_k(j,i)\right\}^2$$

These three factors will decide about the image noise ratio, retrieval quality and ratio of compression for the digital image. The PSNR is most familiarly used as a measure of quality of reconstruction of lossy image compression. The MSE (Mean Square Error) is the cumulative squared error between the compressed and the original image, whereas PSNR is a measure of the peak error.

The experimental results that evaluate the performance of the proposed approach by comparing it with the self-organizing map are tabulated. Table 1 shows the experimental results applied for Lena image and table 2 shows the experimental results of proposed approach applied for Cameraman image.

From table1 and table 2, it can be observed that the bits per pixel (bpp) are more for the proposed modified SOM when compared to the standard SOM. That is bpp for Lena image using modified SOM is 1.84, 2.54, 4.62, 5.12 and 6.76 for cluster size 4, 8, 16, 32 and 64 respectively, whereas in standard SOM less bpp is obtained. With this analysis it can be said that the bits obtained after reconstruction of compressed image will be similar to the original image which undergoes compression.

Then if PSNR is considered, the proposed modified SOM produces PSNR value as 22.15,23.14, 24.23, 26.78 and 28.32 for cluster size 4, 8, 16, 32 and 64 respectively, whereas in standard SOM the PSNR value is minimum when compared with proposed one. This clearly indicates that the noise produced in reconstructed image after compression will be minimum than the noise obtained in the previous methods

Table.1 Experimental Results for Lena Image

Number of Clusters	bpp		PSNR (dB)		MSE	
	Modified SOM	Standard SOM	Modified SOM	Standard SOM	Modified SOM	Standard SOM
4	1.84	1.70	22.15	19.24	503.89	342.45
8	2.54	2.17	23.14	21.33	368.65	187.24
16	4.62	3.82	24.23	22.42	210.42	97.24
32	5.12	5.66	26.78	24.22	111.81	62.21
64	6.76	4.26	28.32	27.10	46.42	36.97

Table.2 Experimental Results for Cameraman Image

Number of Clusters	bpp		PSNR (dB)		MSE	
	Modified SOM	Standard SOM	Modified SOM	Standard SOM	Modified SOM	Standard SOM
4	2.32	1.79	23.65	20.67	453.34	402.52
8	3.34	3.25	24.77	22.65	199.68	190.54
16	4.54	4.32	25.72	25.76	117.84	109.63
32	5.11	5.03	27.91	26.04	68.94	84.32
64	7.47	5.78	30.88	27.57	48.52	73.98

Then the Mean Square Error is considered for analysis. From table 1 and 2 , it can be observed that Mean Square Error (MSE) for Lena image using modified SOM is higher than the standard SOM (ie,. 503.89, 368.65, 210.42,111.81 and 46.42 for cluster size 4, 8, 16, 32 and 64 respectively for modified SOM, whereas 342.45, 187.24, 97.24, 62.21 and 36.97 in case of standard SOM). As the cumulative squared error between the compressed and the original image is minimum in the proposed approach, it reduces the possibility of increasing the noise ratio for the decompressed image

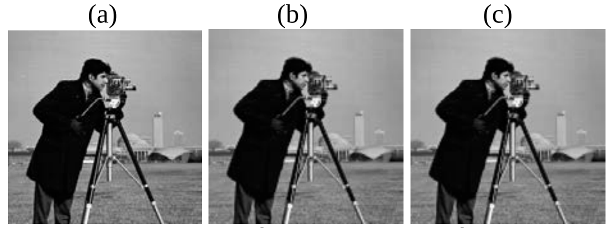
Figure 1 Lena Image

(a) (b) (c)

(a) Sample (b) Resultant image by existing approach (c) Resultant image by proposed approach

Figure 2 Cameraman Image

(a) (b) (c)



Cameraman Image Resultant image by existing approach Resultant image By proposed approach

The Lena and Cameraman Image given for compression and the decompressed image by the existing and the modified approaches are presented above. Figure 1(a) and 2(a) shows the sample input image of Lena and Cameraman respectively and the retrieved image when the existing and proposed approaches used for compression are given in Fig (b) and (c) respectively. The experimental results of the proposed approach of image compression using modified self-organizing feature map algorithm based vector quantization codebook generation revealed the fact that the compression ratio of the proposed approach is high when comparing with other conventional image compression techniques. The decompressed image obtained when the proposed approach used resembles the original image. Thus the proposed approach performs better than the other image compression techniques.

5. Conclusion

Self-organizing map is a popular learning based method and has been widely applied for image compression. This proposed paper introduced a modified self-organizing map algorithm for image compression. This modification overcomes the limitation of the standard SOM algorithm. The Vector Quantization (VQ) codebook is generated by a modified SOM algorithm. This proposed paper modifies the standard SOM algorithm that integrates both the local and the non-local information into the standard SOM algorithm using a novel dissimilarity index in place of the usual distance metric. To evaluate the performance of SOM based vector quantization for image compression some standard image set are considered. The experimental results revealed the fact that the compression ratio of the proposed approach is high when comparing with other conventional image compression techniques. The major limitation of the proposed approach is that it is computationally expensive, and this may limit its applicability to large 3D volume data. Implementation of some suppression technique during the process of iteration helps to overcome this limitation. The future work relies on implementing a suppression technique that can reduce the number of iterations and increase convergence speed of our proposed algorithm effectively.

References

[1] M. J. Nadenau, J. Reichel, and M. Kunt, "Wavelet Based Color Image Compression: Exploiting the Contrast Sensitivity Function," IEEE Transactions Image Processing, Vol. 12, no.1, Pp. 58-70, 2003.

[2] K. H. Talukder, and K. Harada, "Haar Wavelet Based Approach for Image Compression and Quality Assessment of Compressed Image," IAENG International Journal of Applied Mathematics, 2007.

[3] M. Egmont-Petersen, D. de. Ridder, and Handels, "Image Processing with Neural Networks – a review,"Elsevier, Journal on Pattern Recognition, Vol. 35, no. 10, Pp. 2279-2301, 2002.

[4] S. Osowski, R. Waszczuk, and P. Bojarczak, "Image compression using feed forward neural networks — Hierarchical approach," Lecture Notes in Computer Science, Book Chapter, Springer-Verlag, Vol. 3497, Pp. 1009-1015, 2006.

[5] M. Liying, and K. Khashayar, "Adaptive Constructive Neural Networks Using Hermite Polynomials for Image Compression," Lecture Notes in Computer Science, Springer-Verlag, Vol. 3497, Pp. 713-722, 2005.

[6] B. Karlik, "Medical Image Compression by Using Vector Quantization Neural Network," ACAD Sciences press in Computer Science, Vol. 16, no. 4, Pp. 341-348, 2006.

[7] Christopher J. C. Burges, Henrique S. Malvar, and Patrice Y. Simard, "Improving Wavelet Image Compression with Neural Networks," MSR-TR-2001-47, Pp. 1-18, August 2001.

[8] T. Kohonen, "Self Organizing Maps," 3rd ed. Springer, 2001.

[9] D.-S. Q. Hong Wang, Ling LU and X. Luo, "Image compression based on wavelet transform and vector quantization," IEEE Proceedings of the First International Conference on Machine Learning and Cybernetics, Beijing, China.

[10] Jung-Hua Wang, and Ker-Jiang Gou, "Image compression using wavelet transform and self-development neural network," IEEE International Conference on Systems, Man, and Cybernetics, Vol. 4, Pp. 4104-4108, 1998.

[11] Wilford Gillespie, "Still Image Compression using Neural Networks," 2005.

[12] V. Singh, N. Rajpal, and K. S. Murthy, "Neuro-Wavelet Based Approach for Image Compression," Computer Graphics, Imaging and Visualization, CGIV apos'07, Pp. 280-286, 2007.

[13] Barbalho, J.M.; Duarte, A.; Neto, D.; Costa, J.A.F.; Netto, M.L.A., "Hierarchical SOM applied to image compression", Neural Networks, 2001. Proceedings. IJCNN '01, Vol. 1, Publication Year: 2001 , Pp: 442 – 447

[14] C. Ben Amar, and O. Jemai, "Wavelet Networks Approach for Image Compression," ICGST, 2003.

[15] Kwang-Baek Kim; Sung-Kwan Je; Gwang-Ha Kim, "Vector quantizer of medical image using wavelet transform and enhanced neural network", Proceedings of 7th International Workshop on HEALTHCOM 2005, Publication Year: 2005 , Pp: 202 - 207

[16] Adnan Khashman, and KamilDimililer, "Image Compression using Neural Networks and Haar Wavelet," WSEAS Transactions on Image Processing, Vol. 4, no. 5, Pp. 330-339, 2008.

[17] Debnath, J.K.; Rahim, N.M.S. and Wai-keung Fung, "A modified Vector Quantization based image compression technique using wavelet transform", Neural Networks, 2008. IJCNN 2008. (IEEE World Congress on Computational Intelligence), Publication Year: 2008 , Pp: 171 – 176

Mr. G.Boopathi received the Master of Computer Application in 1998. He is currently working towards the Ph.D. Degree in the Department of Computer Application, Karunya University, and Coimbatore. His research interest includes Image Compression using Neural Networks. He is life member of CSI.

Dr.S.Arockiasamy, Head, Information System, University of Nizwa. He received M.Sc., M.Phil. and Ph.D. in Computer Science. He is specialized in Applications of Image processing. He has published a considerable number of Research papers and articles in various leading International journals and International Conferences. He is also Chief Editor of a Quarterly Computer Magazine CSI TIMES (Computer Society of India) and Voice of IT.

Impact of Outlier Removal and Normalization Approach in Modified k-Means Clustering Algorithm

Vaishali R. Patel[1] and Rupa G. Mehta[2]

[1] Department of Computer Engineering, Shri S'ad Vidhya Mandal Institute of Technology
Bharuch-392 001, Gujarat, India

[2] Department of Computer Engineering, Sardar Vallabhbhai National Institute of Technology
Surat-395 007, Gujarat, India

Abstract

Clustering technique is mainly focus on pattern recognition for further organizational design analysis which finds groups of data objects such that objects in a group are similar to one another and dissimilar from the objects in the other group. It is important to preprocess data due to noisy data, errors, inconsistencies, outliers and lack of variable values. Different data preprocessing techniques like cleaning method, outlier detection, data integration and transformation can be carried out before clustering process to achieve successful analysis. Normalization is an important preprocessing step in Data Mining to standardize the values of all variables from dynamic range into specific range. Outliers can significantly affect data mining performance, so outlier detection and removal is an important task in wide variety of data mining applications. k-Means is one of the most well known clustering algorithms yet it suffers major shortcomings like initialize number of clusters and seed values preliminary and converges to local minima. This paper analyzed the performance of modified k-Means clustering algorithm with data preprocessing technique includes cleaning method, normalization approach and outlier detection with automatic initialization of seed values on datasets from UCI dataset repository.

Keywords: *Clustering, k-Means, Normalization Approach, Outlier Removal, Preprocessing.*

1. Introduction

Data mining techniques automate the process to extract hidden patterns from the heterogeneous data sources and to analysis the results which is helpful to the organization for decision making with the development of number of technologies. Data mining is one of the most important research fields that are due to the expansion of both computer hardware and software technologies, which has imposed organizations to depend heavily on these technologies [1]. According to [2], the obtained clusters should react some mechanism at work in the domain from which instances or data points are drawn, a mechanism that causes some instances to bear a stronger resemblance to one another than they do to the remaining instances." Clustering is one solution to the case of unsupervised learning, where class labeling information of the data is not available. It is a method where data is divided into groups (clusters) which 'seem' to make sense. Clustering algorithms are usually fast and quite simple. They need no beforehand knowledge of the used data and form a solution by comparing the given samples to each other and to the *clustering criterion* [3]. Various clustering algorithms according to different techniques have been designed and applied to various data mining problems successfully. Accomplishment of clustering algorithms in many areas, it causes many precincts to the researchers when no or little information are available. There is also no universal clustering algorithm developed; that's why it is very crucial job to choose appropriate clustering technique and algorithm considering above precincts. A good survey on clustering techniques and algorithms found in [4]. A simple and commonly used algorithm for producing clusters by optimizing a criterion function, defined either globally (over all patterns) or locally (on a subset of the patterns), is the k-means algorithm [5]. The k-Means algorithm [5] is effective in producing clusters for many practical applications. This algorithm results in different types of clusters depending on the random choice of initial centroids. Several attempts were made by researchers to improve the performance of the k-Means clustering algorithm. k-Means use Euclidean distance measure centroids of the clusters and distortion among the data objects. These distances are not computed from standardized data. In method of Euclidean distance, the measured distance between data objects is not affected by addition of new objects to the analysis [6]. Missing value is a common problem in almost every real world data. The presence of missing values in data results in datasets that [7] refers to as "incomplete" datasets since some information will not be available. Data pre-processing techniques are applied on raw data to make the data clean, noise free and consistent. Data Normalization standardize the raw data by converting them into specific range using linear transformation which can generate good quality clusters and improve the accuracy of clustering algorithms. The outlier detection is searching for objects in the database that do not obey laws valid for the major part of the data [8]. In clustering, outliers are considered as observations that should be removed in order to make clustering more reliable. Different approaches have been proposed to detect outliers and some of these are discussed in literature survey explained in section 2. Section 3 explains the traditional k-Means with pros and cons. Section 4 proposes modify k-Means clustering algorithm which detect outlier using 5-95% method, apply different normalization techniques like Min-Max, Z-Score and Decimal Scaling to improve the performance and accuracy of the k-Means algorithm. The proposed method first checks to ensure that the data apply to the algorithm are clean and standardized then apply 5-95% method which discard the data and consider it as outlier of the given dataset. Section 5 discusses the implementation of modify k-means and result analysis is done on dataset from the UCI dataset repository which shows that outlier detection and removal with normalization approach improve the effectiveness and performance of the modified k-Means clustering algorithm.

2. Literature Survey

Clustering algorithms generate clusters having similarity between data objects based on characteristics belongs to same cluster. Clustering is extensively used in many areas such as pattern recognition, computer science, medical, machine learning. Result of clustering is dependent on the type of data and application area. An outlier is an observation that deviates so much from other observations as to arouse suspicion that it was generated by a different mechanism [9]. In outlier detection methods based on clustering, outlier is defined to be an observation that does not fit to the overall clustering pattern [10]. This section discusses the various approaches proposed by many researchers to detect outliers in k-Means and other clustering algorithms to handle noise and generate successful results.

Authors [11], proposed a new clustering based approach, which divides the stream in chunks and clusters each chunk using k-median into variable number of clusters. Instead of storing complete data stream chunk in memory, they replace it with the weighted medians found after mining a data stream chunk and pass that information along with the newly arrived data chunk to the next phase. The weighted medians found in each phase are tested for outlierness and after a given number of phases, it is either declared as a real outlier or an inlier. This technique is theoretically better than the k-means as it does not fix the number of clusters to k rather gives a range to it and provides a more stable and better solution which runs in poly-logarithmic space. This approach works only for numeric dataset.

Ville Hautam aki et al. [12] proposed an Outlier Removal Clustering (ORC) algorithm which detects outlier and data clustering simultaneously. The method employs both clustering and outlier discovery to improve estimation of the centroids of the generative distribution. During the first stage of this algorithm, basic k-Means algorithm executes, while during the second stage it iteratively removes the vectors which are far from their cluster centroids. This approach outperforms, particularly in the case of heavily overlapping clusters. In this approach, setting of correct parameter depends on the type of dataset.

Authors [13] have proposed shortest distance method for detecting outliers in k-Means and k-Medians clustering algorithm. In this algorithm, outliers are detected by computing its distance which is far away from the rest of the data objects in the dataset.

Moh'd Belal and Al- Zoubi [14] have proposed an algorithm based on clustering approaches to detect outliers. This algorithm first performs the PAM clustering algorithm. Small clusters are then determined and considered as outlier clusters. The rest of outliers are then detected in the remaining clusters based on calculating the absolute distances between the medoid of the current cluster and each one of the points in the same cluster. This algorithm can be easily implemented on other clustering algorithms that are based on PAM.

During the first phase of proposed two phase clustering algorithm [15], traditional k-Means algorithm is modified using a heuristic "if one new pattern is far enough away from all clusters' centers, then assign it as a new cluster center". As a result, data objects in the same cluster may be most likely all outliers or all non-outliers. During second phase of this proposed algorithm, Minimum Spanning Tree (MST) is constructed and removes the longest edge and small clusters are considered as outliers.

3. Naive k-Means Algorithm

This section discusses the working of traditional k-Means clustering algorithm. k-Means algorithm is one of the most popular clustering algorithms due to its efficiency and simplicity in clustering large data sets. In traditional k-

Means algorithm, a set of data set D is classified using a certain number of clusters (k clusters) which are initialized apriori. It define k centroids, one for each cluster and then consider data object belonging to the given data set and associate this data objects to the closest centroid. Euclidean distance generally considered to determine the distance between data objects and the centroids [16]. First step is completed when there is no data object is remaining and early group is done. Here, there is need to re-calculate new centroids. After obtaining new centroids same data objects are binded with the closest centroid and generate a loop. At the end of loop, k-centroids change their point step by step until centroids do not move any more. This algorithm works on basis of minimizing squared error function. k-Means algorithm suffers from the problems of giving number of clusters and initial seed values preliminary. The k-Means algorithm always converges to a local minimum and it depends on the initial cluster seed values. We have make modifications in traditional k-Means algorithm to initialize the seed values with data preprocessing and data normalization techniques like Min-Max, Z-score and decimal scaling to improve the accuracy and efficiency of traditional algorithm. Section 4 discusses the modifications in k-Means clustering algorithm with normalization approach.

4. Modified k-Means Algorithm with Outlier Detection and Removal

k-Means algorithm can generate better result after modification on the datasets. We apply the modified k-Means algorithm with automatic initialization of number of seed values on river dataset for number of iterations and clusters and compute MSE. Next, we preprocess and normalize river dataset before apply on modify k-Means algorithm. This proposes method works in two stages. During the first stage, we preprocess the dataset then compute and discard 5-95% data from the dataset. Store and normalize these discarded data separately which we consider outliers and use it as new cluster with modified k-Means algorithm. During the second stage, apply modify k-Means algorithm to remaining data to generate clusters. Block diagram and flow chart of the proposed modified k-Means algorithm is shown in Figure 1 and Figure 2 respectively:

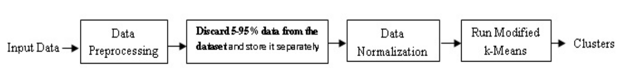


Fig. 1. Block Diagram of modified k-Means clustering algorithm with Outlier Detection

Platform used: VB 6.0 and MS SQL Server 5.0

Input: RIVER Datasets from UCI Machine Leaning

Transform Module: This module accept the dataset in text format and convert it in database file.

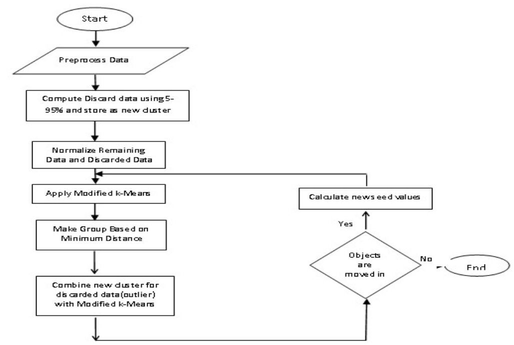


Fig. 2: Flow Chart of Modified k-Means with Outlier Detection and Removal

Data Preparation Module: This module works in two parts. First, data preprocessing technique apply on the dataset received by transform module. The clean data is then passed to second part; Data Normalization which transform the clean raw data into specific range using different techniques.

Data Pre-processing: This is a very important step since it can affect the result of a clustering algorithm. This module calculates tuples with missing values using different options like maximum, minimum, constant, average and standard deviation for the treatment of missing values tuples before we apply normalization approach on the dataset. This process gives the treatment of missing value data and then it applies to the second part (data normalization) of data preparation.

Normalization Approach: Data Mining can generate effective result if proper and effective data mining technique can apply to the dataset. According to authors [17], normalization is used to standardize all the features of the dataset into a specified predefined criterion so that redundant or noisy objects can be eliminated and use made of valid and reliable data which can effect and improve accuracy of the result. k-Means clustering algorithm uses Euclidean distance measures which are highly susceptible to inconsistencies in the size of the features. k-Means algorithm which is using Euclidean distance measure, data normalization is an essential step to prevent larger features from randomized value to the specific range. The importance of normalization is that it enhances the accuracy of the results that are obtained during clustering [18]. Better results are generated when data preparation with data preprocessing and normalization is carried out with different techniques. Data normalizations techniques include min-max normalization, Z-Score normalization, and Decimal Scaling normalization. There is no universally defined rule for normalizing datasets and thus the choice of a particular normalization rule is largely left to the discretion of the user [17]. The proposed Mk-Means algorithm has utilized the three normalization techniques and compares analysis of achieved results. The Min-Max normalization technique involves the linear transformation on raw data. MinA and MaxA are minimum and maximum value for the attribute A. This technique maps the value of attribute A into range of [0, 1]. Equation (1) shows the computation of Min-Max normalization technique.

$$v' = \frac{v - MinA}{MaxA - MinA} \qquad (1)$$

Z-Score normalization technique is useful when the actual minimum and maximum value of attribute A are unknown the value of an attribute using standard deviation and mean of the attribute A. Equation (2) shows the computation of Z-Score normalization technique. In equation (2), $\bar{A}$, σA and v are mean, standard deviation and value of attribute A.

$$v' = \frac{v - \bar{A}}{\sigma A} \qquad (2)$$

Decimal scaling normalization technique normalize the data by moving the decimal point of values of attribute A. Number of decimal points moved depends on the maximum absolute value of A. Equation (3) shows the computation of decimal scaling normalization technique. In equation (3), v is the value of attribute A and j is the smallest integer where Max (|v'|) <1.

Outlier Detection: Outliers detection is a task that finds objects that are dissimilar or inconsistent with respect to the remaining data. It has many uses in applications like fraud detection, network intrusion detection and clinical diagnosis of diseases. Clustering algorithms are frequently used for outlier detection. The clustering algorithms consider outlier detection only to the point they do not interfere with the clustering process. In this proposed approach, outliers are detected using 5-95% method in which 5% of data from minimum side and 5% data from maximum side are detected and removed from the dataset. The experimental results prove that Modified k-Means clustering algorithm with outlier detection and removal improves the accuracy and increases the time efficiency of Mk-Means algorithm.

$$v' = \frac{v}{10^j} \qquad (3)$$

Run modified k-Means Algorithm: The modified k-Means algorithm is implemented and applies it on River dataset which contains 133 tuples and 9 attributes. We have applied this algorithm using two dimensional on RIVER dataset using 4 iterations and different number of clusters. We have applied the same algorithm with data preparation (cleaning method) and data normalization techniques and compute the MSE for the same.

5. Result Analysis

To analyze the accuracy of proposed Mk-Means clustering algorithm, result is taken on River dataset from UCI machine learning data repository. Comparison of computed MSE of Mk-Means with computed MSE of Mk-Means algorithm with cleaning method and various normalization techniques like Min-Max, Z-Score and Decimal Scaling is done on River dataset. This Analysis shows the best result for Mk-Means with normalization approach. Computed MSE and graph of proposed Mk-Means algorithm with outlier removal and cleaning method is shown in Table 1 and Figure 3 respectively. Computed MSE and graph of Mk-Means with different normalization techniques like Min-max, Z-score and Decimal scaling is shown in Table 2 and Figure 4 respectively. Computed MSE and graph of Mk-Means with different normalization techniques like Min-max, Z-score and Decimal scaling and outlier removal is shown in Table 3 and Figure 5. If there are N tuples in the dataset, then, the similarity matrix can be computed in O(KNT). Let N is the number of tuples in the dataset. K is the number of clusters or centroids and T is the time to calculate the distance between to data objects. Time complexity of each iteration is O(KNT). I is the number of iterations in k-Means algorithm. So, during I number of iteration the time complexity of this algorithm is O(IKNT). The performance analysis of modified k-Means clustering algorithm shows that decimal scaling normalization technique gives the best results for the modified k-means clustering algorithm and secondly min-max data normalization generates the best results for modified k-means clustering algorithm. The analysis shows that outlier detection and removal with generates the best and most effective and accurate results than other techniques used in this paper. Comparison of MSE of proposed Mk-Means algorithm with MSE of Data Normalization techniques with Mk-Means algorithm. MSE of proposed modified k-means clustering algorithm with min-max, z-score and decimal scaling. Analysis shows that by applying data preparation techniques like cleaning method for missing value treatment, different normalization approaches and outlier detection and removal improve the performance of modified k-Means clustering algorithm.

Table 1: MSE of proposed Mk-Means Clustering Algorithm with Outlier Removal

DataSet	No. of Samples	Data Preprocessing	Number of clusers	MSE of proposed Mk-Means Algorithm	MSE of proposed Mk-Means Algorithm with Outlier Removal
RIVER	133	Average	2	4.967	4.332
			3	114.597	103.703
			4	55.284	32.095
			5	30.124	18.414
			6	32.531	28.879

Table 2: MSE of proposed modified k-means clustering algorithm with min-max, z-score and decimal scaling normalization techniques

DataSet	No. of Samples	Data Preprocessing	Number of clusers	MSE of proposed Mk-Means with Normalization techniques		
				Min Max	Z-Score	Decimal Scaling
RIVER	133	Average	2	0.769	0	0.88
			3	25.701	83.701	4.377
			4	3.567	64.744	0.761
			5	4.619	22.833	1.19
			6	4.0	18.384	1.308

Table 3: MSE of Proposed Mk-Means algorithm with Outlier Removal and Normalization Techniques

DataSet	No. of Samples	Data Preprocessing	Number of clusers	MSE of Mk-Means with Outlier Removal and Normalization Techniques		
				Min Max	Z-Score	Decimal Scaling
RIVER	133	Average	2	0.76	0	0.839
			3	24.566	62.225	4.197
			4	3.203	43.957	0.655
			5	3.554	15.475	0.979
			6	5.111	10.616	1.096

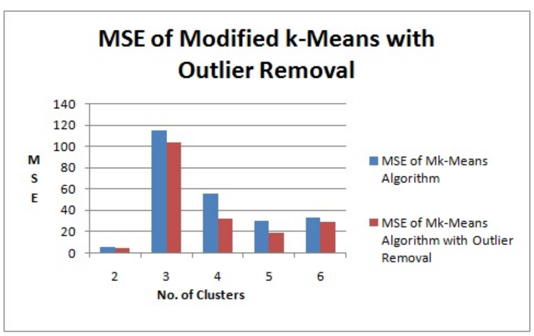


Fig 3: MSE of Mk-Means with Outlier Removal

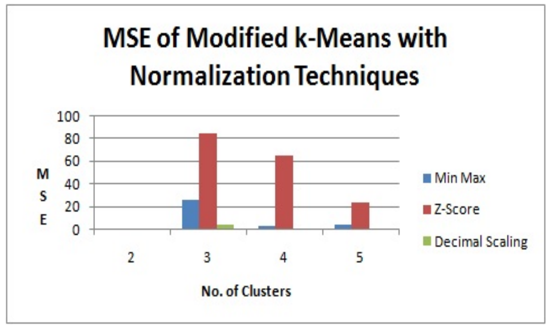


Fig 4: MSE of Mk-Means with Normalization Techniques

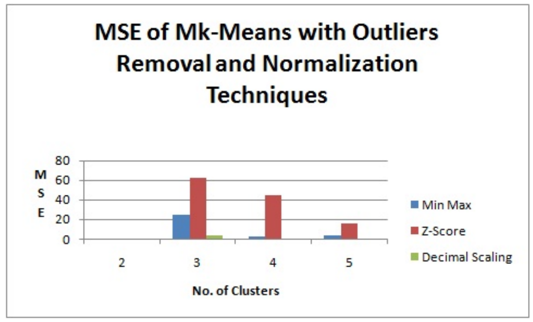


Fig 5: MSE of Mk-Means with Outlier Removal and Normalization Techniques

6. Conclusion

Data Mining is the step in KDD to extract useful pattern. Clustering organize the data in group having similarity. k-Means clustering is the well known partition based clustering algorithm. K-Means suffers from one of the problem of initializing seed values. This paper give explore of data mining and literature survey of methods proposed by many researchers to remove initialization seed values in k-Means. Real world data may be noisy, with missing values or inconsistent. There are a number of data preprocessing techniques to clean the data. Data cleaning can be applied to remove noise and correct inconsistencies in the data. Data transformations, such as normalization, may be applied to improve the accuracy and efficiency of mining algorithms. These data processing techniques, when applied prior to mining, can substantially improve the overall quality of the patterns mined and/or the time required for the actual mining. This paper also propose modifications in naive k-Means to auto initialize seed values with data preparation which preprocess the data with cleaning method and transform the data into specific range using min-max, z-score and decimal scaling data normalization techniques. Outlier is a noise in the clustering algorithm which is detected using 5-95% method and removes from the dataset. Performance analysis of computed MSE for Mk-Means and Mk-Means with three normalization techniques with outlier removal shows best and effective result for Mk-Means which generate minimum MSE and improve the efficiency and quality of result generated by this algorithm.

Acknowledgments

The authors wish to thank Computer Engineering department of Sardar Vallabhbhai National Institute of Technology for the support and providing an environment for this research work.

References

[1] R. Agrawal, T. Imielinksi and A. Swami, “Mining association rules between sets of items in large database, “The ACM SIGMOD Conference, Washington DC, USA, 1993, pp. 207-216.

[2] Babu S. and Widom J. (2001), “Continuous Queries over Data Streams”, Stanford University, SIGMOD Record, 30, pp. 109–120.

[3] Jukka Kainulainen, ”Clustering Algorithms: Basics and Visualization”, HELSINKI UNIVERSITY OF TECHNOLOGY, Laboratory of Computer and Information Science,T-61.195 Special Assignment 1, 2002.

[4] Vaishali R. Patel, Rupa G. Mehta, “Clustering Algorithms: A Comprehensive Survey”, International Conference on Electronics, Information and Communication Systems Engineering, MBM Engineering College, JNV University, Jodhpur, 2011.

[5] McQueen J, “Some methods for classification and analysis of ultivariate observations,” Proc. 5th Berkeley Symp. Math. Statist. Prob., Vol. 1, 1967, pp. 281–297.

[6] Grzymala-Busse and J. W., “Three Approaches to Missing Attribute Values – A Rough Set Perspective”, http://lightning.eecs.ku.edu/c97-brighton.pdf.

[7] N. Karthikeyani Visalakshi and K. Thangavel, ”Impact of Normalization in Distributed K-Means Clustering”, International Journal of Soft Computing 4, Vol. 4, 2009, pp. 168-172.

[8] Su, C.M., Tseng, S.S., Jiang, M.F., Chen and J.C.S., ”A Fast Clustering Process for Outliers and Remainder Clusters”, Lecture Notes in Artificial Intell, 1999, pp. 360-364.

[9] D.Hawkins: “Identification of outliers”. Chapman and Hall, London, 1980.

[10] Zhang T., Ramakrishnan, R. and Livny M., “BIRCH: A new data clustering algorithm and its applications”, Data Mining and Knowledge Discovery, 1997, pp. 141–182.

[11] Parneeta Dhaliwal, MPS Bhatia and Priti Bansal, ”A Cluster-based Approach for Outlier Detection in Dynamic Data Streams (KORM: k-median OutlieR Miner)”, Journal of Computing, Vol. 2, No. 2, 2010, ISSN 2151-9617, pp. 74-80.

[12] Ville Hautam aki, Svetlana Cherednichenko, Ismo Karkk ainen,Tomi Kinnunen, and Pasi Fr anti, ”Improving k-Means by Outlier Removal”, SCIA, LNCS 3540, 2005, pp. 978–987.

[13] Sairam, Manikandan and Sowndarya, ”Performance Analysis of Clustering Algorithms in Detecting Outliers”, International Journal of Computer Science and Information Technologies, Vol. 2, No. 1, 2011, ISSN: 0975-9646, pp. 486-488.

[14] Moh’d Belal Al- Zoubi, ”An Effective Clustering-Based Approach for Outlier Detection”, European Journal of Scientific Research, Vol. 28, No. 2, 2009, ISSN 1450-216X, pp.310-316.

[15] M.F. Jiang and S.S. Tseng, ”Two Phase Clustering Process for Outlier Detection”, Pattern Recognition Letters 22, Elsevier Science B.V., 2001, pp. 691-700.

[16] K. A. Abdul Nazeer and M. P. Sebastian, ”Improving the Accuracy and Efficiency of the k-means Clustering Algorithm”, Proceedings of the World Congress on Engineering¸ Vol. 1, 2009.

[17] N. Karthikeyani Visalakshi and K. Thangavel, ”Distributed Data Clustering: A Comparative Analysis. Foundations of Computational Intelligence (6)”, 2009, pp. 371-397.

[18] M arcilio CP de Souto, Ivan G Costa, Daniel SA de Araujo, Teresa B Ludermir and Alexander Schliep, ”Clustering cancer gene expression data: a comparative study”, BMC Bioinformatics”, 2008.

Vaishali R. Patel is M.Tech (Research) scholar and currently working as an Assistant Professor in Department of Computer Engineering and Information Technology at Shri S’ad Vidya Mandal Institute of Technology, Bharuch, Gujarat, India. Her research areas of interest include Data mining, Clustering, Database Management System` and Software Testing and Quality. She is Microsoft Certified Solution Developer.

Rupa G. Mehta is Ph. D. Scholar and currently working as an associate professor in Department of Computer Engineering at Sardar Vallabhbhai National Institute of Technology, Surat, Gujarat, India. Her research areas of interest include data mining, classification techniques, database management systems, data structures and formal language.

Energy Efficient Routing Scheme for Sensor Networks Using Connected Dominating Set

K. P. Sampoornam[1], Dr. K. Rameshwaran[2]

[1]Department of ECE, K.S.R. College of Engg.
Tiruchengode, Tamil Nadu, India

[2]Principal, JJ College of Engg,
Tiruchirappalli, Tamil Nadu, India

Abstract

A power efficient scheduling scheme extending the life time of sensor nodes on wireless sensor networks has been proposed in this paper. This power efficient sleep scheduling algorithm is based on the connected dominating set approach. This approach first constructs a connected dominating set for parent selection and uses the conjugative sleep scheduler scheme for data aggregation. The re-calculation of connected dominating set when the node is in ON condition and in OFF conditions have been discussed. Finally, the performance of connected dominating set approach is compared with minimum spanning tree approach. By simulating the network with different node density, it is observed that, our proposed approach performs better than spanning tree topology. Whenever a sensor node fails or a new node is added to the network, the conjugative scheduling scheme evaluates the route maintenance. Remaining power levels of nodes have been examined in order to increase the network lifetime.

Keywords: *Sensor Networks, Data Aggregation, Power, Sleeping Schedule, CDS.*

1. Introduction

Wireless Sensor Networks are powerful in that they are amenable to support a lot of very different real world applications They are also a challenging research problem in Engineering, because of their flexibility. In many WSN applications, individual nodes cannot easily be connected to a wired power supply. Hence, energy efficiency along with life time of a proposed solution has been a very important figure of merit as a long operation time is usually desirable. Realizing such a wireless sensor networks is a crucial step.

Based on the application requirements, designing of wireless sensor networks is not unique. Various sensor nodes that are sensed by the environmental monitoring are selected and placed at a base station where numerous processing can be carried out. Design of the sensor nodes varies accordingly so that in certain cases the possibility of recharge or replacement of batteries cannot take place. A common example of sensor networks allows the involvement of hundreds or thousands of sensor nodes from where any amount of information can be collected frequently. For the sake of end-user queries, the information collected can be processed further. The data received can be combined to reduce the message size and further to reduce the communication overhead and energy consumption.

In order to increase the life time of the network by finding an energy efficient routing scheme for accumulating all the data at the sink periodically, which is the main problem discussed in this paper. A simple way of doing that is aggregating the data. An important problem studied here is finding an energy efficient routing scheme for gathering all data at the sink periodically so that the lifetime of the network is prolonged as much as possible [1][2]. The life time of the network can be measured in terms of rounds where a round denotes the difference of the time taken between two sensing activities of the sensor nodes. Studies also show us that a protocol named Minimum Spanning Tree (MST) based routing provides good performance in terms of lifetime when the data is gathered with the help of aggregation. Here the authors proposed a new centralized protocol called PEDAP [3].

The proposed routing approach includes two phases and satisfies the requirements stated above. The first phase discusses the sparse topology over the visibility graph of the network in a localized manner. The second phase, computes the data gathering tree over the edges of the computed sparse topology. All the methods discussed above are based on flooding which is a special packet which uses only the edges of the computed structure. Based on the results obtained the resultant decisions are made.

2. Literature Review

Several routing protocols have been proposed for data gathering without aggregation.

2.1 Routing Protocols

Shortest weighted path approach is one of the main methods using several combinations of transmission power, reluctance, hop count, and energy consumption metrics. Classical routing algorithms such as AODV [4] or Directed Diffusion [5] can be considered also for this case. As the transmission distance increases, this method is also not very efficient, since the transmission distances can be quite long and finding a minimum distance chain is very difficult. Results also shows that the protocol PEDAP which routes the packets on the edges of a minimum spanning tree improves the system lifetime when compared to its alternatives discussed by Tan and Korpeoglu [6].

With the help of Prim's MST algorithm [7], PEDAP protocol using the link costs computes the minimum energy cost tree. The MST algorithm helps to design a centralized algorithm that focuses at finding a spanning tree whose maximal degree is the minimum among all possible spanning trees, as the energy consumed at each node depends only on the number of messages received from children nodes, that is, on the number of children. As a result, these trees reduce the load on bottleneck nodes [8].

2.2 Power Efficient Topologies

Discussion on energy efficiency of topologies using location information of one-hop neighbors for various topologies is available in the literature. A localized topology that helps to identify the shortest weighted path, called as enclosure graph, has been discussed by Rodoplu and Meng [9]. An edge $\mathbf{e_{ij}}$ is called as the enclosure graph if the direct transmission between node i and node j consumes lesser energy than the total energy of all links of any path between them. When shortest weighted path tree uses routing without aggregation, the enclosure graph exhibits a good performance. The topologies that we have concentrated in this work are supersets of Euclidean MST. One among them is the relative neighborhood graph (RNG) [10].The main advantage of using structures like RNG and LMST [11] is that they can be constructed very efficiently in a localized manner. The deletions of nodes and additions of nodes do not change the structure globally. The changes to be performed in the structure are only local changes which can be efficiently computed [12] when a node fails or when a new node is introduced to the network.

3. Efficient data aggregation for sensor networks with Connected Dominating Set.

This section mainly focuses on formation of a dominating set [13]. The main objective of this paper is to combine the features of Connected Dominating Set (CDS) with conjugative sleep schedule algorithm. Some desirable features are: (1) The formation process should be distributed and simple. (2) The resultant dominating set should be connected and close to minimum. (3) The resultant dominating set should include all intermediate nodes of any shortest path. In order to achieve this, the following algorithm is proposed.

Algorithm

- Create a sorted sequence of edges based on their transmission – reception energy loss.
- Randomly shuffle the sequence of sorted edges.
- Construct a spanning tree based on this shuffled sequence of edges.
- Remove the edges that are in the previously generated spanning tree(s) and obtain a reduced graph.
- Repeat the above procedure on the reduced graph and generate spanning trees until no more spanning trees can be generated.
- The internal nodes of a spanning trees form a Connected Dominating Set (CDS).Select the root node to be the node with the highest energy among the nodes in the CDS and form a rooted directed tree among the CDS nodes.
- The rooted directed tree along with the links from the leaf nodes to their nearest CDS node forms the directed graph tree.
- Repeat the data gathering process for every T rounds and then switch to the next edge disjoint spanning tree in the k- tree list.
- Compare this approach with just a MST based DG tree for every round. Choose the node with the highest energy as the root and form a rooted directed tree. The rooted directed tree along with the links from the leaf nodes to their nearest CDS node forms the MST-based DG tree .

In order to construct CDS, the following two rules are followed. Consider G and G' are the connected and disconnected dominating set respectively. In G', *v* and *u* are two disconnected vertices .In order to reduce the size of a connected dominating set, Jie Wu et.al. Proposed two rules. For this, we first assign a distinct id, *id(v)*, to each vertex *v* in G'. **N [*v*] = N(*v*) U {*v*}** is a closed neighbor set of v, as oppose to the open one N (*v*)

RULE 1: Consider two vertices v and u in G'. If N[v] $\subseteq$ N[u] in G and id(v) < id(u), change the marker of v to F if node v is marked, i.e., G' is changed to G' - {v}.

RULE2: Assume u and w are two marked neighbors of marked vertex v in G'. If $N(v) \subseteq N(u) \cup N(w)$ in G and id(v) = min{id(v), id(u), id(w)}, then change the marker of v to F.

After developing a sparse topology, which is suitable for energy aware rooting, Relative Neighbor Graph (RNG) topology is developed. RNG needs only location information of 1- hop neighbors. It develops Localized Minimum Spanning Tree (LMST), which needs a second message for the sake of informing to local neighbors.

3.1 Recalculation of Connected Dominating Set:

There are two different topological conditions such as sensor node in ON condition and node is in OFF condition are considered for CDS recalculation.

a) Sensor node Vs switch on.

When a node v switches on, only its non-gateway neighbors along with host
V needs to update their status, because any gateway neighbor will still remain as gateway after a new vertex v is added.
The corresponding marking process can be the following:

(i) Node v broadcasts to its neighbors about its switch on.
(ii) Each host $w \in N(v)$ exchanges its open neighbor set N (w) with its neighbors.
(iii) Host *v* assigns its marker m (*v*) to T if there are two unconnected neighbors.
(iv) Each non gateway host w E N (v) assigns its marker m (w) to T if it has two unconnected neighbors.
(v) Whenever there is a newly marked gateway, host *v* and all its gateway neighbors apply Rule 1 and Rule 2 to reduce the number of gateway hosts.

b) Sensor Node Vs switch off.

When a node v switches off, only gateway neighbors of that switched off host need to update their status, because any non-gateway neighbor will still remain as non-gateway after vertex *v* is deleted.
The corresponding marking process can be the following:

(i) Node v broadcasts to its neighbors about its switching off.
(ii) Each gateway neighbor $w \in N(v)$ exchanges its open neighbor set N (*W*) with its neighbors.
(iii) Each gateway neighbor *w* changes its marker m (*w*) to F if all neighbors are pair wise connected.

Since the underlying graph G is connected, we can easily prove by contradiction that the resultant dominating set (using the above marking process) is still connected when a host (gateway or non-gateway) switches off.

3.2 Data Gathering and Route Maintenance

The job of the sensor node is to periodically sense its nearby environment and tries to generate the data that has to be sent to the sink. Transmission occurs only once in the sensor in a round robin fashion and saves its energy. Then we have to wait for all the data coming from its child nodes to aggregate the data coming from them together with its own data and then the aggregated data is send to the parent node [14]. Whenever the power-aware functions are used while considering route maintenance, there is a possibility of changes to takes place in routing plan and route re-computation [15]. Next we broadcast a new ROUTE-DISCOVERY packet with a new sequence ID. In the node addition process new node broadcasts a HELLO message [3]. The neighboring node updates the local structure and message is being replied to its neighboring node which helps to determine its neighbors.

3.3 Conjugative Power Efficient Sleep Scheduling Scheme

This phase discusses about the Conjugative sleep scheduling scheme in order to efficiently compute routing for the localized energy levels in wireless sensor networks. Better connectivity is being maintained which satisfies the user defined coverage target conjugative to the sensor node and the sink. The main job of sensor nodes is to give the feedback from the sink for the purpose of scheduling activity of the neighbors with the help of residual energy. Sleep scheduler mode reduces randomness in scheduling and the sleeping nodes run in a synchronized schedule. It also maintains a backbone composed of active nodes which deliver data to sleeping nodes when they wake up and bound the communication delay between any two nodes in the order of one duty cycle. Sensor node motion predicts future pickup points where the user expects a query result. In case of pre-fetching, the sensor nodes send pre-fetched messages to future pickup points ahead of time. In the sleep schedule mode, query dissemination is done for collector node which creates a routing tree, to alert the sleeping nodes. The data collection is carried out by the waking up the nodes in time and send data to user through the tree

3.3.1 Conjugative Sleep Schedule operation

The conjugate transpose matrix of a-by-b matrix X with the complex entries is b-by-a matrix X^T which is the transpose. It is an iterative method and can be applied to methods which have enormous data to be handled manually. The resultant transpose matrix is applied with the complex conjugate for each entry. Here, we negate the imaginary parts without affecting the real parts. Defining the conjugate transpose as

$$(X^T)_{mn} = \text{Negation}(X_{mn}) \qquad (1)$$

where the subscripts denote the m, n-th entry, for $1 \leq m \leq a$ and $1 \leq n \leq b$, and the negation denotes a scalar complex conjugate value. (The complex conjugate of x + iy, where x and y are real's, is x − iy). We say that two non-zero vectors a and b are conjugate (with respect to X) if

$$a \times b = 0 \qquad (2)$$

Further this conjugate relation is also a symmetric one. (i.e.) If **x** is conjugate to **y,** then **y** is conjugate to **x**.
The operation of the sensor network involves two states: A) A state in which no target is involved. B) A state in which target is involved. Our work involves energy efficient sleep scheduling algorithm for tracking state. We assume that all the nodes follow the random sleep technique. As the sleep pattern is random, all the nodes assume the same toggling period TC at the same time and each node calculates its own random starting time. For instance, a node will wake up and keep active for TC * DC, and then enter the sleep state for TC * (1 - DC).

3.3.2 Algorithm: Conjugative Power Efficient Sleep Scheduling

1. For each node **n** in the tracking neighbors do
2. Compute the distance from the neighbors root node to node **n**
3. Compute t_{start} and t_{end}.
4. Compute the angle θ between the 'root node ' and 'node n' and the instant velocity **v**.
5. Compute DC_{max}.
6. Set the next cycle value as DC_{max} for node **n.**
7. Set n's duty cycle recovery number as Round $(T_{end} - T_{start})$ / TC where Round refers to the rounding function and TC refer to Toggling Cycle.
8. If T_{start} = minimum_sleep_time then
   a. Set node **n** state as SLEEP
   b. Reset node **n** waking up state as T_{start}

3.4 Performance on Sensor Network Density and Power efficiency

The performance measure of distributed adaptive sleep scheduling algorithm in wireless sensor network is performed by applying different density with the help of variation in the sensor nodes. With the help of centralized sleep scheduling algorithm the performance metric is measured. Network lifetime of about 92% is attained along with better energy savings.

4. Experimental Results and Discussion on Power Efficient Scheduler for Sensed Data Aggregation

The experimental evaluation starts from choosing the best parent selection strategy using CDS and then continue the data aggregation with sleep scheduler strategy. Running the experiments with the different topologies for our scenario with three parameters i.e., number of nodes **N**, maximum transmission radius **R** and side-length of the square area **L.** Another parameter defined in this topology is density ***d***, which is defined as the average number of neighbors per node.

Then generate a network with parameters $N = 80, R = 30m$, $L = 120m$. We repeated the experiments on this network with sleep scheduler topologies using spanning tree based parent selection methods. Then compare the efficiency of these two parent selection methods. We conducted experiments with different values of *N*, *R* and *l*.

The life time of the network is obtained for various density set up. The results are tabulated in Table 1.. From the table it is clear that, the Conjugate Sleep Schedule algorithm with CDS extends the life time of the network as the density of the node increases. Both the attributes have been tested with CDS based routing tree construction and without CDS.

Table 1: Performance Comparison of Network Life Time

Conjugate Sleep schedule		***Conjugate Sleep schedule with CDS***	
Density	***lifetime***	***Density***	***lifetime***
10	69	10	78
20	75	20	83
30	79	30	86
40	77	40	79
50	76	50	78

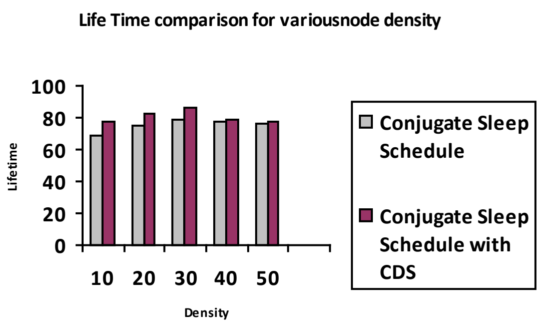


Fig 1. Effect of CDS based Conjugate Sleep Schedule on Network Life Time

Fig1.shows the performance of conjugate sleep scheduler scheme with CDS is better than conjugate sleep scheduler scheme with out CDS for increasing data nodes. As the area enlarges, connectivity decreases, and distances get longer induces the sensed data time reduction.

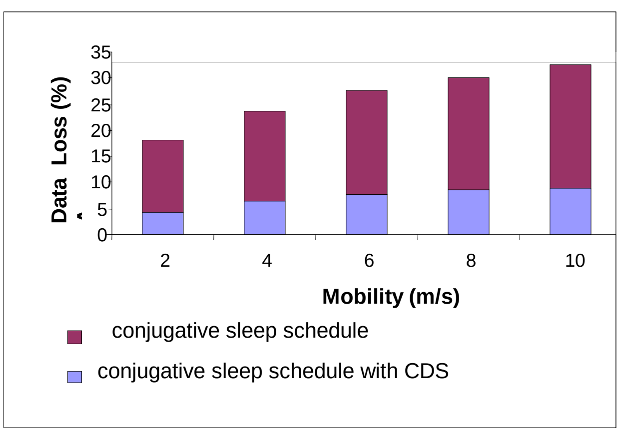


Fig. 2: Mobility Vs data loss

Fig.2 shows the simulation output of data loss for various mobility of nodes in WSN. The mobility factor of the node

Table 2: Comparison of Data loss performance

Mobility (m/s)	***Data loss (%)***	
	Conjugate Sleep schedule with CDS	***Conjugate Sleep schedule***
2	4.2	13.9
4	6.4	17.2
6	7.8	19.8
8	8.5	21.7
10	9	23.6

is measured in terms of meters per second which usually causes the loss of data being transited at higher levels if the messages are sent before data redundancy reduction process. While transmitting the data, several portions of data values gets aggregated and stored in aggregator node. After removing the redundant data, aggregator node sent data. As the mobility increases, data loss also increases. The performance results obtained through simulation on data loss characteristics during transmission scenario shows that Conjugate Sleep Schedule with CDS Scheme proposed in our work decreases the data loss compared to that of existing Conjugative Sleep Scheduling Scheme.

Fig.3 depicts the output of the simulation by varying the number of nodes in WSN. There is an appreciable change in the throughput the of number of nodes. In our proposed scheme, due to the elimination of redundant data values, the load aggregation gets normalized, which in turn improves the throughput of the sensed data transmission to the sink. Fig. 3 shows that as the number of nodes increases, throughput decreases. By comparing it with Conjugative sleep scheduling model, the throughput is high when CDS is used.

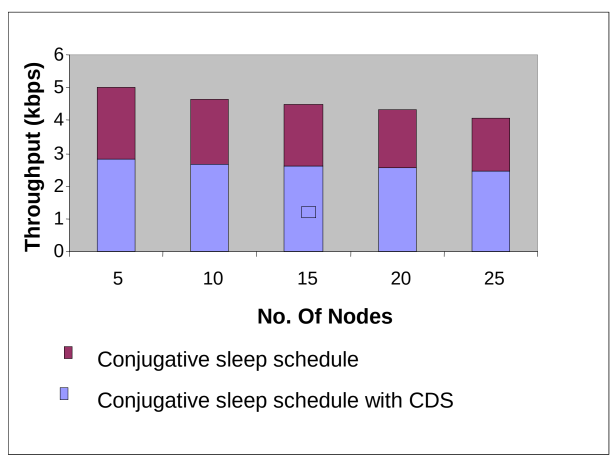


Fig 3. No. of Nodes Vs Throughput

Table 3: Throughput values of proposed and existing model

Number of Nodes	*Throughput (kbps)*	
	Conjugate Sleep schedule with CDS	*Conjugate Sleep schedule*
5	2.8	2.2
10	2.65	2
15	2.6	1.9
20	2.55	1.8
25	2.45	1.6

5. Conclusion

The conjugative scheduler scheme for data aggregation in wireless sensor network proposed an energy efficient routing approach. The properties of both connected dominating set and Minimum Spanning Tree for shortest path tree routing schemes are used. Results shows us that the life time increases with the application of CDS even though node density increases.. So sleep schedule algorithm with CDS is more suitable for the application requiring more life time.

This paper combines the features of Connected Dominating Set (CDS) with Conjugative Sleep Schedule Algorithm. To update the connected dominating set, two different topological conditions are considered. An example of changing re-computation period dynamically in a centralized solution can be found in the area size and the maximum transmission range is usually set by the application itself. By the application of distributed sleep scheduling algorithm, performance improved on high density networks.

Based on the simulation topologies and by variation of the sensor nodes to different densities (i.e., 20, 40, 60 and 80) the performance of conjugative power efficient scheduling scheme with and with out CDS is evaluated. As a result of better lifetime, better energy savings achieved in wireless sensor networks.

References

[1] J. Chang and L. Tassiulas, “Maximum lifetime routing in wireless sensor networks,” in Proceedings of Advanced Telecommunications and Information Distribution Research Program, College Park, MD, 2000.
[2] K. Kalpakis, K. Dasgupta, and P. Namjoshi, “Maximum lifetime data gathering and aggregation in wireless sensor networks,” in Proceedings of the 2002 IEEE International Conference on Networking (ICN’02), August 2002, pp. 685–696.
[3] Huseyin Ozgur Tan, Ibrahim Korpeoglu and Ivan Stojmenovic,” Computing Localized Power Efficient Data Aggregation Trees for Sensor Networks “ in IEEE Transactions on Parallel and Distributed Systems.
[4] C. E. Perkins and E. M. Royer, “Ad-hoc on-demand distance vector routing,” Proceedings of the 2nd IEEE Workshop on Mobile Computing Systems and Applications, p. 90, 1999.
[5] C. Intanagonwiwat, R. Govindan, and D. Estrin, “Directed diffusion: a scalable and robust communication paradigm for sensor networks,” in Mobile Computing and Networking, 2000, pp. 56–67.
[6] H. O¨ . Tan and I. Korpeoglu, “Power efficient data gathering and aggregation in wireless sensor networks.” SIGMOD Record, vol. 32, no. 4, pp. 66–71, 2003.
[7] R. Prim, “Shortest Connecting Networks and Some Generalizations,” Bell Syst. Tech. J., vol. 36, pp. 1389–1401, 1957.
[8] S. Upadhyayula and S. K. S. Gupta, “Spanning tree based algorithms for low latency and energy efficient data aggregation enhanced convergecast (dac) in wireless sensor networks,” Ad Hoc Netw., vol. 5, no. 5, pp. 626–648, 2007.
[9] V. Rodoplu and T. Meng, “Minimum energy mobile wireless networks,” IEEE J. Select. Areas Commun. vol. 17, no. 8, pp. 1333 – 1344, Aug. 1999.
[10] G. Toussaint, “The relative neighborhood graph of a finite planar set,” Pattern Recognition, vol. 12, pp. 231–268, 1980.
[11] N. Li, J. C. Hou, and L. Sha, “Design and analysis of an MST-based topology control algorithm.” in INFOCOM, 2003.
[12] J. Bachrach and C. Taylor, “Localization in sensor networks,” in Stojmenovic I, editor. Handbook of sensor networks: algorithms and architectures. Wiley, 2005, pp. 277–310.
[13] Jie Wu and Hailan Li,” On Calculating Connected Dominating Set for Efficient Routing in Ad Hoc Wireless Networks,” in ACM, 1999.
[14] H. O. Tan, I. Korpeoglu, and I. Stojmenovic, “A distributed and dynamic data gathering protocol for sensor networks,” in AINA ’07: Proceedings of the 21st International Conference on Advanced Networking and Applications. Niagara Falls, CA: IEEE Computer Society, 2007, pp. 220–227.
[15] S. Lindsey and C. S. Raghavendra, “Pegasis: Power-efficient gathering in sensor information systems,” in IEEE Aerospace Conference, March 2002.

K.P.Sampoornam was born in Erode on 16th May 1969.She Received the B.E degree in Electronics and Communication Engineering from V.L.B.Janakiammal College of Engineering, Coimbatore and M.E degree in VLSI Design from Government College of Technology, Coimbatore , TamilNadu on 1990 and 2005 respectively. She has been doing Ph.D in Electronics Engineering under Anna University of Technology, Coimbatore in part time.She worked as a lecturer in ECE department in Al-Ameen Polytecnic College from 1993 to 1997.Then she worked as a lecturer in ECE department in MPNMJ Polytechnic College & MPNMJ Engineering College from 1997-2007.After that she worked as an Assistant professor in ECE department from 2007 to 2009 and an Associate professor in ECE department from 2009 to till now in K.S.R. College of Engineering, Tiruchengode, TamilNadu, India. She has presented a paper in National conference which was conducted by S.A.Engineering College-Chennai. Her current research interests are in the areas of wavelet analysis, watermarking, image processing and Wireless sensor networks. Ms. K. P. Sampoornam is a life member of ISTE.

K.Rameshwaran was born in Ramanathapuram, Tamilnadu on 1st June 1958. He obtained his B.E. degree in Electronics & Communication Engineering from the University of Madras in 1980. He obtained his M.E.Degree in Electronics Engineering from Anna University, Chennai in 1982 and his Ph.D. degree from I.I.T.Madras, Chennai.

He started his professional career with a brief stint at I.I.T. Madras during 1982-1983 as a Project Engineer. He joined the department of Electrical Engineering at the Thiagarajar College of Engineering, Madurai as an Associate Lecturer in July 1983. Later, he joined the department of Electronics and Communication Engineering at the erstwhile Regional Engineering College **(Presently known as National Institute of Technology),** Tiruchirappalli in 1987. During the period between July 2006 and June 2008, he worked as the Principal of K.S.R. College of Engineering, Tiruchengode in Namakkal (District), and Tamilnadu. Consequent to his retirement on Voluntary basis (VRS) from NITT in December 2009, he joined as the Principal of R.M.K. Engineering College, Kavaraipettai-601 206 and worked for a brief period of 7 (Seven) months. Now he has been working as the principle of JJ College of Engineering and Technology, Ammapettai, Tiruchirappalli-620 009. He has published several research papers in International and National Journals. He has also presented research papers in National and International conferences. His areas of interest are: Digital system and Microprocessors, Digital Filters and Control theory.

A Fuzzy Neural Clustering approach for Fingerprint Recognition

Mahendra Kondekar[1], Uday Kulkarni[2] and Santosh Chowhan[3]

**[1] College of Computer Science and
Information Technology, Latur-413512
Maharshtra, India.**

**[2] SGGSIE&T, Nanded,
Maharastra, India.**

**[3] College of Computer Science and
Information Technology, Latur-413512
Maharshtra, India.**

Abstract

Fingerprint recognition is one of the most promising and evergreen biometric recognition technique used presently. In this paper we have proposed Extended Fuzzy Hyperline Segment Clustering Neural Network (EFHLSCNN) with its learning algorithm, which utilizes fuzzy sets as pattern clusters. In this extended version of paper we have used Manhattan distance for calculating distance of hyperline segments. The performance of EFHLSCNN when verified with fingerprint feature vectors, it is found superior than fuzzy hyperline segment clustering neural network (FHLSCNN) proposed by Kulkarni and Sontakke in terms of higher recognition rate and generalization.

Keywords: *Biometrics, Pattern Clustering, Fuzzy Neural Network, FingerCode.*

1. Introduction

Biometrics is the science of identifying individuals by a particular physical characteristic such as voice, eye, fingerprints, height, facial appearance, iris texture, or signature. Fingerprint based personal identification is routinely used in forensic laboratories and identification units around the world [1] and it has been accepted in the courts of law for nearly a century [2, 3]. Fingerprint features are permanent and fingerprints of an individual are unique [15]. Here in this paper we have used one of the most enthusiastic approaches to computer-based pattern recognition i.e. use of fuzzy neural networks for clustering feature patterns. They have been successfully used in many pattern recognition problems [4], [5], [6].

Cluster is a group of patterns having some common properties. Patterns can be grouped into clusters by some predefined criterion. As mentioned by Bezdek [7] the clusters can be formed according to some criterion like distance, angle, curvature, symmetry, connectivity, and intensity. Patterns which are similar are allocated to the same cluster, while the patterns which differ significantly are put in different clusters. Regardless of the clustering method the final result is always a partition of patterns in disconnected or overlapped clusters [10]. The choice of the proper grouping metric is only one aspect of the clustering problem. The fuzzy min-max (FMM) clustering and classification neural network algorithms [11], [12], with their representation of classes as hyperboxes in n-dimensional pattern space and their conceptual simplicity simple but powerful learning process, provided a natural basis for our paper. The derivatives of the original FMM can also be found in [13] and [14]. U. V. Kulkarni and T. R. Sontakke [8] also have proposed FHLSCNN.

In this paper they have used Euclidian distance for calculating distance between hyperline segments. In this extended version of FHLSCNN we have improved the recognition rate of patterns by using Manhattan distance metric. The patterns used for classification and clustering are of Poly U HRF Fingerprint database images of 320*240 sizes at 1200 dpi resolution. The feature extraction process is based on FilterBank based FingerCode feature extraction algorithm. In this algorithm they have used eight different values for with respect to the x-axis θ $(0^0, 22.5^0, 45^0, 67.5^0, 90^0, 122.5^0, 135^0, \text{and } 157.5^0)$. The normalized region of interest in a fingerprint image is convolved with each of these eight filters to produce a set of eight filtered features. These eight directional-sensitive filters capture most of the global ridge directionality information as well as the local ridge characteristics present in a fingerprint. The mean of each sector in each of the eight filtered features defines the components of FingerCode feature vector.

2. Topology of the EFHLSCNN

The EFHLSCNN consists of three layers as shown in Figure 1. The F_R layer accepts an unlabeled input pattern and consists of n processing elements, one for each dimension of the pattern. The F_E layer consists of m hyperline segments that are constructed during training and each node is characterized by the extended hyperline segment membership function as shown in Figure 2. One connection represents one end point for that dimension and the other connection represents another end point of that dimension, for a particular hyperline segment as shown.

The end points of hyperline segments are stored in V and W matrices. Each node of F_C layer represents a cluster and is constructed during training. The transfer function of F_C node performs the union of appropriate hyperline segments. The weights assigned between F_E and F_C layers are stored in the U matrix.

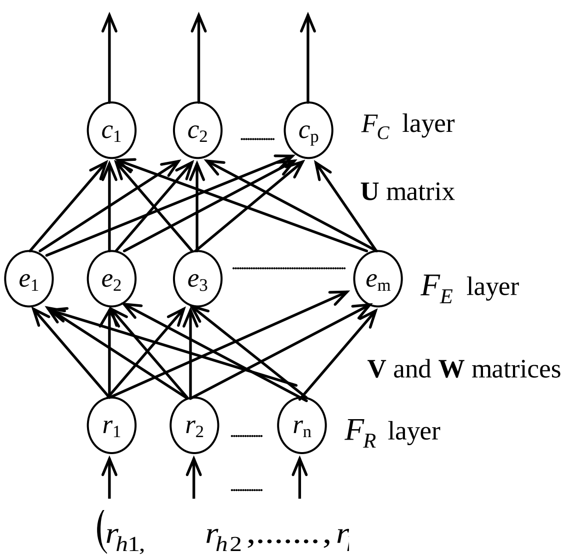


Fig. 1 Extended Fuzzy hyperline segment clustering neural network.

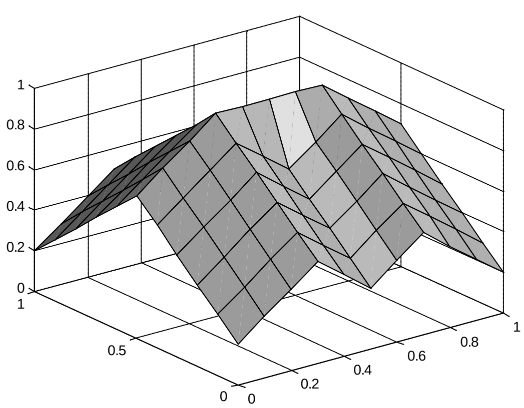


Fig. 2 The plot of extended fuzzy hyperline segment membership function for $\gamma_1 = 1$ with end points w = [0.5 0.3] and v = [0.5 0.7].

3. The EFHLSNN Learning Algorithm

The learning algorithm consists of three steps, creation of hyperline segments, clustering hyperline segments and intersection test. These three steps are described below in detail.

Creation of hyperline segments: The maximum length of hyperline segment is controlled by the parameter, ξ, bounded by $0 \le \xi \le 1$. The value of ξ should be moderately high so that the created hyperline segments will include the patterns, which are close to each other and possibly falling in the same cluster. Assuming $R \in \{R_h \mid h = 1, 2, .., P\}$, where $R_h = (r_{h1}, r_{h2}, ..., r_{hn}) \in I^n$ is the h_{th} pattern belonging to the training set R, the learning process begins by initializing first hyperline segment by the first pattern and then applying unlabeled patterns one by one from the pattern set. The applied pattern is tested for inclusion by calculating the fuzzy membership value with the already created hyperline segments having same end points.

Let $R_h = (r_{h1}, r_{h2}, ..., r_{hn})$ is the h_{th} input pattern, $V_j = (v_{j1}, v_{j2}, ..., v_{jn})$ is one end point of the hyperline segment e_j and $W_j = (w_{j1}, w_{j2}, ..., w_{jn})$ is the other end point of e_j. The fuzzy hyperline segment membership function of j_{th} F_E node is defined as

$$e_j(R_h, V_j, W_j) = 1 - f^3(x, \gamma_1, l), \qquad (1)$$

in which $x = l_1 + l_2$ and the distances l_1, l_2 and l are defined as (2), (3) and (4).

Here in this paper we have used Manhattan distance for computing the values of l_1, l_2 and l as shown in equation 5, 6 and 7 which has given best performance in comparison with Euclidian distance [8] as shown below.

$$l_1 = \max\left(\sum_{i=1}^{n} \left| w_{ji} - r_{hi} \right|\right) \qquad (2)$$

$$l_2 = \max\left(\sum_{i=1}^{n} \left| v_{ji} - r_{hi} \right|\right) \qquad (3)$$

$$l = \max\left(\sum_{i=1}^{n} \left| w_{ji} - v_{ji} \right|\right) \qquad (4)$$

and $f^3(\cdot)$ is a three-parameter ramp threshold function defined as,

$$f^3(x,\gamma_1,l)=0 \qquad \text{if } x=l \text{ otherwise}$$

$$f^3(x,\gamma_1,l)=\begin{cases} x\gamma_1 & \text{if } 0\le x\gamma_1\le 1 \\ 1 & \text{if } x\gamma_1>1. \end{cases}$$

The parameter γ regulates how fast the membership value decreases when the distance between R_h and e_j increases. If the fuzzy membership value calculated is grater than or equal to ξ for any one hyperline segment then the pattern is included by extending that hyperline segment else new hyperline segment is created.

Clustering hyperline segments: The number of clusters constructed depends on the parameters α_2 and β, called as centering and bunching factors, respectively and the values of these parameters are problem dependent. The clustering consists of three steps, I: Determining the centroid, II: Bunching and III: Removal of bunched patterns and hyperline segments. These three steps are described below in detail.

Determining the centroid: To determine the centroid of the cluster, all the patterns are applied to each of the hyperline segments and the patterns that give fuzzy membership larger than α are counted for all hyperline segments. If e_j is the hyperline segment with the maximum count then the centroid is computed as

$$(w_{ji}+v_{ji})/2 \quad \text{for} \quad i=1,2,..........,n. \tag{5}$$

Bunching: The hyperline segments, which are falling around the centroid and give fuzzy membership value larger than β are bunched in a cluster. Thus the cluster boundaries (i.e. the number of clusters) are governed by the value of bunching factor. As the clusters are being formed the weights in U matrix are also updated as

$$u_{jk}=\begin{cases} 1 & \text{if } e_j \text{ is the hyperline segment} \\ & \text{of the cluster } c_k, \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

for $k=1,2,......,p$ and $j=1,2,.......,m$.

Removal of bunched patterns and hyperline segments: The clustered hyperline segments in previous step and the patterns included by these hyperline segments are eliminated. Thus, the next pass uses remaining unclustered hyperline segments and pattern set consisting of remaining patterns for clustering. These three steps are repeated till all the created hyperline segments are clustered.

Let R_p, R_c and R_n represent set of patterns used in the current pass, set of patterns clustered in the current pass and set of patterns that will be used in the next pass, respectively. Then R_n can be described as,

$$R_n=R_p-R_c=\left\{R_n \mid R_n\in R_p \text{ and } R_n\notin R_c\right\} \tag{7}$$

The R_n calculated in the current pass becomes R_p for the next pass.

Each node of F_C layer represents a cluster. It gives soft decision and the output of *k*th F_C node represents the degree to which the input pattern belongs to the cluster c_k. The transfer function of each F_C node performs the union of appropriate (of same cluster) hyperline segment fuzzy values, which is described as,

$$c_k=\max_{j=1}^{m} e_j\, u_{jk} \qquad \text{for } k=1,2,........,p. \tag{8}$$

Intersection test: The learning algorithm allows the intersection of hyperline segments from the same cluster and eliminates the intersection between the hyperline segments from separate clusters. If the two hyperline segments from different clusters are intersecting then the intersection is removed by breaking one of the hyperline segment.

4. Simulation Results

The EFHLSCNN is trained with fingerprint feature vector data by setting $\alpha=0.9$, and $\gamma=1$. The value of ξ is set moderately large so that the EFHLSCNN algorithm will create hyperline segments of patterns possibly falling in the same cluster. We have adjusted the value of α close to one so that while computing the centroids, patterns falling around and close to hyperline segment are counted. These values have resulted in the creation of 295 hyperline segments.

The performance of EFHLSCNN algorithm is tested with fingerprint feature vector data. The experiments are carried out with $\alpha=0.9$, $\xi=0.9$, $\gamma=1$ and by varying the bunching factor β. These results are tabulated in the Table 1. The centering and bunching factors are fixed to moderately high value so that EFHLSCNN creates hyperline segments of patterns possibly falling close to each other and belonging to the same cluster and while finding the centroids it will count patterns falling around hyperline segment under consideration. As β increases the number of hyperline segments bunched in cluster decreases, which leads to increase in number of clusters. The performance of EFHLSCNN algorithm is compared with FHLSNN [9], EFHLSNN [14] and FHLSCNN algorithm. The results are depicted in Table 2

The experiments are carried out using fingerprint feature vector and the results obtained are tabulated in Table 2. The timing analysis is also depicted in Table 2.

The FHLSNN algorithm created 200 hyperline segments when trained with the parameters $\theta = 0.14$ and $\gamma = 1$. The results delivered by FHLSNN algorithm are tabulated in the first row of Table 2. This row indicates that the FHLSNN algorithm gives better results with less number of hyperline segments.

Recognition Rate	β	No. of clusters
60.25	0.6	124
68.25	0.65	148
77.75	0.7	174
88.25	0.75	206
92.75	0.8	237
98.00	0.85	263
100	0.9	286

Table 1: The percentage recognition rates obtained from EFHLSCNN algorithm using fingerprint feature vector with number of created clusters

The results obtained from FHLSCNN algorithm for 268 clusters are depicted in the third row of Table 2 with 87.75 recognition rate with 5.133608 seconds recall time. The proposed algorithm gives the recognition rate of 100 percent with 128.9859 and recall time of 4.279710 which is less as compared to FHLSCNN.

Algorithm	Avg.	Training Time	Recall time per pattern
FHLSNN	100	0.1648	0.811566
EFHLSNN	100	0.1631	0.743732
FHLSCNN	87.75	128.9859	5.133608
EFHLSCNN	100	246.4701	4.279710

Table 2: The percentage recognition rates and timing analysis using fingerprint feature vector

The results are depicted in fourth row of Table 2. The experimental result confirms that the EFHLSCNN algorithm generalizes well and yields highest average percentage recognition rates than the FHLSCNN algorithm except increase in training time and less recall time.

These results indicate that EFHLSCNN algorithm gives best average percentage recognition rate as compared to FHLSCNN algorithms using Manhattan distance for calculating distance of hyperline segments. The improvement in generalization performance indicates that clusters are created properly in the pattern space.

Finally, the Table 1 and 2 indicates that the results obtained using fingerprint feature vectors are superior in terms of recognition rate and generalization when we are solving clustering problems.

5. Conclusions

A new extended approach of clustering that utilizes hyperline segments as fuzzy sets that are aggregated into fuzzy set clusters with revision is introduced. The performance of EFHLSCNN algorithm is found superior compared to FHLSCNN algorithm when applied to clustering of fingerprint feature data.

It is observed that the EFHLSCNN algorithm generalizes fit and produces highest average percentage recognition rates as compared with the FHLSCNN algorithm except increase in training time.

6. References

[1] H.C. Lee and R.E. Gaensslen, “Advances in Fingerprint Technology”, second ed., New York: CRC Press, 2001.
[2] Latent Print Examination, http://onin.com/fp/, 2002.
[3] Robert S. Germain, Andrea Califano, and Scott Colville, “Fingerprint Matching Using Transformation Parameter Clustering”, IEEE Computational Science & Engineering, October-December 1997.
[4] Bezdek, J.C. “Computing with uncertainty”, IEEE Communication Mag., pp. 24–36, Sept. 1992
[5] M. H. Hassoun, “Fundamentals of Artificial Neural Networks”, Cambridge, MA: MIT Press, 1995.
[6] R. Krishnapuram, “Generation of membership functions via possibilistic clustering”, in Proc. 1994 IEEE 3rd Int. Fuzzy Systems Conf., vol. 2, June 1994, pp. 902–908.
[7] J. Bezdek, “Pattern recognition with fuzzy objective function algorithms”, New York: Plenum Press, 1981.
[8] Kulkami U. V, Sontakke T. R, Kulkami A. B, “Fuzzy hyperline segment clustering neural network”. IEEE Electronics Letters, 37 (S) , pp 301-303,2001.
[9] U.V. Kulkarni, T.R. Sontakke and G. D. Randale, “Fuzzy Hyperline Segment Neural Network for Rotation Invariant Handwritten Character Recognition”, Neural Networks, 2001. Proceedings. IJCNN '01. International Joint Conference on Neural Network, Washington, DC, USA, 2001, pp. 2918 - 2923 vol.4.
[10] Bogdan Gabrys and Andrzej Bargiela, “General Fuzzy Min-Max Neural Network for Clustering and Classification”, IEEE

TRANSACTIONS ON NEURAL NETWORKS, VOL. 11, NO. 3, MAY 2000

[11] Patrick K. Simpson, "Fuzzy min-max neural networks—Part 1: Classification", IEEE, Trans. Neural Networks, vol. 3, pp. 776–786, Sept. 1992.

[12] Patrick K. Simpson, "Fuzzy min-max neural networks—Part 2: Clustering", IEEE Trans. Fuzzy Syst., vol. 1, pp. 32–45, Feb. 1993.

[13] A. Joshi, N. Ramakrishman, E. N. Houstis, and J. R. Rice, "On neurobiological, neurofuzzy", machine learning, and statistical pattern recognition techniques," IEEE Trans. Neural Networks, vol. 8, Jan. 1997.

[14] M. Meneganti, F. S. Saviello, and R. Tagliaferri, "Fuzzy neural networks for classification and detection of anomalies", IEEE Trans. Neural Networks, vol. 9, Sept. 1998.

[15] M. H. Kondekar, Dr. U. V. Kulkarni & S. S. Chowhan, "Fingerprint Recognition Using Extended Fuzzy Hypersphere Neural Network", JOURNAL OF COMPUTING, VOLUME 3, ISSUE 4, APRIL 2011, ISSN 2151-9617,pp. 101-105.

M. H. Kondekar received the M.C.A. degree from Government Engineering College, Aurangabad in the year 2000. He received the M.Phil. Degree in Computer Science from Y.C.M.O. University, Nashik in the year 2009. He is currently working as Lecturer in the College of Computer Science and Information Technology, Latur, Maharastra, India. His current research interests include various aspects of Neural Networks and Fuzzy Logic, Pattern Recogntion and Biometrics.

Dr.U.V.Kulkarni received the Ph.D. degree in Electronics and Computer Science Engineering from S.R.T.M. University, Nanded in the year 2003. He is currently working as professor in Computer Science and Engineering Department, SGGS Institute of Engineering and Technology, Nanded, Maharastra, India.

S. S. Chowhan received the M.Sc.(CS) degree from Dr. BAM University, Aurangabad in the year 2000. He received the M.Phil. Degree in Computer Science from Y.C.M.O. University, Nashik in the year 2008. He is currently working as lecturer in the College of Computer Science and Information Technology, Latur, Maharastra. His current research interests include various aspects of Neural Networks and Fuzzy Logic, Pattern Recogntion and Biometrics.

Secure Biometric Key Generation Scheme for Cryptography using Combined Biometric Features of Fingerprint and Iris

Mr.P.Balakumar[1] and Dr.R.Venkatesan[2]

[1] Associate Professor, Department of Computer Science and Engineering, Selvam College of Technology, Namakkal, Tamilnadu, India

[2] Professor and Head, Department of Computer Science and Engineering, PSG College of Technology, Coimbatore, Tamilnadu, India

Abstract

Exact and automatic recognition and authentication of users are a essential difficulty in all systems. Shared secrets like Personal Identification Numbers or Passwords and key devices such as Smart cards are not presently sufficient in few situations. What is required is a system that could authenticate that the person is actually the person. The biometrics is improving the capability to recognize the persons. The usage of biometrics system permits the recognition of a living person according to the physiological features or behavioral features to be recognized without human involvement. This leads to the world wide usage of biometrics to secure the system. The various biometrics used in securing system are fingerprint, iris, retina, etc. The construction of cryptographic key from biometrics is used generally to secure the system. The efficiency and the flexibility of the cryptographic make it suitable for securing purpose. In some times, biometrics can be stolen; this makes the attackers to access the system for any time. This problem is diminished in this paper by using two biometrics features. The biometrics used in this paper is fingerprint and iris. These two features are combined with the help of fusion algorithm. From the combined features, cryptographic key is generated. The experimental result shows that the proposed techniques results in better security than the existing techniques.

***Keywords*—**Biometrics, Cryptography Key Generation, Minutiae Points, Security Analysis

1. Introduction

Information security and privacy has become an important factor in the present world. Biometric recognition is one of the most important techniques for the security privacy due to its distinctive nature of biometric [5] traits such as fingerprints, iris and faces [7]. As a result, this technique is used with many other applications to enhance the security. Cryptographic techniques have gained its popularity due to its security purpose. In the cryptographic technique the original data is encoded by using any key so that it is not in a understandable format for the attacker. The original data can be obtained by decoding the encoded data using the same key. Thus the privacy is well protected in this cryptographic approach. Several cryptographic techniques like DES, AES and public key architectures like RSA are widely used for the authentication purpose.

The characteristic feature of cryptographic security is conditioned by an authentication step that depends on long pseudo-random keys (128 bits in symmetric encryption), which are very impossible to keep in mind. This feature of inability to remember cryptographic keys [8] has been restraining the security of systems for a long time. The inability of human users to remember powerful cryptographic keys has been a feature restraining the security of systems for decades.

It's the natural tendency of humans to set passwords [6, 13] that are usually recognized or deduced by any social engineering methods. Typically people usually store keys in a place that is insecure and can possibly be shared among users and therefore it is not capable of ensuring non-repudiation. Moreover it's a natural human tendency to use same keys or password for a variety of applications and as a result, if one system is hacked it is very easy to hack all the systems corresponding to that key. This practically reduces the security privacy and makes the work easy for the hacker. Cryptographic techniques when combined with the biometric approach are used to solve these problems and provide security. The cryptographic keys are produced from the biometric data and are used in the authentication checking.

Biometric technique [17] provides the distinct characteristics of a person which is always prevalent. A person's individuality can be differentiated from one or more behavioral or physiological features by this authentication technique. Various techniques that are under the biometric research include facial, palm prints, retinal and iris scans, and hand geometry, signature capture and vocal features.

Biometric cryptosystems is a new technique which combines biometrics and cryptography [2], and is popularly known as crypto-biometric systems. The integration of biometrics [16] and cryptography is broadly carried out in two distinct steps. In case of biometrics-based key generation, a biometric matching amid an input biometric signal and a registered template is utilized in the release of the secret key.

In biometrics resetting is very much complicated. One of the huge merits of the biometric data over time is its uniformity which is also the demerit at the same instant. In case of any conventional techniques, like credit card, it is possible to issue a new one, if it is lost. But it is impossible to substitute the biometric characteristics and it is fully evident since it is not feasible to provide a person with a fresh biometric feature once it is stolen.

This problem can be solved by the approach called cancellable biometric. This procedure uses a predefined transform and thus provides the intended and repeatable distortion of a biometric signal. This approach thus makes expose cross matching unachievable by facilitating the every incidence of enrollment to utilize a distinct transform. Furthermore, it is just enough to merely change the transform operation to produce a new variant for re-enrollment, if a variant of the transformed biometric data is comprised. Generally, the non-invertible transforms are utilized for distortion. Thus it is impossible for the hacker or the unauthenticated user to recover the original biometrics without knowing the transform method and the resulting transformed biometric data.

This paper uses two biometrics features to generate the cryptography key [3, 4]. The biometrics used in this paper is fingerprint [11] and iris [18]. These two biometrics features are combined using a technique called fusion. From these combined features, cryptography key is generated in this paper.

2. Related Works

The proposed work is inspired from a number of researches which are related to cryptography and cancellable biometric techniques. Goh and Ngo combined have proposed a new system based on face biometrics [5]. The work adopted the biometric locking approach of Soutar et al. Here the features are the Eigen-projections which are extracted from the face image, each of which is then mixed with a random string and quantized into a single bit.

Cancellable biometrics gives a better performance of security as it facilitates with more than one template for the same biometric data. Ang et al. [1] proposed the measurement of the success of a particular transformation and matching algorithm for fingerprints. A key-dependent cancellable template for the fingerprint was produced by employing a key dependant geometric transform on the obtained fingerprint features. Besides, the performance evaluation of an authentication system that utilizes the cancellable biometric is studied and it was found that the performance of the cancellable biometric was significant.

Hao et al. [10] presented a realistic and secure way to incorporate the iris biometric into cryptographic applications. They deliberated on the error patterns within iris codes and developed a two-layer error correction technique that merges Hadamard and Reed-Solomon codes. The key was produced from the iris image of the subject through the auxiliary error correction data that do not disclose the key and can be saved in a tamper-resistant token like a smart card. The evaluation of the methodology was performed with the aid of samples from 70 different eyes, 10 samples being obtained from every eye. It was established that an error-free key can be reproduced reliably from genuine iris codes with a success rate of 99.5 percent. It is possible to produce up to 140 bits of biometric key, more than adequate for 128-bit AES.

An on-line signature-based biometric authentication system, where non invertible transformations were applied to the acquired signature functions ruling out the possibility to derive the original biometrics from the stored templates at the same time maintaining the same recognition performances of an unprotected system was projected by Maiorana et al. Precisely the probability of producing cancellable templates from the same original data, thereby offering an appropriate solution to privacy concerns and security problems was intensely explored.

Teoh et al. [15] have presented a two-factor cancellable formulation that facilitates data distortion in a revocable yet non-reversible manner by first converting the raw biometric data into a fixed-length feature vector followed by the projection of the feature vector onto a sequence of random subspaces that were obtained from a user-specified Pseudorandom Number (PRN). The process was revocable making the replacement of biometrics seem as easy as replacing PRNs. This formulation was confirmed under numerous scenarios (normal, stolen PRN, and compromised biometrics scenarios) with the aid of 2400 Facial Recognition Technology face images.A cancellable biometric approach called PalmHashing was projected by T. Connie et al. [5] in order to address the non revocable biometric issue. This technique hashes palmprint templates with a set of pseudo-random keys to acquire a unique code known as the palmhash.

A fuzzy commitment method working on lattice mapping for cryptographic key generation from biometric data was proposed later. Despite providing high entropy keys as output the method as well obscures the original biometric data such that it becomes unfeasible to recover the biometric data besides the stored information in the system being open to an attacker. Results of simulation illustrated that the method's authentication accuracy was analogous to that of the renowned. For cancelable biometrics, the main scheme is to store an irreversibly transformed version of the biometric template which provides a high privacy and security level by allowing multiple templates to be associated with the dame biometric data [7].

Jo et al. [12] proposed a simple technique for the generation of digital signatures and cryptography communication with the aid of biometrics. The generation of the signature is necessary in such a way that it becomes possible to verify the same with a cryptographic algorithm in existence like the RSA without altering its own security constraint and infrastructure.

3. Proposed Methodology

Biometric cryptosystems combines cryptography and biometrics to afford the advantages of both for security. This technique will provide the advantages like better and modifiable security levels which are the advantages of cryptography and advantages like eliminating the must to memorize passwords or to carry tokens etc which are the advantages of using biometrics. This paper combines the features of fingerprint and iris and with that combined feature, cryptography key is generated.

3.1. Feature Extraction from Fingerprint

Figure.1 represents a general procedure of extracting the minutiae points from fingerprint taken from the user.

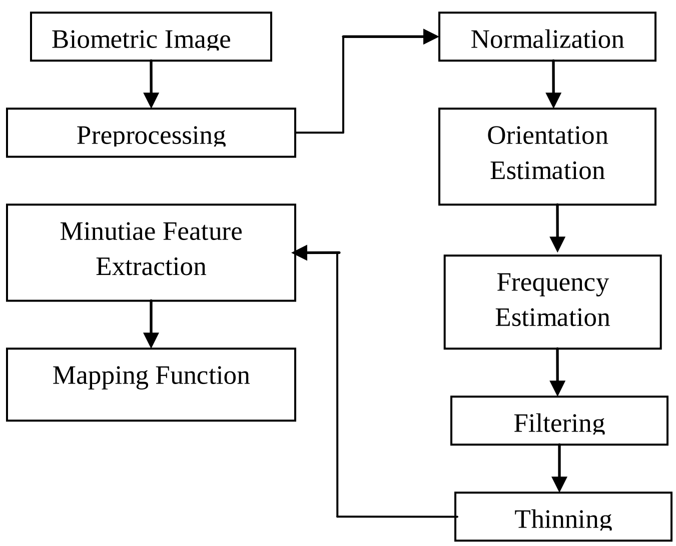


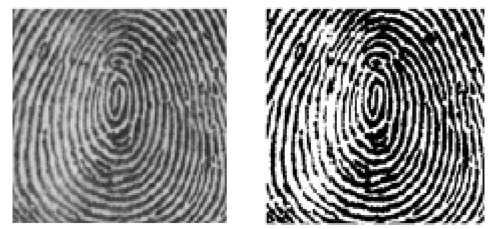
Figure 1. Steps Involved in Extracting Feature Point

A. Fingerprint Enhancement

This is usually required process in creating a security system with the help of biometrics. This process includes subsequent processing on the gathered fingerprint image. Fingerprint consists of sequence of ridges and furrows on the finger surface. This provides the individuality of the users fingerprint. No two fingerprints can have the similar existence of ridges and furrows. Minutiae points are local ridge features that appear at either a ridge bifurcation or a ridge ending. The ridges hold the information of features mandatory for minutiae extraction. Hence the clarity of the ridge occurrences in a fingerprint image must be very important. The gathered image is then enhanced with the help of image enhancement methods in order to diminish the noise in the image. The image enhancement methods used to enhance fingerprint image are normalization, orientation estimation, local frequency estimation, Gabor filtering, and thinning.

1 Normalization

Normalization technique is nothing but the standardization of the intensity values in an image by altering the range of gray-level values with the intention that it occurs within a preferred range of values. Additionally the ridge structure in the fingerprint does not undergo any alterations in its structure because of this processing. This process is performed in order to standardize the dynamic levels of dissimilarity in gray-level values that assist the processing of subsequent image improvement processes. Figure 2 represents an image of the fingerprint before and after normalization.

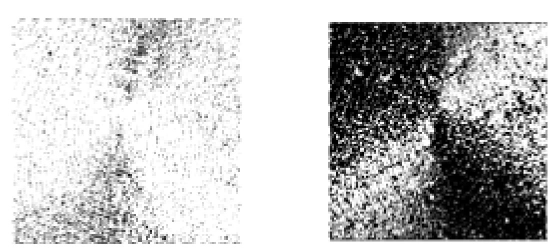


(a) (b)

Figure 2. (a) Original Image (b) Image after normalization

2 Orientation Estimation

The orientation estimation is a necessary process in the improvement process as the successive Gabor filtering stage depends on the local orientation for the purpose of effectively improve the fingerprint image. Figure 3 (a) and 3 (b) indicates the outcome of orientation estimation and smoothed orientation estimation of the fingerprint image correspondingly. Other than the orientation image, another significant parameter that is utilized in the building of the Gabor filter is the local ridge frequency.



(a) (b)

Figure 3. (a) Orientation Image (b) Smoothened Orientation Image

3 Gabor Filtering

As the ridge orientation and ridge frequency parameters are determined, the even-symmetric Gabor filter [9] can be created with the help of those parameters. Gabor filters are engaged since they contain frequency-selective and orientation selective assets. These assets permit the filter to be adjusted to provide maximal response to ridges at a particular orientation and frequency in the fingerprint image. Hence, a appropriately adjusted Gabor filter can be utilized to successfully maintain the ridge structures during noise removal process. Figure 4 represents the outcome of applying Gabor filter to a fingerprint image.

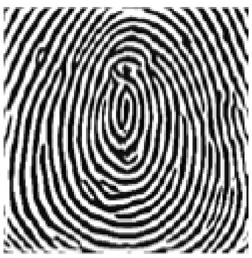


Figure 4. Filtered Image

4 Thinning

The final image improvement pace normally performed before minutiae extraction is thinning [14]. Thinning is a morphological process that consecutively takes away the foreground pixels till they are one pixel apart. By applying the thinning technique to a fingerprint image maintains the connectivity of the ridge structures during the formation of a skeleton stage of the binary image. This skeleton image is subsequently utilized in the following extraction of minutiae. Figure 5 shows the results of thinning to a fingerprint image.

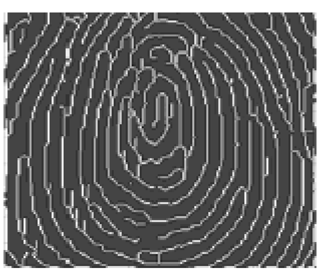


Figure 5. Thinned Image

B. Minutiae Feature Extraction

The next step is to obtain the minutiae from the thinned image. The most commonly used technique of minutiae extraction is the Crossing Number (CN) model. This process involves the utilization of the skeleton image in which the ridge flow pattern is eight-connected.

The minutiae are obtained by examining the local neighborhood of every ridge pixel in the image by means of a 3x3 window. The CN value is then calculated which is defined as partially the addition of the differences among the pairs of neighboring pixels in the eight-neighborhood. Figure 6 indicates the list of minutiae in a fingerprint image.

Figure 6. Minutiae extraction on a fingerprint image.

C. Mapping Function

The coordinate system utilized for the purpose articulating the minutiae point locations of a fingerprint is a Cartesian coordinate system. The X and Y coordinate of the minutiae points are in pixel units. Angles are represented in regular mathematical format, with zero degrees to the right and angles rising in the counter-clockwise direction. The obtained minutiae points are stored as below

$$F_1 = [x_1, x_2, \ldots x_n]$$

$$F_2 = [y_1, y_2, \ldots y_n]$$

Feature Extraction from Iris

Iris Localization and Normalization:

Canny edge detection is performed mutually in vertical direction and horizontal directions of the provided iris image. The radius of the iris image is determined and provided to the Hough transform. For better accuracy, the Hough transform is carried out initially for iris/sclera boundary and then for iris/pupil boundary. The outcome of this step results in storing the radius and x, y parameters of inner and outer circles.

Canny edge detection is utilized to build edges in horizontal direction and then Hough transform is applied on it. If the maximum Hough space is below the threshold then it indicates the non occlusion of eyelids. For isolating eyelashes it is very easy by utilizing thresholding. This is because they are darker while comparing with further elements in eye. The contrast of the eye image is improved with the help of histogram equalization.

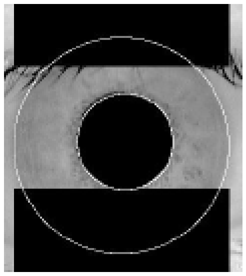


Figure 7.Localized iris image

The normal Cartesian to polar transformation is recommended which maps all the pixel in the iris area into a pair of polar coordinates(r, θ), where r and θ represents the intervals of [0 1] and [0 2π].

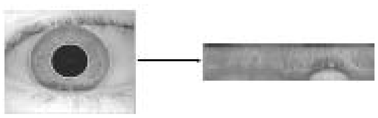
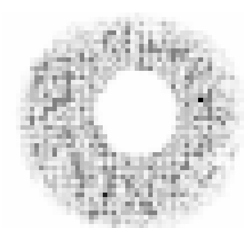


Figure 8. Normalized Iris

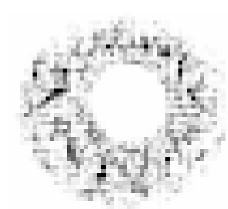
D. Extraction of Lock/Unlock Data

On the emphasized iris structures as a whole, the following order of morphological operations is utilized to mine the pseudo structures.

- Close - by - reconstruction top - hat (figure 9(a))
- Opening (figure 9(b)), area opening to remove structures in according to its size resulting image with structures disposed in layers (figure 9(c))
- Thresholding is applied to obtain binary image.



(a) Closing-by-tophat



(b) Opening

(c) Thresholded images

Figure 9. Morphological operations on Iris Textures

For suitable depiction of structures, thinning is utilized so that it presents all the structure itself as an agglomerate of pixels. It is represented in figure 10.

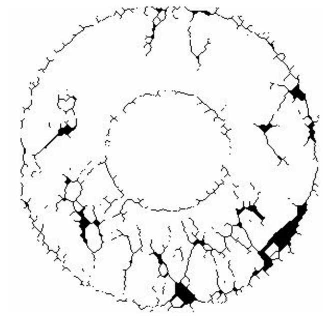


Figure 10. Iris textures after thinning operation

From the above iris rim containing iris pseudo textures, the polar coordinates of minutiae (nodes and end points of iris textures) are obtained by resizing the image into a standard format as represented in figure 11.

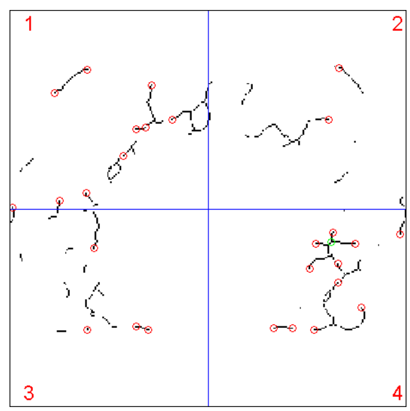


Figure 11. Minutiae representation of Nodes and end points are shown in circles.

These obtained Minutiae points are kept as

$$I_1 = [x_1, x_2, \dots\ x_n]$$

$$I_2 = [y_1, y_2, \dots\ y_n]$$

Fusion of Biometric Features

This phase will perform the fusion process for the gathered fingerprint and iris features. The input to the fusion process will be four vectors F_1, F_2, I_1 and I_2 which are obtained from fingerprint and iris. The steps involved in fusion of biometric feature vectors are as follows.

1. *Shuffling of Individual Feature Vectors*:

The initial step in the fusion process is the shuffling of all the individual feature vectors F_1, F_2, I_1 and I_2. The steps for performing shuffling of vector F_1 are as below,

i. A random vector R of size F_1 is created. The random vector R is controlled by the seed value.
ii. For shuffling the ith component of fingerprint feature vector F_1,
a) The ith component of the random vector R is multiplied with a large integer value.
b) The product value resulted is modulo operated with the size of the fingerprint feature vector F_1.
c) The obtained value is the index say 'j' to be interchanged with. The components in the ith and jth indexes are interchanged.
iii. Step (ii) is repeated for all the component of F1. The shuffled vector F1 is indicated as S1.
This procedure is repeated for all other vectors F2, I1 and I2 and represent it as S1, S2 and S3 respectively, where S2 is shuffled F2 and S3 is shuffled I1. The shuffling process results with four vectors S1, S2, S3 and S4.

2. *Concatenation of Shuffled Feature Vectors*:

The next process is to concatenate the shuffled vectors process S_1, S_2, S_3 and S_4. In this process, the shuffled fingerprints S_1 and S_2 are concatenate with the shuffled iris features S_3 and S_4 correspondingly. The concatenation of the vectors S_1 and S_3 is performed as below:

i. A vector M_1 of size $|S_1| + |S_3|$ is generated and its initial $|S_3|$ values are filled with S_3.
ii. For all the components in S_1 ,
a) The equivalent indexed component of M_1 say't' is selected.
b) Logical right shift operation is performed in M_1 from index't'.
c) The component of S_1 is inserted into the emptied tth index of M_1.
The above mentioned procedure is performed among shuffled vectors S_2 and S_4 to obtain a vector M_2. In this manner, the concatenation process yields two vectors M_1 and M_2.

3. *Merging of the Concatenated Feature Vectors*:

The final process in creating the biometric template B_T is the merging of two vectors M_1 and M_2. The process for merging the concatenated feature vectors is provided below.

i. For all the component of M_1 and M_2,
a. The components M_{11} and M_{21} are converted into their binary form.
b. Binary NOR operation is carried out among the components M_{11} and M_{21}.
c. The obtained binary value is then transformed back into decimal form.
ii. These decimal values are stored in the vector B_T that serves biometric template.

3.2 Generation of Cryptographic Key from Fused Features

The final process of the proposed technique is the creation of the k-bit cryptographic key from the obtained biometric template B_T. The template vector B_T can be indicated as,

$$B_T = [b_{T_1}\ b_{T_2}\ b_{T_3} \ldots\ b_{T_h}]$$

The set of different components in the template vector B_T are recognized and are stored in another vector U_{BT}.

$$U_{BT} = [u_1\ u_2\ u_3 \ldots\ u_d]\ ;\ |U_{BT}| \leq\ |B_T|$$

The vector U_{BT} is then resized to k components appropriate for creating the k-bit key. The resize procedure utilized in the proposed technique is

$$B = \begin{cases} [u_1\ u_1 \ldots\ u_k], & if\ |U_{BT}| > k \\ [u_1\ u_1 \ldots\ u_d] \ll u_i\ ; d+1 \geq i \geq k, & if\ |U_{BT}| < k \end{cases}$$

Where,

$$u_i = \frac{1}{d}\sum_{j=1}^{d} u_j$$

Finally, the key K_B is created from the vector B,

$$K_B \ll B_i\ mod\ 2, i = 1,2,3, \ldots k$$

This finally obtained key serves as an authentication key for the individual in the system. This key is definitely very difficult for the theft to generate. Therefore, a better secure system is created using the proposed technique. The evaluation for the proposed technique is provided in experimental result section.

4. Experimental Results

This section provides the evaluation of the proposed biometrics techniques. The fingerprint and iris are obtained from 100 persons are used for evaluation. Then the feature points are obtained from the gathered biometrics using the techniques presented in this paper. The biometrics features are obtained for fingerprint and iris separately. Then, these fingerprint features and iris features are combined using the technique fusion. The fusion process is carried out according the fusion method presented in this paper. Then the proposed system is evaluated using the parameters such as False Rejection Rate (FRR) and False Acceptance Rate (FAR)

Table 1: Average False Rejection Rate (FRR) (%) Comparison

User	Fingerprint	Iris	Proposed
1-10	92.9	89.5	85.6
11-20	92.1	89.6	84.6
21-30	93.6	89.9	85.1
31-40	94.5	92.1	86.2
41-50	92.8	90.6	86.7
51-60	89.6	87.6	84.2
61-70	90.6	89.4	83.5
71-80	92.7	90.9	84.1
81-90	91.1	90.1	85.6
91-100	91.6	89.2	85.1

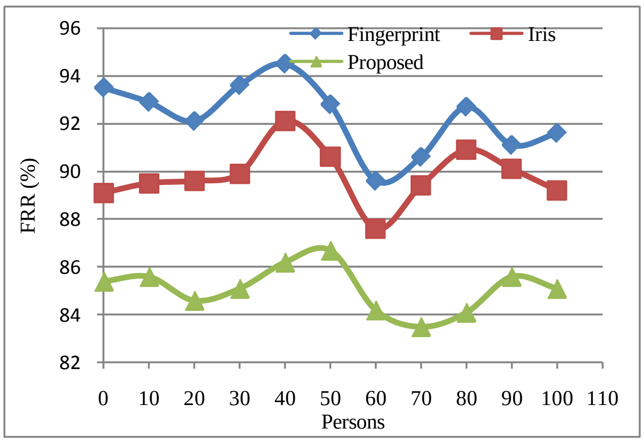


Figure 12. Resulted False Rejection Rate

Table 2 .False Acceptance Rate (FAR) (%) Comparison

User	Fingerprint	Iris	Proposed
1-10	0.45	0.39	0.11
11-20	0.42	0.38	0.12
21-30	0.43	0.39	0.09
31-40	0.39	0.38	0.11
41-50	0.38	0.37	0.08
51-60	0.39	0.37	0.08
61-70	0.44	0.41	0.09
71-80	0.41	0.38	1.02
81-90	0.39	0.37	1.12
91-100	0.40	0.39	0.98

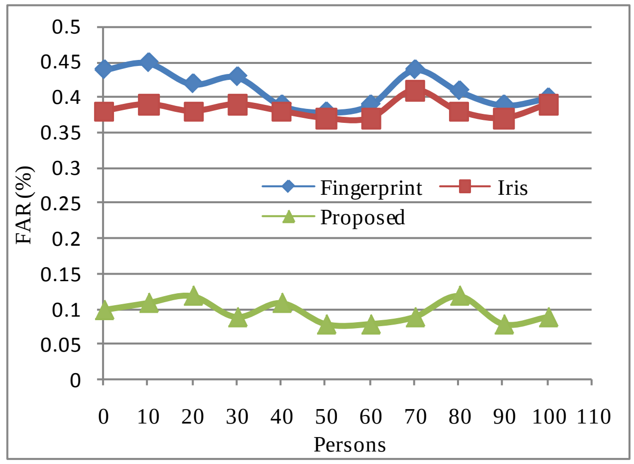


Figure 13. Resulted False Acceptance Rate

Table 1 and figure 12 shows the resulted False Rejection Rate (FRR) for the proposed and existing technique. From the result, it can be observed that the proposed technique results in lesser False Rejection Rate when compared to the existing technique. Table 2 and figure 13 shows the resulted False Acceptance Rate (FAR) for the proposed and existing technique. From the result, it can be observed that the proposed technique results in lesser False Acceptance Rate for all the persons, whereas the existing techniques results with higher percentage of False Acceptance Rate. From all the results obtained, it can be said that the proposed technique results in better security than the existing technique.

5. Conclusion

Securing the information system becomes most challenging task because of the increased number of theft. The conventional security system uses password or security key for authentication; but those password and security key can be easily stolen by the theft. To overcome these issues, biometrics of a person is used to secure the system. But, if the biometrics is stolen one time, it can be used by theft to access the system until it exists. This provides huge difficulty for the researchers to develop a new secure technique. One solution to this problem is usage of more than one biometrics for securing the system. This is because it is mostly impossible for the theft to steal more than one biometrics. This paper used fingerprint and iris biometrics to secure the system. The features obtained from these two biometrics are combined using fusion technique. From these fused features, cryptographic key is generated which is more secure than other techniques. The experimental result shows that the proposed security scheme results in better security than the existing techniques.

References

[1] R. Ang, R. Safavi-Naini, L. McAven, "Cancellable key-based fingerprint templates," ACISP 2005, pp. 242-252.
[2] Announcing the Advanced Encryption Standard (AES), Federal Information, Processing Standards Publication 197, Nov. 26, 2001.
[3] Y. J. Chang, Z. Wende, and T. Chen, "Biometrics- based cryptographic key generation," IEEE International Conference on Multimedia and Expo, vol. 3, p. 2203-2206, 2004
[4] Chen, and V. Chandran, "Biometric Based Cryptographic Key Generation from Faces," Proceedings of the 9th Biennial Conference of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications, pp. 394-401, 2007.
[5] T. Connie, A. Teoh, M. Goh, and D. Ngo, " Palm hashing: A novel approach for cancellable biometrics," Information processing letters, vol. 93, no. 1, pp. 1-5, 2005.
[6] Feldmeier and P. Karn. "UNIX password security Ten years later," Advances in Cryptology Crytpo '89, LNCS 435, pp. 44-63, Springer-Verlag, 1990.
[7] Y. C. Feng, P. C. Yuen, and A. K. Jain, "A hybrid approach for face template protection," Proceedings of SPIE Conference of Biometric Technology for Human Identification, Orlando, USA, vol. 6944, pp. ca. 325, 2008.
[8] M. F. Santos, J. F. Aguilar, and J. O. Garcia, Cryptographic key generation using handwritten signature," Proceedings of SPIE, vol. 6202, pp. 225-231, Orlando, Fla, USA, Apr. 2006.
[9] Gabor Filter. (http://en.wikipedia.org/wiki/Gaborlter)
[10] F. Hao, R. Anderson, and J. Daugman, "Combining crypto with biometrics effectively," IEEE Transactions on Computers, vol. 55, pp. 1081-1088, 2006.
[11] L. C. Jain, U. Halici, I. Hayashi, S.B. Lee, and S.Tsutsui, "Intelligent Biometric Techniques in Fingerprint and Face Recognition", CRC Press, 1999.
[12] J. G. Jo, J. W. Seo, and H. W. Lee, "Biometric digital signature key generation and cryptography communication based on fingerprint," First Annual International Workshop 2007, LNCS 4613, pp. 38-49, Springer Verlag, 2007.
[13] Klein, "Foiling the cracker: A survey of, and improvements to, password security," Proceedings of the 2nd USENIX Security Workshop, pp. 5-14, Aug. 1990.
[14] L. Lam, S. W. Lee, and C. Y. Suen, "Thinning methodologies-A comprehensive survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 14, no. 9, pp. 879, Sep. 1992.
[15] B. Teoh, and C. T. Yuang, "Cancellable biometrics realization with multispace random projections," IEEE Transactions on Systems, vol. 37, no. 5, pp.1096-106, 2007.
[16] Alexander P. Pons, and Peter Polak, "Understanding user perspectives on biometric technology," Communications of the ACM, vol. 51, no. 9, pp. 115-118, September 2008.
[17] N. K. Ratha, J. H. Connell, and R. M. Bolle "Enhancing security and privacy in biometrics based authentication systems", IBM Systems Journal, vol. 40, pp. 614-634, 2001.
S. C. Chong, A. B. J. Teoh, and D. C. L. Ngo, "Iris Authentication Using Privatized Advanced Correlation Filter," in ICB, pages 382–388, 2006

P.Balakumar received the B.E. and M.E. degrees in Computer Science and Engineering from PSG College of Technology, Coimbatore, in 1997 and Anna University, Chennai in 2004 respectively. During 1999-2001, he worked as Lecturer in PSG College of Technology in Coimbatore. Later during 2003-2008, he worked as Lecturer & Assistant Professor in AMS Engineering College, Namakkal. He now with Selvam College of Technology, Namakkal, Tamilnadu, India as Associate Professor in Department of Computer Science and Engineering.

Dr.R. Venkatesan was born in Tamilnadu, India, in 1958. He received his B.E (Hons) degree from Madras University in 1980. He completed his Masters degree in Industrial Engineering from Madras University in 1982. He obtained his second Masters degree MS in Computer and Information Science from University of Michigan, USA in 1999. He was awarded with PhD from Anna University, Chennai in 2007. He is currently Professor and Head in the Department of Information Technology PSG College of Technology, Coimbatore, India. His research interests are in Simulation and Modeling, Software Engineering, Algorithm Design, Software Process Management.

Image Transference & Retrieval over SMS

Muhammad Fahad Khan[1], Saira Beg[2] and Fakhra Kashif[3]

[1] **Computer Science Department, Federal Urdu University of Arts, Science and Technology Islamabad, 46000, Pakistan**

[2] **Computer Science Department, COMSATS Institute of Information and Technology Islamabad, 46000, Pakistan**

[3] **Computer Science Department, Federal Urdu University of Arts, Science and Technology Islamabad, 46000, Pakistan**

Abstract

Paper describes the method of sending colorful images and animations through SMS (Short Message Service) over GSM (Global System for Mobile Communications) network. Presented method has two main blocks; first block converts images or animations into characters. And second block applies Huffman coding compression method on those characters and set them as a payload text of SMS. Generally, SMS is text based service and cannot send multimedia contents.

For delivering multimedia contents, two other services are utilized which are; MMS (Multimedia Messaging Service) and EMS (Extended Messaging Service). EMS is an application level extension; can transfer predefined sounds, animations and images etc. But a drawback is that EMS is not supported as widely as SMS. On the other hand, MMS have richer contents than SMS and EMS. But for large files it needs higher network capability like 3G etc. Enabling colored images (raster/vector) and animation features in SMS we develop an application using J2ME platform.

Keywords: *GSM (Global System for Mobile Communications) network, SMS (Short Message Service), Raster and Vector Images, Animations, Huffman Coding compression.*

1. Introduction

GSM network provides different services which can be grouped in three major categories; Tele services (TS), Bearer services (BS) and Supplementary services (SS). Short messaging service belongs to the BS category [1]. Generally, GSM has three types of messaging services; SMS, EMS and MMS [1-3]. SMS is text based service and its size is limited to 140 bytes or 160 characters. And those characters should be alphanumeric or binary non-text only [3]. On the other hand, EMS is an application level extension of SMS. And it can transfer richer contents; predefined sounds, animations, images etc. But a drawback is that EMS is not supported as widely as SMS, and it's all content is present in message header which will be ignored in unsupported mobile phones [4].

For richer contents delivery, MMS service is used. MMS can transfer videos, sounds, pictures, animations etc. MMS size may be up to 1000 bytes, but for delivering such a large size files it needs higher network capability like 3G, Edge. Enabling MMS service may need enhancement in existing network infrastructure, new billing structure, content adaption and mobile capability (must support MMS service) etc [5].

In this paper we present a method which enables transferring of color images and animations through SMS. For this we develop an application which brings this feature in all GSM devices, even those which does not have EMS/MMS/GPRS/EDGE or other 3G, 4G capability. In this method, our main focus is to reduce hardware dependencies (as in case of MMS and EMS) and provide an alternative method of transferring one's emotion and pictures to the receipts. This Paper is divided into following sections. Section 2 described related work section 3 is about proposed methodology, section 4 discusses the results and lastly we conclude the paper.

2. Related Work

[6] discuss the method of transferring Stereo images through SMS over GSM network. In this method, authors convert the stereo images into characters and then set these characters as a payload text of SMS. They also compared the lightening effect on stereoscopic images. Image displaying with cordless phone is presented in [7]. It is a standalone image display device which has telephone and internet capability. [8] presents a method of sending and displaying animation through SMS over GSM network. Major information required for animated SMS are user identification number and animation flag. User identification number is used to identify the receiving party, flag value determined that SMS has animation or

not. If yes than specify content pointer location value which suggests animation location at receiver side and animation type.

[9] is about transferring voice through SMS. They test their method with three different formats; PCM, ULAW and AMR. And results suggested that their method could be improved by using compression method. [10] Discuss the method of transferring voice using SMS. In this, first get the utterance generated by the encoder card which present in a mobile phone, then converts it into non text representation and insert that text into the body of SMS. Reverse procedure is done on received data after receive the SMS at receiver end. [11] discussing the new J2ME RMI package, which makes use of object compression in order to minimize the transmission time. RMI EOP makes use of GZip and PPM for object compression.

3. Proposed Methodology

If user wants to send colored still and animated images, just browse the content through mobile application and press 'send' button. Our application hides the procedure which converts the selected content into SMS. Application has two main components; first component converts the selected content into characters and second component applies Huffman Coding compression and after that it set those characters as a payload text of SMS. First component has following steps;

1. first, it checks the type of selected content and then saves the content into signed ByteArrayOutputStream.
2. In second step, converts that ByteArrayOutputStream into unsigned integer array.
3. Now, convert unsigned integer array into their respective Extended ASCII character. But practically, some ASCII character values cannot send through SMS. Those ASCII character values are 0-31, so add 256 in those values which falling in the range to move them on to the range 256 to 287 and then converts each integer value into their respective Extended ASCII character.

Second component of application has following steps;

1. First, it takes input of first component and applies Huffman Coding compression method.
2. After compression, convert these compressed characters into strings.
3. Lastly, set these strings as payload text of SMS. There is huge possibility that those strings may consume multiple number of messages; so keep the order of SMS, we reserved first three characters which gives 0-999 indexing. Now add index number to all consumed SMS.

When SMS received at receiver side, our application buffer all incoming messages in order with the help of message index number. Now extract data from SMS and apply all above procedures in reverse order. After this an image or animation will displayed in a mobile screen. Fig 1 shows the sending procedure of an image/ animation.

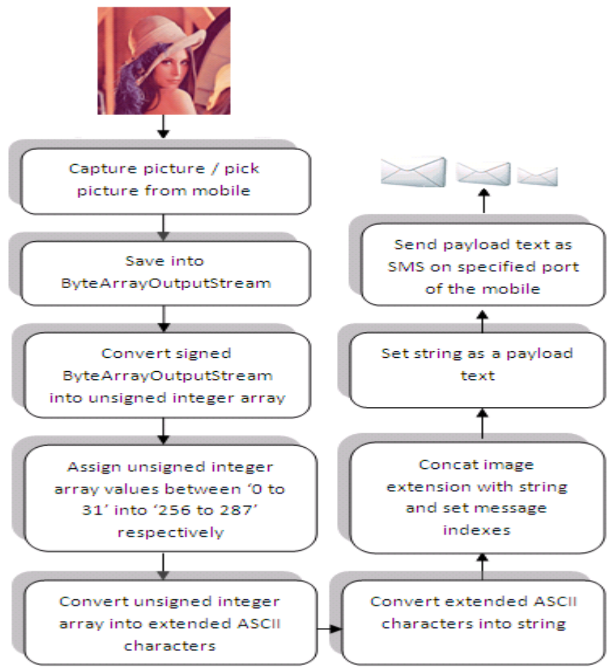


Fig. 1sending procedure of an image/ animation

4. Results, Discussion and Comparison

We developed an application using J2me platform and used Nokia 3110c mobile phones for testing. In this we consider 3 main factors, number of characters, number of connected messages (concatenated message) and unique colors. As SMS has limited data size, solution of this limitation is concatenated SMS [12]. Generally, there are two major image formats; Raster and Vector. For experiment, we used both image formats. Fig 2 shows the test structure which we conducted.

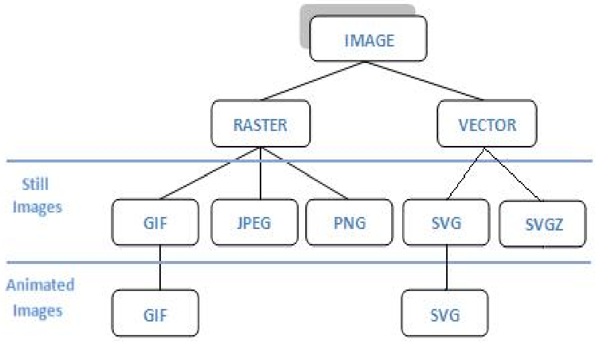


Fig. 2 Test hierarchies

Raster images are consists of pixels. Each pixel contains value which represents the brightness of image at any specific point. Two dimensional array of integer is used to store raster image, known as raster map. One drawback of raster image is that, they cannot scale up very well because it's directly effects the image quality. In contrast, vector images are consists of lines and curves which are points having direction and length. Vector images have small size than raster images because they does not keep track of small pixels as raster images do and they can easily scale up without compromising the quality [13]. In case of raster file format, we selected the three formats which are GIF, JPEG and PNG and for vector [15] we selected SVG and SVGZ formats.

4.1 Images Results

During experiment, for Raster format we keep track of four elements; image resolution, number of supported/unique colors, number of characters and number of connected messages. On other hand for vector images we focused on number of characters and number of connected messages. In case of raster, we select Lena image in JPEG, GIF and PNG formats respectively. Such images are taken from photo database [14].

We divided the experiments in two cases; with "WHCoding" and without "WOHCoding" compression. Fig 3 a, b and c represents the comparison of with and without compression. Fig 3 'a' discusses the comparison of raster images in terms of characters. According to results JPEG format generates small number of characters. On the other hand PNG has highest number of characters.

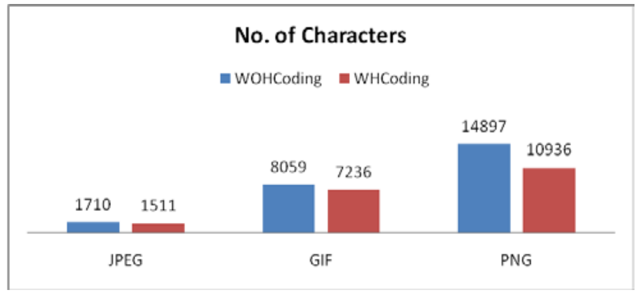


Fig 3a: Comparison of JPEG, GIF and PING in terms of Characters

Fig 3 b and c discusses the compression of raster images in terms of unique colors and connected messages. Connected messages and characters are directly proportional to each other.

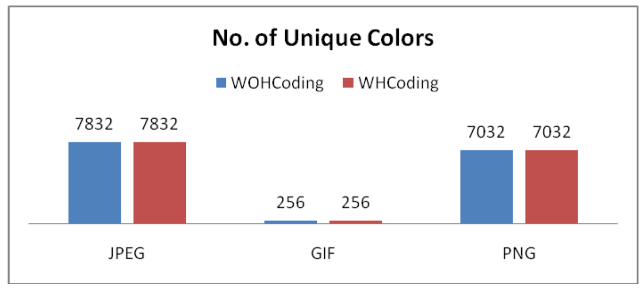


Fig 3b: Comparison of JPEG, GIF and PING in terms of Unique Colors

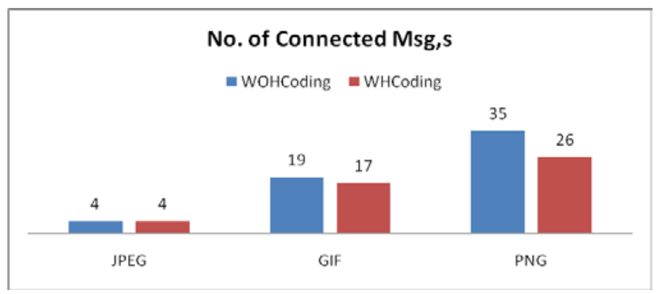


Fig 3c: Comparison of JPEG, GIF and PING in terms of Connected Messages

Fig 4 a and b shows the comparison of both vector formats in terms of connected messages and number of characters, while number of unique colors is same for both formats. Results suggested that SVGZ takes smaller number of messages than SVG; because SVGZ is compressed SVG with most general compression technique.

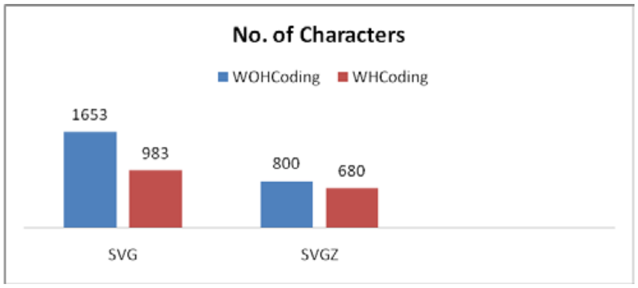


Fig 4a: Comparison of SVG and SVGZ in terms of characters

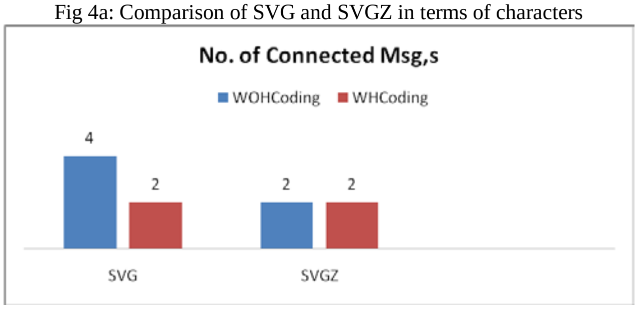


Fig 4b: Comparison of SVG and SVGZ in terms of connected messages

4.2 Animation Results

In second phase, we test our application with aniamtions. For this purpose, we select two aniamtions in SVG and GIF format. Table 1 shows the selected aniamtions. One important note: "GIF format is not supported by Microsoft Word, for this we convert such images in SWF format in order to show those images, similarly, SVG format for animation we convert it into SWF respectively".

Table 1: Selected animations

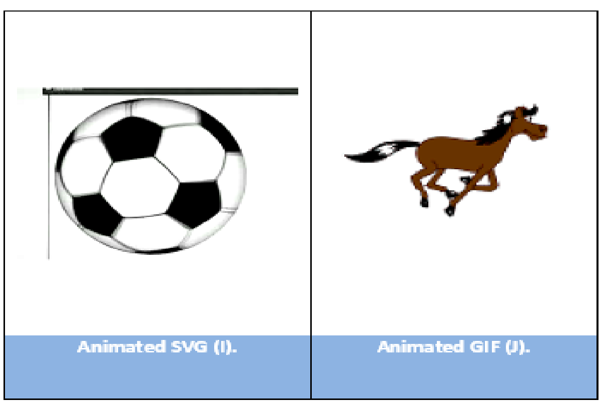


Table 2 shows the results of animations. SVG based animation generates smaller number of characters and messages, while GIF based animation produce larger characters and messages. In case of compression, SVG shows better performance than GIF. Only 4 connected messages

Table 2: Results of each animation without compression

Results of Animated Raster & Vector Without Compression			Results of Animated Raster & Vector With Compression	
Source Images	No. Of Characters	No. Of Messages	No. Of Characters	No. Of Messages
Animated SVG (I)	2927	7	1723	4
Animated GIF (J)	5165	12	4520	11

5. Conclusion

This paper presents a method which can transfer color images and animations through SMS. In this framework, our main focus is to reduce hardware and service dependencies, using existing GSM-SMS architecture. In other words, it is the only way of transferring images in absence of other services like, EMS, MMS, GPRS, and EDGE etc.

References

[1] http://e-articles.info/e/a/title/Features-and-Services-of-GSM/
[2] http://www.mobile-phones-uk.org.uk/mms.htm
[3] http://e-articles.info/e/a/title/The-Main-Protocols-used-by-Mobile-Phones-(SMS-EMS-MMS-WAP)/

[4] http://www.developershome.com/sms/smsIntro.asp
[5] Gordon A. Gow and Richard K. Smith, "Mobile and Wireless communication ", pp no:66, section 5.3.1, Published in 2006, Online available at: http://books.google.com.pk/books?id=RptTSXPgRekC&pg=PA66&dq=challenges+of+Multimedia+MEssaging+Service&hl=en&ei=6Jr1Tc2LI4HtrQfk0_XzDA&sa=X&oi=book_result&ct=result&resnum=6&ved=0CEkQ6AEwBQ#v=onepage&q=challenges%20of%20Multimedia%20MEssaging%20Service&f=false
[6] M. Fahad Khan and Saira Beg, "Stereo image Transference & Retrieval over SMS", Journal of Computing (JoC), pp 20-22, Volume 3, Issue 7, July 2011. ISSN-2151-9617
[7] LI et al. "image display with cordless phone", United States Patent Application, Publication Number US 2009/0128502 AI, May 21, 2009.
[8] Camp Jr, William O "Device and method for providing and displaying animated SMS messages", United States Patent Application, Publication Number 7756536 B2, Primary Class 455/456, (Chapel Hill, NC, US), 2010. Online Available: http://www.freepatentsonline.com/pdfb/documents/uspt/patent_pdf/7756/US7756536/pdf/US7756536.pdf
[9] M. Fahad Khan and Saira Beg, "Transferring Voice using SMS over GSM Network", Journal of Computing (JoC), pp 50-53, Volume 3, Issue 4, April 2011. ISSN-2151-9617
[10] Daniel L. ROTH "Voice over Short Message Service", United States Patent Application, Publication Number US2009/0017849 AI, and Origin: BOSTON, MA US, IPC8 Class: AH04Q720FI, USPC Class: 455/466, 2009. Online Available: http://www.freepatentsonline.com/20090017849.pdf

[11] Jalal Kawash, Ahmad El-Halabi and Ghassan Samara, "Utilizing Object Compression for Better J2ME Remote Method Invocation in 2.5G Networks", Journal of Computing and Information Technology pp 255–264, - CIT 14, 2006, 3.
[12] Sun Microsystems, Inc"Wireless Messaging API (WMA) for Java™ 2 Micro Edition Reference Implementation", Version 1.0, JSR 120 Expert Groups, 2002.
[13] http://www.signindustry.com/computers/articles/2004-11-30-DASvector_v_raster.php3
[14] http://www.petitcolas.net/fabien/watermarking/image_database/
[15] http://www.svgopen.org/2010/papers/3-Compressing_SVG_with_EXI/index.html

MR. Muhammad Fahad Khan is working as a Lecturer and In charge of Management Information System in Federal Urdu University of Arts, Science and Technology, Islamabad. He is interested in Mobile Application and Information System Development, Information Security and Embedded Systems. He did his Bachelor Degree from Federal Urdu University of Arts, Science and Technology in 2008 and now pursuing his Master Degree from IQRA University, Islamabad Campus.

MS. Saira Beg is working as a Research Associate (May 2009- up to date) in COMSATS Institute of Information Technology, Islamabad. Her Interest areas are Networks, Network Security and Artificial Intelligence and other their related Fields. She is the member of Artificial Intelligence Group at CIIT, Islamabad. She did her Bachelor Degree (Gold medalist) from Federal Urdu University of Arts, Science and Technology, Islamabad, Pakistan in 2008 and now doing her Master Degree from COMSATS Institute of Information Technology, Islamabad.
.

MS. Fakhra Kashif is working as Assistant Professor in Federal Urdu University of Arts, Science and Technology, Islamabad. Her Interest areas are Data Mining and RDBMS. She did her Bachelor Degree and MS Degree from International Islamic University Islamabad, Pakistan.

Object-Net Approach for Data Extraction and Reporting

M.Munusamy[1], G.Tholkappia Arasu[2]

**[1] Anna University-Coimbatore,
Tamil Nadu, India.**

**[2] Jayam College of Engineering and Technology,
Anna University-Coimbatore, Tamil Nadu, India.**

Abstract

This paper attempts to present an Object-Net method for understanding the meaning of a natural language (i.e. English), and generate the SQL query based on underrating, and produce the data report as per the requirement understanding. It is proposed to model the elementary meanings which assist the machine to autonomously undertake the analysis and synthesis processes of meaning. In the proposed methodology disambiguation process is performed in context manner: starting from natural text, the context of the sentence is identified, then the actual meaning is identified using correlation of elementary object meanings exist in Object-Net database, It is because even ambiguous word will have only one meaning based on the context or object or domain on which the sentence is written. The data warehouse (DW) is a database. The data stored in the warehouse is uploaded from the operational systems. Data warehouse provides the information required by the decision makers. Business intelligence (BI) mainly refers to computer-based techniques used in identifying, extracting, and analyzing business data, such as sales revenue by products and/or departments, or by associated costs and incomes. BI technologies provide historical, current and predictive views of business operations using the data from data warehouse (DW). The cost of building a data warehouse (DW) & business intelligence (BI) is expensive for any organization as it requires data warehouse tools for building data warehouse and extracting data using data mining tools from data warehouse. The proposed method called Object-Net uses the English language for getting the requirement for business intelligence reporting and identifies meaning of the sentence, internally creates the interface layer, generates query, gets the data and reports for analyzing the data.

Keywords: *Context, Database, Domain, NLP- Natural language processing, Object, Parse tree, WSD- Word sense disambiguation, DW- Data Warehouse, BI- Business Intelligence.*

1. Introduction

Word sense disambiguation (WSD) is the process of identifying which sense of a meaning is used in any given sentence, when the word has a number of distinct senses [4]. For a long time the WSD is an open problem in natural language processing (NLP). The solution of this problem impacts other tasks such as discourse, engines, anaphora resolution, coherence, inference, information retrieval, machine translation and others. This paper attempts to present an object net method for data extraction from database and reporting using natural English language. It is proposed to model to autonomously undertake the analysis and synthesis processes of meaning the elementary meanings of English sentence and mapping it as per the actual database structure from which user requires extracting the data, and writing the SQL query for the input English sentence or the requirement which is written in regular English. It is different from existing approach on database query language like SQL, PLSQL, Oracle, Sybase, because any of the database engine can able understand only its own query language but if we try to enter our requirement interims of English sentence the database engine will not process our sentence and it will give error message. The proposed here an algorithm understands the English language and maps the equivalent corresponding database query language and in such a way that it can be used to interact with the database engine with its language to retrieve the data and then report it to user.

2. Word sense disambiguation (WSD) approaches, Data Warehouse (DW) and Business Intelligence (BI)

There are two main types of approach for WSD in natural language processing called as deep approaches and shallow approaches.

Deep approaches: these approaches involve the intention to understand and create meaning from what is being learned, Interact vigorously with the content, Make use of evidence, inquiry and evaluation, Take a broad view and relate ideas to one another, and Relate concepts to every time experience [3], [6], [9]. These approaches are not very successful in practice, mainly because such a body of knowledge does not exist in a computer-readable format, outside of very limited domains. There is a long tradition in computational linguistics, of trying such approaches in terms of coded knowledge and in some cases; it is hard to say clearly whether the knowledge involved is linguistic or world knowledge. The first attempt was that by Margaret Masterman, at the Cambridge Language Research Unit in England, in the 1950s, and Yarowsky's machine learning optimization of a thesaurus method in the 1990s.

Shallow approaches: These approaches are not concerned of learning the text instead they deal with the surrounding words of the ambiguous word and try to identify only parts of interest for a particular application. They just consider the surrounding words, using a training corpus of words tagged with their word senses the rules can be automatically derived by the computer. This approach, while theoretically not as powerful as deep approaches, gives superior results in practice, due to the computer's limited word knowledge.

In addition to deep approaches and shallow approaches, there are four conventional approaches to WSD:

Dictionary and knowledge-based methods: These approaches make use of dictionaries, thesauri, and lexical knowledge bases, without using any corpus evidence.

Supervised methods: These approaches make use of sense-annotated corpora already been trained from semantically disambiguated corpus.

Semi-supervised or minimally-supervised methods: These approaches make use of both labelled and unlabeled data for training - typically a small amount of labelled data with a large amount of unlabeled data [10].

Unsupervised methods: These eschew (almost) completely external information and work directly from raw corpora (i.e. not annotated).

Data Warehouse: A data warehouse is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision making process. A data warehouse maintains its functions in three layers: staging, integration, and access. Staging is used to store raw data for use by developers. The integration layer is used to integrate data and to have a level of abstraction from users. The access layer is for getting data out for users. The data from data warehouse is used for many purposes like, business professionals for data mining, online analytical processing, market research and decision support. However, the means to retrieve and analyze data, to extract, transform and load data, and to manage the data dictionary are also considered essential components of a data warehousing system.

Business Intelligence: It is usually refers to make decisions based on the information that is available in data warehouse. A data warehousing (or data mart) system is the backend, or the infrastructural, component for achieving business intelligence. Business intelligence also includes the insight gained from doing data mining analysis, as well as unstructured data. Business intelligence aims to support better business decision-making. Thus a BI system can be called a decision support system (DSS). Though the term business intelligence is sometimes used as a synonym for competitive intelligence, because they both support decision making, BI uses technologies, processes, and applications to analyze mostly internal, structured data and business processes while competitive intelligence gathers, analyzes and disseminates information with a topical focus on company competitors. Business intelligence is a subset of competitive intelligence.

The method proposed here is a semi-supervised method; it is called as object - net approach which uses the information dynamically gathered from user that is while machine finds any of untrained corpora or unable to solve the disambiguation then those information are reported to user or master, after user understand the problem the related corpora are trained [7]. It differs from previous semi-supervised approaches: the algorithm has a set of disambiguated trained elementary objects, and incrementally builds and resolves the untrained elementary objects. This algorithm can be incorporate into lager applications like machine translation, code generation, search engine, IR, etc.

Resources: The algorithm does not dependant on any other existing WSD resources like WordNet, SemCor, and any BI tools like SAP Business Object, . Instead of that it uses separate database named as Object-Net Database which contains trained elementary objects. Initially the database is stored with limited data, this database updated when new untrained object found in the input text or when fine tuning is required on existing already trained element. The proposed algorithm finds all its required information to identify the meaning of the word on a particular context from this Object-Net database, so precision of word sense disambiguation of proposed algorithm mainly depends on data from this special Object-Net database.

3. Object - Net Approach for Data Extraction and Reporting

3.1 Object –Net Approach Procedures: The algorithm presented in this paper determines, in a given text, a set of nouns and verbs which can be disambiguated with high precision, the semantic tagging is performed using the sense defined in Object-Net Database, and actual meaning of the sentence is identified. But above mentioned task are completed in step by step using methods, so the various methods used to identify the correct sense of a word are presented first, Next presents Object-Net Database architecture, the main algorithm in which these procedures are invoked in an iterative manner, and the method of updating, fine tuning the Object-Net Database.

PROCEDURE 1: This procedure tokenizes the given sentence and creates a parse tree path for the given sentence. Parse tree paths were used for semantic role labelling. Predicates are typically assumed to be specific target words (verbs), and arguments are assumed to be spans of words in the sentence that are dominated by nodes in the parse tree. A parse tree path can be described as a sequence of transitions up from the target word then down to the node that dominates the argument span. The parse tree paths are particularly interesting for automated semantic role labelling because they generalize well across syntactically similar sentences. For example, the parse tree path in fig 1 would still correctly identify the "taker" argument in the given sentence if the personal pronoun "she" were swapped with a markedly different noun phrase.

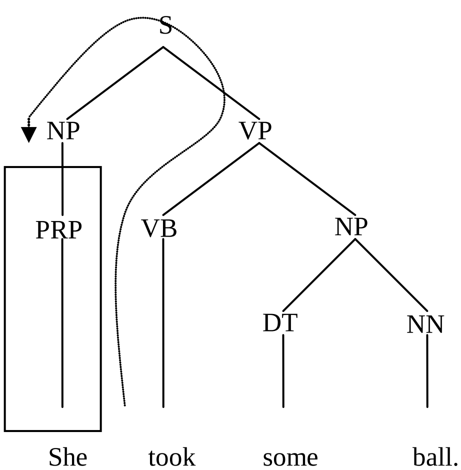


Fig 1: An example parse tree path from the predicate "took" to the argument "She", represented as ↑VB↑VP↑S↓NP

PROCEDURE 2: Identify the words having only one sense(monosemous words) in Object-Net database and make them as having number of sense as #1.
Example: the noun subcommittee has one sense defined Object-Net database. So this is a monosemous word and marked as having sense #1.
PROCEDURE 3: with this procedure, we are trying to get contextual clues regarding the usage of the sense of a word. For a given word Wi, at position i in the text, form two pairs, one with the word before Wi and the other one with the word after word Wi. Then we find out all the occurrences of these pairs found within the Object-Net database. If, in all the occurrences, the word Wi has only one sense as # Wis, then mark the word Wi as having sense #Wis.
PROCEDURE 4: Find the words which are semantically connected to the already disambiguated words for which the connection distance is 0. The semantic distance is computed based on the ObjectNet hierarchy. Two words semantically connected at a distance of zero if they belong to same path of subnet.
PROCEDURE 5: Find words which are semantically connected [2] in ObjectNet and for which the connection distance length is zero. In this procedure none of the words considered by this procedure already disambiguated. We have to consider all the sense of both words in order to determine whether or not the distance between them is zero, this makes this procedure computationally intensive.
PROCEDURE 6: Form the semantic network [8] based on understanding made by the learning done from procedure #1 to procedure #5 and come to the final conclusion about the input sentence and action to be performed. The procedures presented above are applied iterative; this allows us to identify a set of nouns and verbs which can be disambiguated with high precision.
PROCEDURE 7: Using the procedures from PROCEDURE 1 to PROCEDURE 6, the system identifies the meaning of given input sentence. After identifying the actual meaning sentence (i.e. the action to be performed by the system) the system generates the SQL query for given input English text by correlating understudied meaning with actual database field which is exist in user requested database.
PROCEDURE 8: Using the procedures from PROCEDURE 1 to PROCEDURE 7, the system identifies the meaning of given input sentence and extracts the required data using SQL query from database. After the data extraction the query result data is projected in the report format as request in the user input English text.

The procedures presented above are applied iterative; this allows us to identify a set of nouns and verbs which can be disambiguated with high precision.

3.2 Object-Net Database Architecture

The existing knowledge bases in machine readable formats are WordNet, OMCSNet, MindNet, CYC, Thought treasure, VerbNet, Semcor, Open Mind Word Expert, Frame Net, and PropBank. These knowledge bases are useful to serve the purpose of developing information retrieval systems and shallow semantic representation for an input text. They model their elementary meanings only with conceptual world properties and constraints, and taxonomic relations between these words. They do not have synthesis capabilities, but rather their definitions are pre-programmed by humans. They do not make the machine creative enough to master its own language and to compose its own text based on its understood meanings. So a new methodology is required for machine to autonomously undertake the learning, analysis and of both the elementary and composite meanings of natural language, and most importantly, it is to note that the robustness of proposed algorithm by machine relies not only on sophisticated algorithms for knowledge manipulation but also the kind of knowledge it has. (i.e. careful modelling of elementary meanings from an engineering point of view). The new methodology for maintaining trained elementary meaning is called Object-Net database and details of this database is explained in analytical and synthesis capability section.

3.3 Algorithm with an Example

Consider for example to retrieve data from any of user database like "*I need the student report that joined on 04 November 2010 and report the result in the format of student name and course.*"

Procedure#1:
Tokenize the given sentence as below
"*I + need + the + student + report + that + joined + on + 04 + November + 2010 + and + report + the + result + in + the + format + of + student + name + and + course.*"
While categorizing these token words the below result is found.
"*Pro+Ver+Art+Nou+ver+pro+ver+adv+Num+Nov+Num+Con+Ver+Art+Ver+Pre+Art+Nov+Pre+Nov+Nov+Con+Nov*"
Creates the parse tree after tokenizing the sentence.

Procedure#2:
Find the words which are having unique sense of meaning and find object on which the action need to be performed.

"***I*** *(Sense#1) +* ***need*** *(Sense#1) +* ***the*** *(Sense#1) + student + report +* ***that*** *(Sense#1) + joined + (***on*** *+ (***04 + November + 2010***)) (Sense#1) + and + report + the + result +* ***in*** *(Sense#1) +* ***the*** *(Sense#1) +* ***format*** *(Sense#1) + of + student +* ***name*** *(Sense#1) + and + course.*"
In this example the word "I", "the", "name", "that" and "date (04 November 2010)" are having only one sense of meaning, and student is the object on which the sentence related.

Procedur#3:

As per procedur#2 result, the related object or domain of sentence identified (i.e. as per example student), in Object-Net database search for the particular domain which is identified in procedur#2, from the identified object co-relate and identify meaning of the remaining words in sentence. Consider the network exist in object-net database as below fig 2.

While forming the two pairs one with the word previous to the current word and one next to the current word, for our example we will be arrived to the pairs as in below table 1, the last column shows that understanding.

Table 1: parsed tokens and its relation.

S.No	Pairs	Description
1	I + need	Whom->I
2	need + the student	What -> the student
3	The student + report	What -> Report
4	Report + that	Unable to correlate
5	That + joined	Which-> joined
6	Joined + on	Which -> on
7	On + 04 November 2010	Which -> date
8	And +report	What-> report
9	The + result	Which->result
10	Result + in	How -> in
11	In + the	Which -> the
12	The + format	What -> format
13	Format + of	How -> of
14	Of + student	What -> student
15	Student + name	What -> name
16	Name + and	Unable to correlate
17	And + course	What-> course

Procedure#4:
From the procedure#3 we come to know that "need" is the action it required for "whom" is "I", "what" required is "student". From the student node "what" required is "report".
But "report" is ambiguous word in English it is having many meaning, and also by directly correlating words existing object-net is not giving correct path for the pairs "report + that" and "Name + and", as Date is already disambiguated and while considering pervious nodes it gives the meaning like "on" which is some date (ie. 04 November 2010).

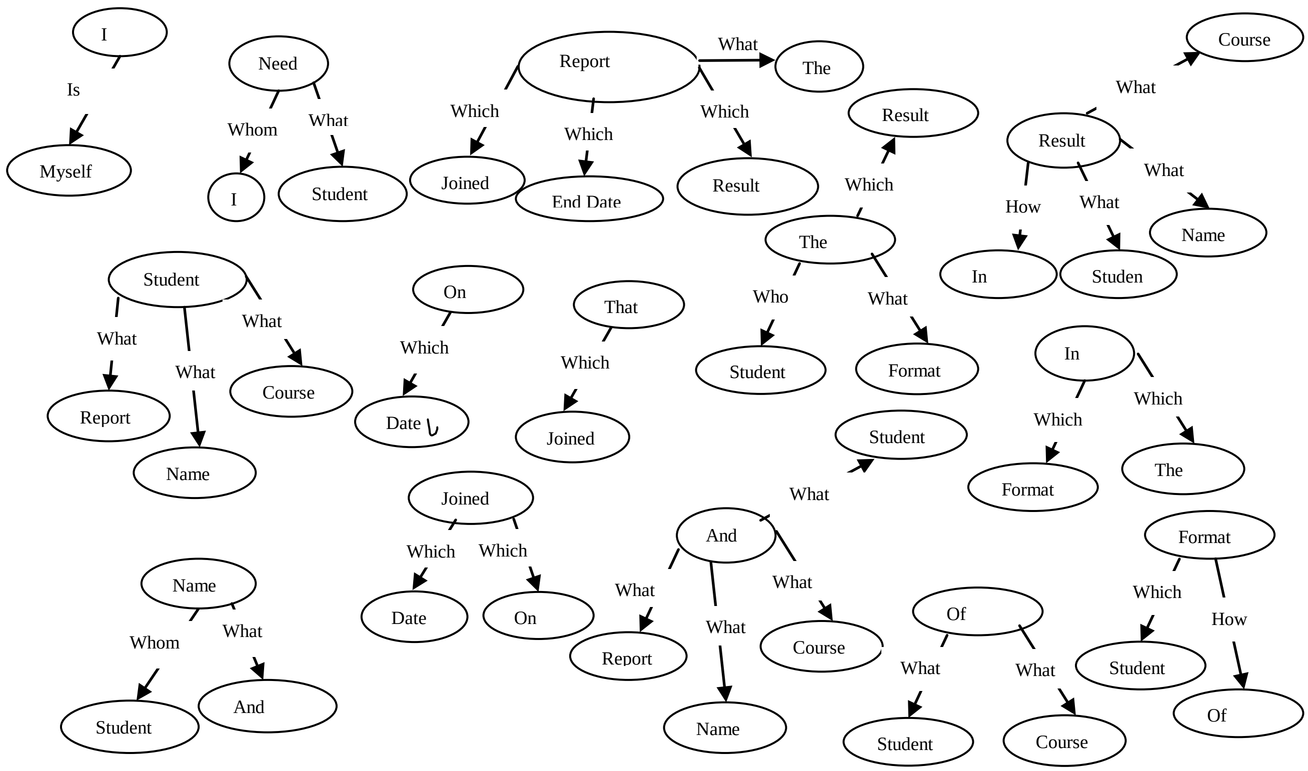


Fig. 2 Sample Object-Net Database

By node with connection distance of zero we will be arrived into the below mentioned paths.

1. I -> need
   a. I -> need -> the student
   b. I -> need-> the student -> report
2. That
   a. That -> joined
   b. That -> joined -> On
   c. That -> joined -> On -> 04 November 2010
   d. That -> joined -> On -> 04 November 2010 -> And
   e. That -> joined -> On -> 04 November 2010 -> And -> report
   f. That -> joined -> On -> 04 November 2010 -> And -> report -> the
   g. That -> joined -> On -> 04 November 2010 -> And -> report-> the -> result
   h. That -> joined -> On -> 04 November 2010 -> And -> report-> the -> result ->in
   i. That -> joined -> On -> 04 November 2010 -> And -> report-> the -> result ->in->the
   j. That -> joined -> On -> 04 November 2010 -> And -> report-> the -> result ->in->the -> format
   k. That -> joined -> On -> 04 November 2010 -> And -> report-> the -> result ->in->the -> format ->of
   l. That -> joined -> On -> 04 November 2010 -> And -> report-> the -> result ->in->the -> format ->of -> student
   m. That -> joined -> On -> 04 November 2010 -> And -> report-> the -> result ->in->the -> format ->of -> student -> name
3. And-> course

Procedur#5:
The word “report” was not clear still Procedur#4, now the report is clear like on “join date” some report is required. The ambiguous word “report” semantically connected with other part of the sentence in three ways as mentioned below.

1. Report -> joined -> 04 November 2010.
2. Report -> joined -> on -> 04 November 2010.
3. Report -> That -> joined -> On -> 04 November 2010 -> And -> report-> the -> result ->in->the -> format ->of -> student -> name.

Here the path 2& 3 are already occurred in Procedur#4 but path 3 is bigger than path 2, so this path is considered, and now it is clear that report of joined date is required.

Procedur#6:
From the procedure#5, the "need" node is connected to "student" node. "student" node is connected to "report", "report" is connected to "joined date" and it is connected to "date", from this we have form a semantic network which gives the meaning as "need" is the action required by "I" and what required is "student", from "student" what required are report, and which report is "join date" report and similarly forms the semantic network.

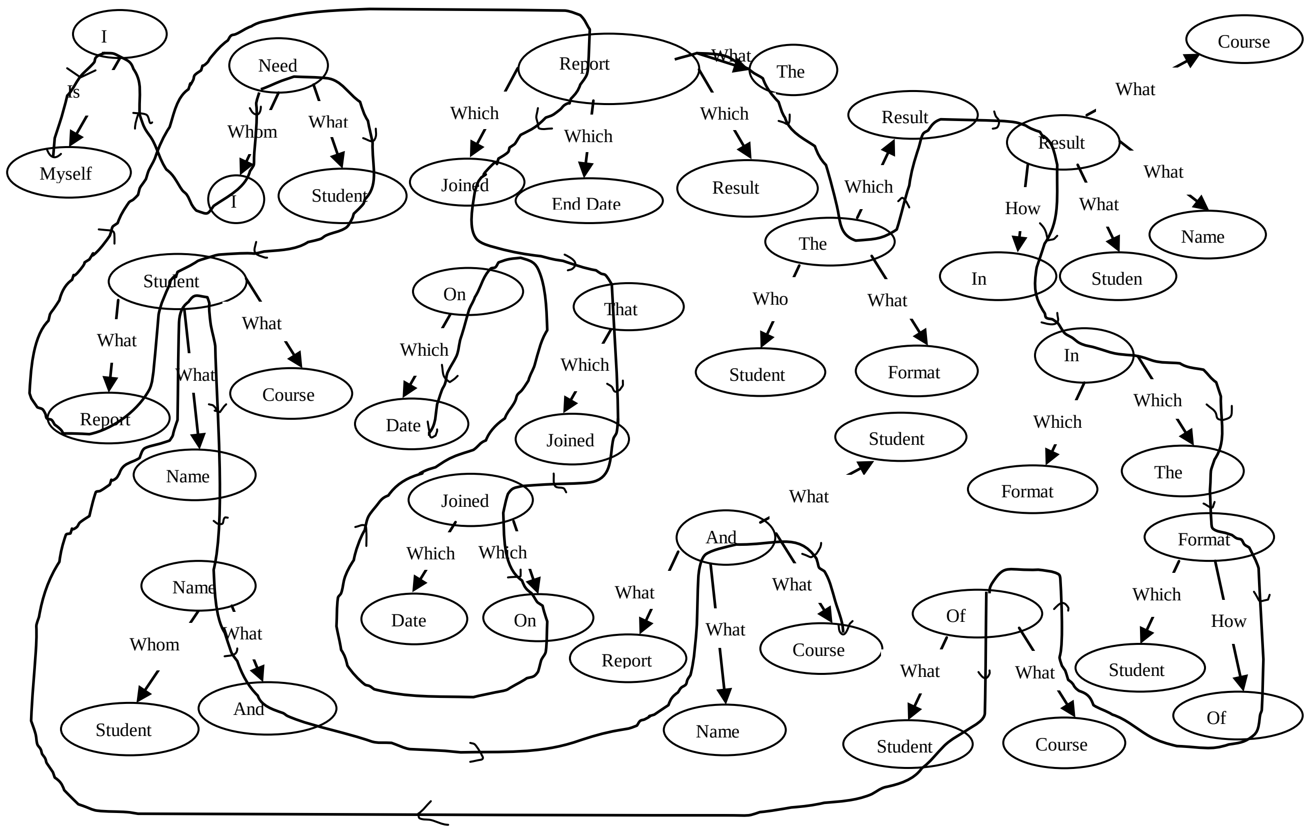


Fig. 3 Forming a semantic network in Object-Net Database

The above fig 3 shows the semantic relation path which gives meaning of the sentence.

Procedur#7:
After identifying the meaning and semantic relation path, system identifies that "Need" is the action and what need is "student" which is also exist in mapping list of user database, now identified the sub-user database as "Student details "but it contains two table as "Personal _details" and "mark details" at the same time from input text it identifies "joined date" which is existing in "Personal _details" so now it identifies table name "Personal _details" and writes the SQL statement as "From Personal _details" , the word "Report" and "student" says that selection of "name" and "course" is required so the system writes the SQL statement as "Select name, course", from the word "Report" it also identifies the SQL where class constraint as report of joined date is required but in input text contains some constant date value of "04 November 2010" so it writs the SQL statement as "Where Joined_date="04 November 2010"". While it recompiles all spices of

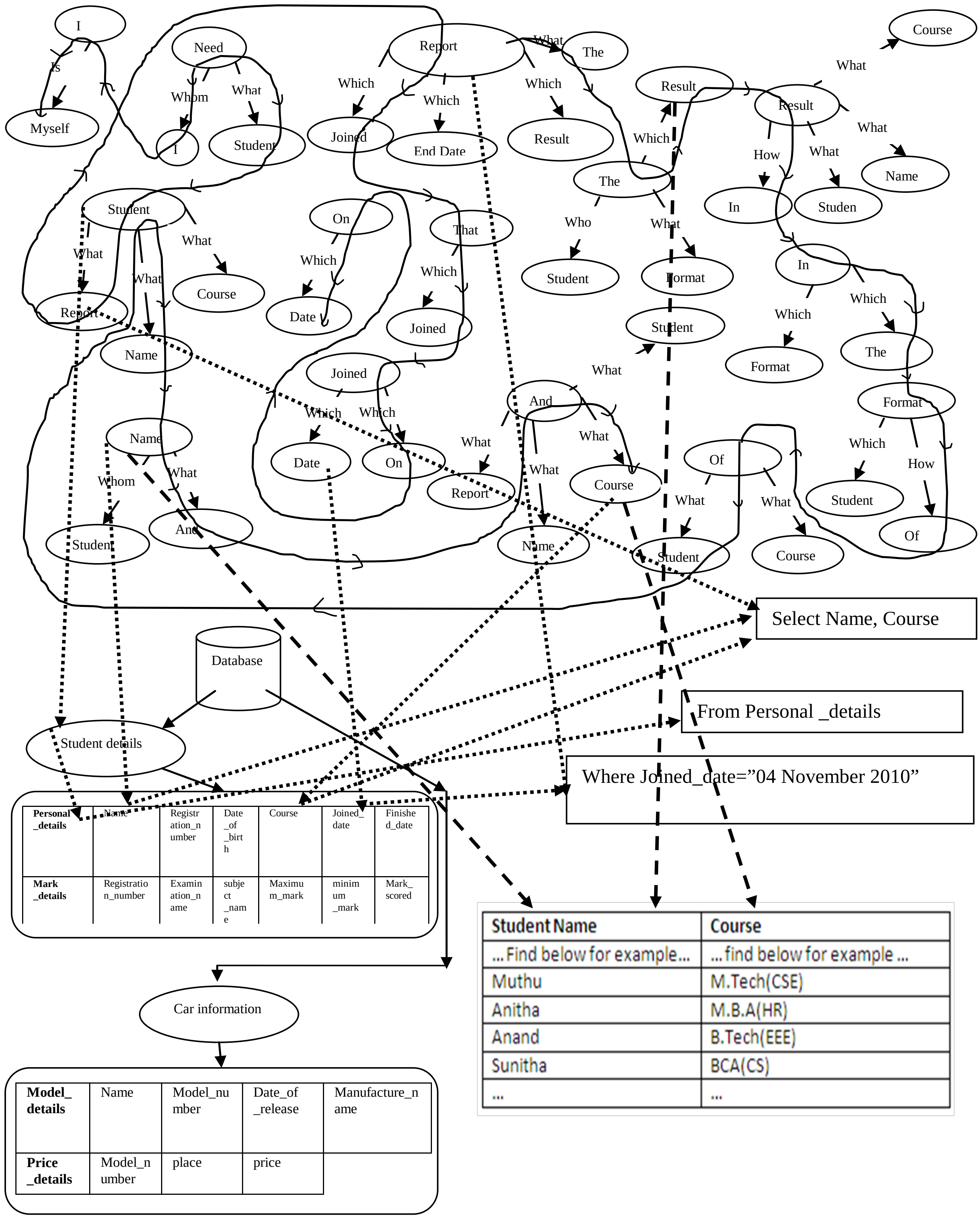


Fig. 4 Generating SQL statement from the semantic network in Object-Net Database

SQL statement which are generated as above procedures. The fig 4 shows the shows the semantic relation path along with SQL generation path. We get the complete SQL query as "*Select name, course From Personal _details Where Joined_date="04 November 2010"*. This SQL statement is executed by database engine where the user requested details are exist and gives result back to user.

Procedur#8:
The PROCEDURE 7 generates the SQL query based on the semantic network for reporting. From semantic network the system identifies the meaning of other part of the given sentence, from the "report" it understood that the "result" needs to be in "format" of "student name" and the "course". The result which came out after executing the SQL query is restricted to display only the student name and course as shown in the fig 4.

4. Analytical and Synthesis Capability In Object-Net Database

The example sentence "I need the student report that joined on 04 November 2010 and report the result in the format of student name and course" can be written in many as like mentioned below to reference same meaning as above sentence says. The possible ways are:

1. I need the student report that joined on 04 November 2010 and report the result in the format of student name and course
2. Need student report joined on 04 November 2010 and report the result in the format of student name and course
3. Report of student joined on 04 November 2010 in the format of student name and course
4. Student report joined on 04 November 2010 in the format of student name and course
5. On 04 November 2010 joined student report in the format of student name and course

The above mentioned sentences are giving same meaning as sentence#1, even though the sentences are not in corrected grammatical. But as a human can understand that meaning of all above sentence as "student name and course report is required who are all joined on 04 November 2010". So similarly we have to make sure that our proposed algorithm is also capable understanding the meaning of sentence as human.
For example the above sentence # 3 "Report of student joined on 04 November 2010 in the format of student name and course", in existing trained Object-Net network does not have direct relation from report and student but already the "what" relation were existing so it makes the new understanding link between "Report" and "student" with relation of "what". Similarly consider the above sentence#5 "On 04 November 2010 joined student report", this sentence starts with a date and it does not have action part like a action verb "need", in existing Object-Net doesn't have any of node starts with "Date" but there is a "Which relationship exists between "Joined" and "Date" so system creates a new node as "Date" to "Joined" with relation of "Which", next for student report there are two relationship exist one is from "Report" and another one from "student" node, now it creates two relation from newly created date "Date" node to "Student" and "Report" with relation of "Where" and "What" respectively. The fig 5 show the updated Object-Net database which will be used for future purpose.

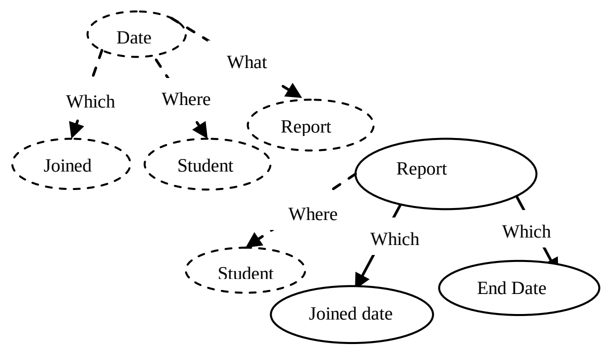


Fig. 5 updating active memory

So the system analyses and keeps updating its database memory there comes the system learning capability. If some words occurred in input text which is not exist in Object-Net database and also system is not able to resolve it internally then it will ask a master to train the relational network there come the human master into picture in order to correct and update the database.

5. Object-Net Approach for Database Extraction

We illustrate here the Object-Net disambiguation algorithm with the help of previous example "*I need the student report that joined on 04 November 2010 and report the result in the format of student name and course*". The system identifies the data meaning of the sentence, and what is the command, and what is action that user is expecting from the system. After identifying the meaning of the sentence, it maps the action to be done along with the trained internal actual database structure so that it can produce exact the SQL query for the input sentence or requirement. The bellow fig 6 shows that "student details" and "car information" databases are exist in a database; this mapping information is shared or trained to our system so that our system knows about where to fetch and which are to be fetched for a given sentence.

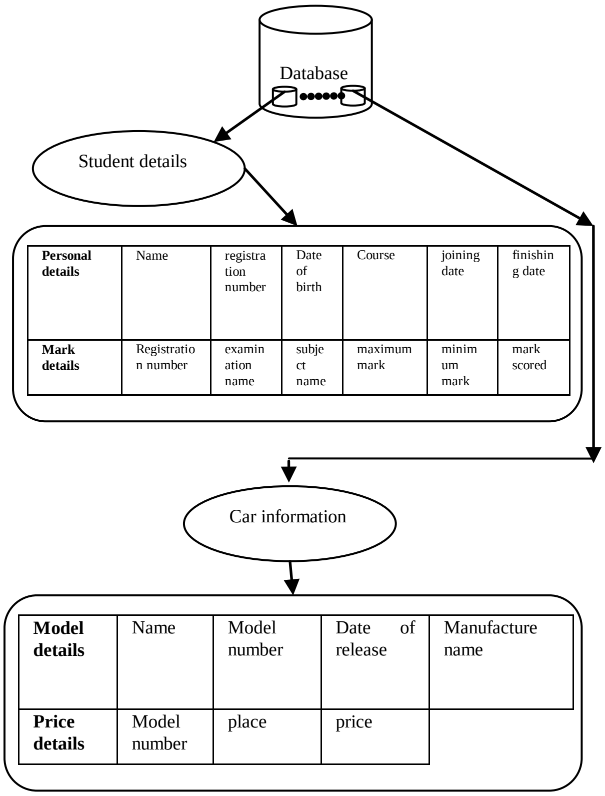


Fig. 6 Actual database information for mapping

As the system is capable of processing the English sentence, identify the meaning and writing the SQL query for the input sentence. In our above example instead of student details if the user asks for car details such as “I need the car details that are released on 01 MAY 2009” then the system will process this sentence and as above mentioned algorithm and it maps and writs SQL query on car details database. So the proposed system can be integrated to any database once the elementary meaning in Object-Net database is trained.

6. Performance of Data Extraction from Database, And Word Sense Disambiguation Based On Object-Net Database

The data extraction from database is the process of analyzing the requirement, and writing the SQL query based on the requirement. In our proposed algorithm the system itself analyzes the user requirement given interim of English sentence and writes the corresponding SQL queries. The performance of accuracy of writing SQL query and extracting the user required data is based on how the system understand the input text. But the system understanding of input sentence is based on the object-net database. The Object-net consists of set of initially trained entity network along with their meaningful representation with their action/behavior/property. The performance of our word sense disambiguation is mainly based on how many trained networks exist in Object-net database. If number of network data are high then number of hit ratio or number of occurrence of word in input text and trained network is high so it helps our algorithm to fetch correct object on which the input sentence is written and what is action or purpose of the sentence in order to give good accuracy on ambiguous words and sentence. When the number of trained network data of words in object-net database is less then number of hit ratio or number of occurrence of word in input text in trained network words is less so the active memory model of object-net database requires the help from master to train the non-trained words into database. So the accuracy the user requested data from user database is based on number trained network in Object-Net database and automatic system mapping between Object-Net database to actual system database entity fields. Fig 7 plots the graph between accuracy of the result of our algorithm versus number trained network word exist in object-net database, and the learning update required of object-net database in active memory model, along with accuracy over data extraction and automatic system mapping.

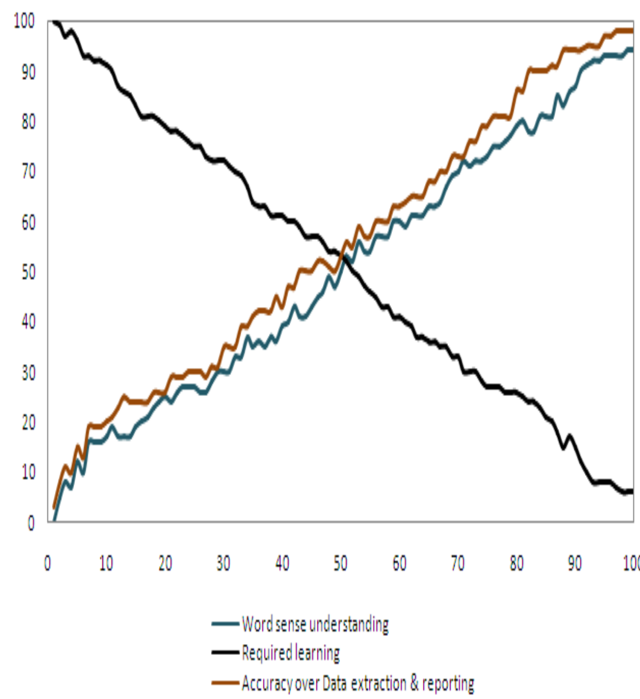


Fig. 7 Accuracy of Data extraction and Reporting Vs Number Trained Network, and automatic system mapping

7. Conclusion and Further Work

The algorithm identifies the meaning of requirement (i.e. user input sentence), analyzes the input sentence, and generates the SQL query statement, and reports the result as per the requested format. As per the above performance analyzes our algorithm extract and reports the data with precision of more than 97%, so it can be used in human computer interaction system, and data warehousing system in place of existing costliest commercial tools like DataStage, Business Object, etc.

8. Conclusion and Further Work

The algorithm identifies the meaning of sentence like human brain. It disambiguates ambiguous words based on object on which sentence is written and it generates the SQL query for data access and reporting. In future we can train our Object-net data base to other object or domains wherever intelligent human-computer interaction is required. And also from understanding of natural text meaning to the actual database query generation process can be implemented for accessing data from user database as per the user requirement.

References

[1] Dan Klein, Kristina Toutanova, H. Tolga Ilhan, Sepandar D. Kamvar, and Christopher D.Manning, 2002 "Combining Heterogeneous Classifiers for Word-Sense Disambiguation." In Workshop on Word Sense Disambiguation: Recent Successes and Future Directions at ACL 40, pages 74-80.

[2] Ezeife, C.I. and Y. Lu, 2005. Mining web log sequential patterns with position coded pre-order linked WAP-tree. Data Min. Knowl. Discov., 10: 5-38. DOI: 10.1007/s10618-005-0248-3

[3] Lee N, Laine A.F, Klein A, 2011.Towards a deep learning approach to brain parcellation, 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Page: 321 - 324, Chicago, IL, ISSN: 1945-7928.

[4] Sanderson, M., 1994. Word sense disambiguation and information retrieval. Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, July 3-6, Springer-Verlag New York, Inc., New York, USA., pp: 142-151. http://portal.acm.org/citation.cfm?id=188548&dl=.

[5] Thomas, M. Connolly, Carolyn E. Begg and Carolyn E. Begg, 2004. Database Systems: A Practical Approach to Design, Implementation and Management. International Computer Science Ser.4th Edn., Pearson, ISBN-13: 9780321210258.

[6] Gordon S. Linoff and Michael J. A. Berry, 2011. Data Mining Techniques for Marketing, Sales and Customer Relationship Management. 3rd Edn., Wiley Computer Publishing, ISBN 0-470-65093-1.

[7] Jie Fan, Pu Wen, Hunan Radio & TV Univ., Changsha, 2007. Application of C4.5 Algorithm in Web-Based Learning Assessment System, IEEE International Conference on Machine Learning and Cybernetics, Issue Date: 19-22 Aug. 2007, On page(s): 4139 - 4143,Hong Kong, DOI: 10.1109/ICMLC.2007.4370871.

[8] Lee, G.N. and H. Fujita, 2007. K-means Clustering for Classifying Unlabelled MRI Data. Proceeding of the Digital Image Computing Techniques and Applications, Dec. 3-5, IEEE Xlpore Press, USA.,pp: 92-98. DOI: 10.1109/DICTA.2007.4426781

[9] Niknafs, A.A., M.E. Shiri and M.M. Javidi, 2005. An intelligent knowledge sharing strategy featuring item-based collaborative filtering and case based reasoning. Proceeding of the Intelligent Systems Design and Applications, Sept. 8-10, IEEE Xlpore Press, USA., pp: 67-72. DOI: 10.1109/ISDA.2005.22

[10] Claudio Giuliano, Alfio Massimiliano Gliozzo, and Carlo Strapparava, 2009. Kernel methods for minimally supervised wsd, ACM on Computational Linguistics, Volume 35 Issue 4, Pages 513-528,MIT Press Cambridge Publisher, MA, USA, ISSN: 0891-2017

[11] M. Barathi, S. Valli, 2010.Ontology Based Query Expansion Using Word Sense Disambiguation, International Journal of Computer Science and Information Security(IJCSIS) on Information Retrieval (cs.IR), Vol. 7, No. 2, pp. 022-027, USA, Computer Science ISSN 19475500.

M.Munusamy
He is a Ph.D research scholar in Department of Computer Science in Anna University- Coimbatore. His area of research is the application of natural language processing in database systems. He has M.Tech degree in Computer Science from SRM University- Chennai. Currently doing research on machine learning in computer science.

G.Tholkappia Arasu
He is having Ph.D on computer science. He is the principal of Jayam college of engineering and technology, Anna University, India. His area of research is the nanotechnology. He has several research publications in journals and presented papers in national and international conferences. Currently he is the chief editor of journal of convergence in engineering technology and science – JCETS.

Implementation of Re-configurable Digital Front End Module of MIMO-OFDM using NCO

Mrs. VEENA M.B.[1] , Dr. M.N.SHANMUKHA SWAMY[2]

[1] *Assistant professor, Vemana I.T. ,Koramangala, Bangalore, Karnataka, India.*

[2] *professor, ECE department, SJCE, Mysore, Karnataka, India*

Abstract

This paper focuses on FPGA implementation of reconfigurable Digital Front end MIMO-OFDM module. The modeling of the MIMO-OFDM system was carried out in MATLAB followed by Verilog HDL implementation. Unlike the conventional OFDM based systems, the Numerically Controlled Oscillators (NCO) is used for mapping modulated data onto the sub carriers. The use of NCO in the MIMO-OFDM system reduces the resource utilization of the design on FPGA along with reduced power consumption. The major modules that were designed, which constitute the digital front end module, are Quadrature Phase Shift Keying (QPSK) modulator/demodulator,16-Quadrature Amplitude Modulation (QAM) modulator/demodulator and NCOs. Each of the modules was tested for their functionality by developing corresponding test benches. In order to achieve real time reconfigurability of the proposed architecture, the proposed approach is realized on FPGAs optimizing area, power and speed. Reconfigurability of the proposed approach is dependent upon user requirement. Hence the proposed approach can support future generation communication technologies that are based on MIMO-OFDM

Keywords: *OFDM, NCO, MIMO, QPSK, QAM, FPGA.*

1. Introduction

The increasing demand on wireless services, both for voice and data communications is a major motivational factor for developing MIMO-OFDM system. In particular the demand for multimedia services such as video-on-demand, downloading music and movies, video conferencing, etc, is expected to diversify services and increase the volume of data traffic. As a result, emerging wireless/mobile networks are those which can integrate voice and data services, opposed to traditional voice-oriented networks. It is necessary to clarify that this system works within a frequency limitation established by the FPGA. This means it will never be able to directly modify the value of the RF carrier frequency. However, a typical RF front-end with only one up-converter will be able to modify the RF carrier frequency to fit the transmission system requirements. The maximum RF that can be fed to the SPARTAN 3 FPGA board is 500 MHz. The number of subcarriers chosen was 64, with the bandwidth of each subcarrier as 7.8 MHz. Cyclic prefix chosen is 25% of the number of subcarriers used. Modulation schemes used are QPSK and 16-QAM. The data rate achieved is 100 Mbps.

1.1 MIMO

To multiply throughput of a radio link, innovative techniques, such as, multiple antennas (and multiple RF chains accordingly) at both the transmitter and the receiver, have been employed. These systems are termed as Multiple Input Multiple Output systems [5]. A MIMO system with similar count of antennas at both the transmitter and the receiver in a point-to-point link is able to multiply the system throughput linearly with every additional antenna. Figure 1.1 illustrates a 2x2 MIMO system which has the ability to double the throughput.

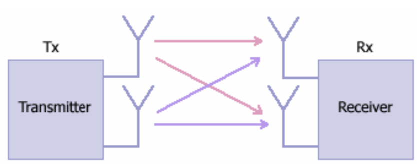


Figure 1.1 Multiple Input Multiple Outputs (MIMO 2x2)

A MIMO system takes advantage of the spatial diversity that is obtained by spatially separated antennas in a dense multi path scattering environment. MIMO systems may be implemented in a number of different ways to obtain either a diversity gain to combat signal fading, or to obtain a capacity gain.STBC is based on orthogonal design and obtains full diversity gain with low decoding complexity (Alamouti code is a special case with double Tx antennas). Space-time block coding (STBC) is a simple yet ingenious transmit diversity technique in MIMO technology.

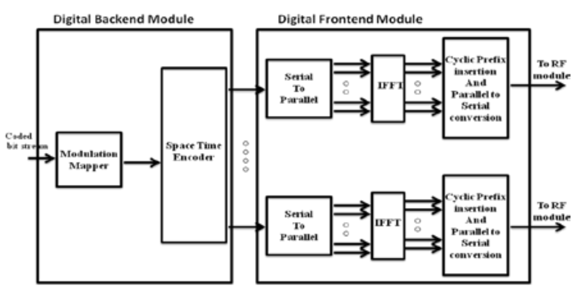
1.2 OFDM

OFDM is a multi-carrier technique that operates with specific orthogonally constraints between its sub-carriers[12]. This orthogonally yields very high spectral efficiency. Although the OFDM principle has been around 40 years, only the present technology level makes it feasible. OFDM allows the spectrum of each tone to overlap because they are orthogonal, and thus they do not interfere with each other. By allowing the tones to overlap, the overall amount of spectrum required is reduced.

The sinusoidal waveforms making up the tones in OFDM have the special property of being the only Eigen-functions of a linear channel. This special property prevents adjacent tones in OFDM systems from interfering with one another. To maintain the orthogonality in an OFDM system, a cyclic prefix is a critical concept.

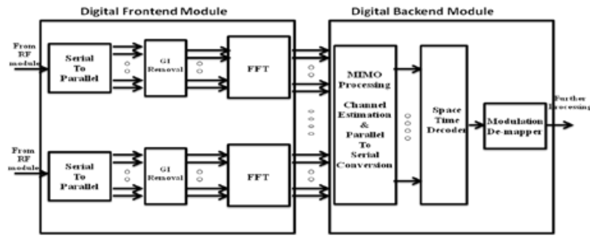
2. MIMO-OFDM

A MIMO-OFDM system [5] takes a data stream and splits it into N parallel data streams, each at a rate 1/N of the original rate, as depicted in fig 2.1 Each stream is then mapped to a tone at a unique frequency and combined together using the Inverse Fast Fourier Transform (IFFT) to yield the time-domain waveform to be transmitted.



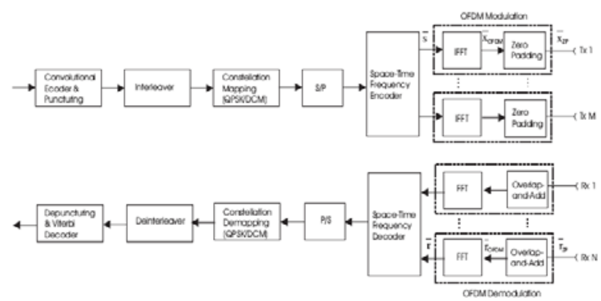
2.1 MIMO-OFDM transmitter section

The figure2.2 illustrates the digital module of receiver section of the MIMO-OFDM system. The OFDM signal is first de-serialized and then fed to the cyclic prefix removal system. After the cyclic prefix is attenuated from the OFDM signal, it is fed to the FFT section which de-maps the data from each of the subcarriers. This is then fed to the demodulator to recover the original data.



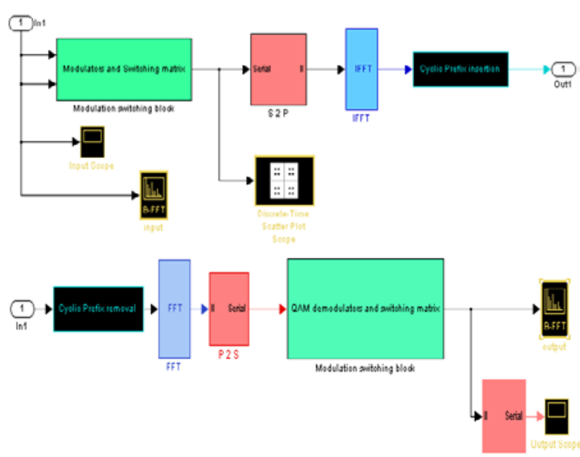
2.2 MIMO-OFDM Receiver section

Figure 2.3 shows the top level architecture of MIMO-OFDM, it consists of convolution encoder, interleaver, QPSK modulator, space time frequency encoder and OFDM modulation. The modified architecture proposed in this work consists of QPSK and QAM modulators that is run time reconfigurable depending upon the data rate and channel performance. Having both QPSK and QAM modulators achieves better performance for MIMO-OFDM modulators. The modulated data is space time-frequency encoded and is given to the OFDM modulator



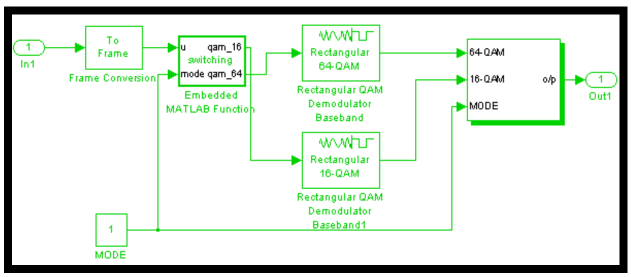
2.3 MIMO-OFDM Software reference model

Figure2.4 shows the Simulink model developed for MIMO-OFDM system. The developed model is as per the standard reference model reported in the literature.



2.4 Simulink model of MIMO-OFDM system

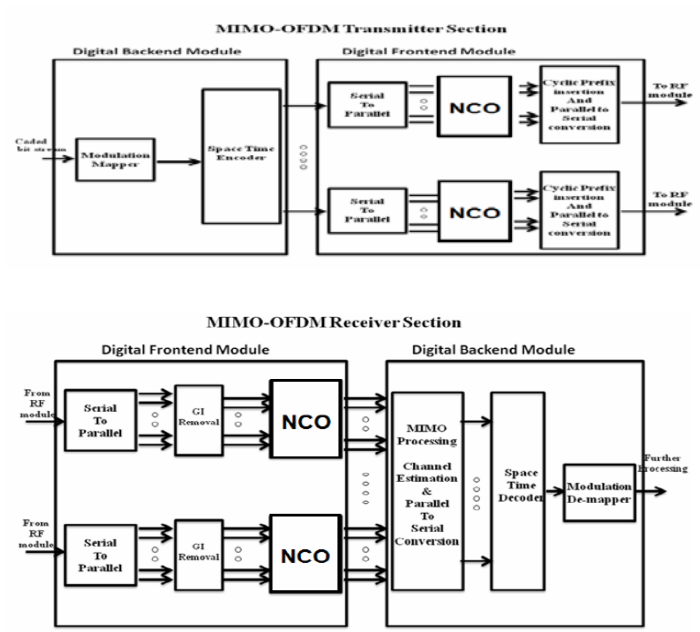
QAM modulator and demodulator is used to replace QPSK modulator and demodulator, QAM is used to enhance the data rate as well as reduce the bandwidth for signal transmission[8]. Figure2.5 shows the Simulink model of QAM modulator, in this work we have developed both 16 QAM and 64 QAM for modulation of data.



2.5 Simulink reconfigurable model of MIMO-OFDM system

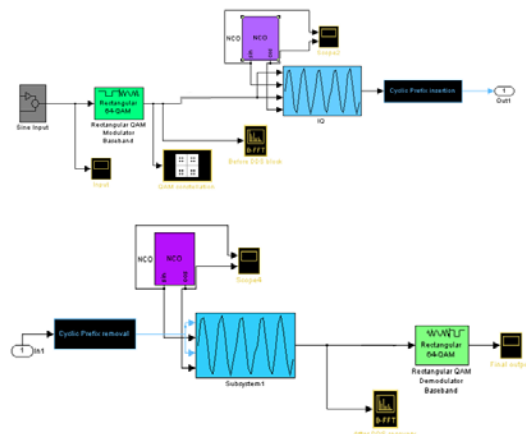
3. MODIFIED MIMO-OFDM

In order to enhance the performance of MIMO-OFDM, the FFT/IFFT blocks are replaced with NCO block. As the IFFT unit at the transmitter converts the incoming signal into multiple frequency band depending upon the phaseinformation, instead of IFFT, replacing with an NCO reduces the circuit complexity and also helps in achieving multiple frequency shifts. An NCO generates multiple frequency components based on corresponding phase information. Hence the modified system is more accurate than the existing architectures. The modified architecture is shown in figure 3.1 below



3.1 Modified MIMO-OFDM system

The modified architecture is modeled in Simulink and the results obtained based on an experimental setup for known sets of inputs are compared with the results of the software reference model. From the comparison of results it is found that the modified architecture achieves better accuracy and also more flexible. Figure 3.2 below shows the Simulink mode for modified MIMO-OFDM architecture.



3.2 Simulink model of modified MIMO-OFDM system

Table 1: Design specifications for OFDM model

Specifications	FFT-OFDM	NCO-OFDM
Maximum Input Frequency	100KHz	100KHz
Sample time	1/30000	1/30000
Modulation	64 QAM	64 QAM
FFT/IFFT size	256	-
Number of data points in look up table	-	17
Quarter wave sine look up table size	-	34 bytes
Spurious dynamic range	-	48dBc
Frequency resolution	-	30.5176 mHz
Channel	AWGN	AWGN
Channel SNR	60dB	60dB

IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 2, September 2011
ISSN (Online): 1694-0814
www.IJCSI.org

4. DESIGN OF NCO

The NCO block generates a multichannel real or complex sinusoidal signal, with independent frequency and phase in each output channel. The amplitude of the created signal is always 1. The implementation of a numerically controlled oscillator (NCO) has two distinct parts. First, a phase accumulator accumulates the phase increment and adds in the phase offset. In this stage, an optional internal dither signal can also be added. The NCO output is then calculated by quantizing the results of the phase accumulator section and using them to select values from a lookup table. Given a desired output frequency F_0, calculate the value of the Phase increment block parameter with

Phase increment = $(F_0 . 2^N)/F_S$

where N is the accumulator word length and

$F_S = 1/T_S = 1/\text{Sample time}$

The frequency resolution of an NCO is defined by

$\Delta f = 1/(T_S . 2^N)$ Hz

Given a desired phase offset (in radians), calculate the **Phase offset** block parameter with

Phase offset $= (2^N . \text{desired phase offset})/2\Pi$

The spurious free dynamic range (SFDR) is estimated as follows for a lookup table with entries, where P is the number of quantized accumulator bits:

SFDR = (6P) dB without dither
SFDR = (6P+12) dB with dither

This block uses a quarter-wave lookup table technique that stores table values from 0 to π/2. The block calculates other values on demand using the accumulator data type, then casts them into the output data type. This can lead to quantization effects at the range limits of a given data type. For example, consider a case where you would expect the value of the sine wave to be –1 at π. Because the lookup table value at that point must be calculated, the block might not yield exactly –1, depending on the precision of the accumulator and output data types. In this work the NCO designed can generate a maximum frequency of 49.9 MHz, which supports 64, 32 and 16 subcarrier applications. The 12-bit quantization along with the LUT of 4096 entries ensures a smoother sine and cosine oscillation generation. Figure 7 shows the NCO model designed. The 32-bit phase accumulator output is quantized to 12 bit output and is used to access the Contents of LUT. The LUT consists of sine and cosine data that is read out to generate the required signals for modulation.

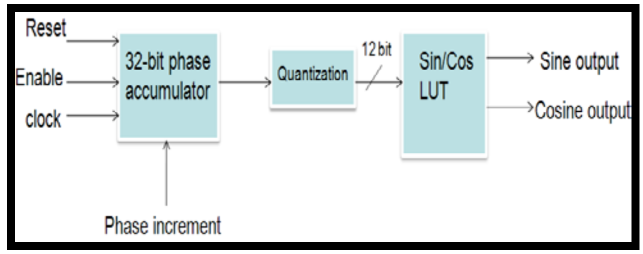


Figure: 4.1 NCO design

5. RESULTS

A sine wave input is applied as a test signal, the system developed is tested for its functionality, and results of the same are shown below.

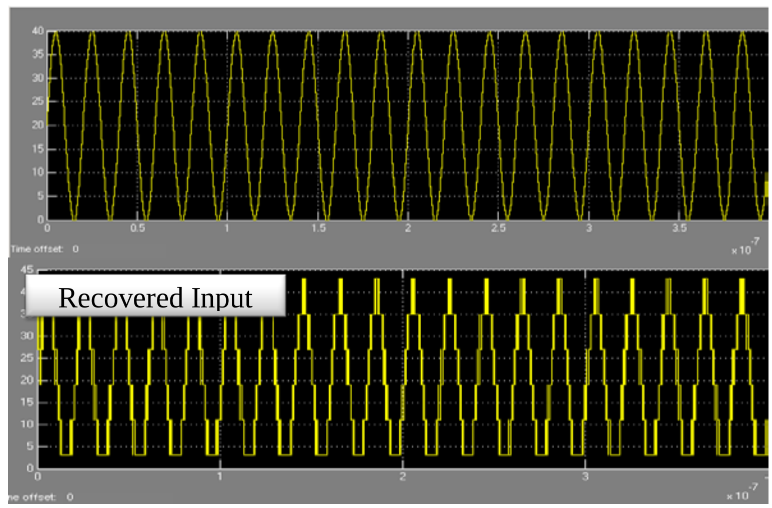


Fig 5.1 Simulation results of FFT based OFDM module

Figure shows the simulation results of OFDM with FFT and OFDM with NCO. The principle of orthogonality exists in the frequency spectrum shown below. The same set of orthogonality is also established in the NCO based OFDM module.

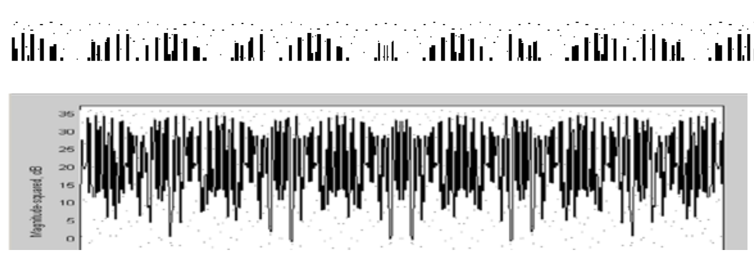


Fig 5.2: Simulation results of NCO based OFDM module.

Thus the proposed OFDM module using NCO achieves the same functionality as that of OFDM with FFT. Various test signals have been used to validate the developed modules. The results obtained are compared with the reference model. The simulink model developed is further modeled using Verilog HDL, the developed model is simulated using ModelSim and synthesized using Xilinx ISE for FPGA implementation. Spartan IIIE FPGA is targeted for implementation of modified architecture. The implementation results of modified architecture are compared with existing architecture and are presented in Table2 shown below.

Table 2: Comparison table of NCO with FFT/IFFT

Device Spartan 3 FPGA	Available resources	IFFT/ FFT	Utilization	NCO	Utilization
Slice Registers	19200	1231	6%	44	0.23 %
Slice LUTs	19200	971	5%	34	0.17 %
LUT-FF pairs	1289	913	70%	33	2.5%
Bonded IOBs	220	4	1%	83	37.7 %
Block RAM	32	0	0%	3	9.3%
Max. frequency		**372.75 MHz**		**469.39 MHz**	

5.1 Simulation results

System in QPSK mode

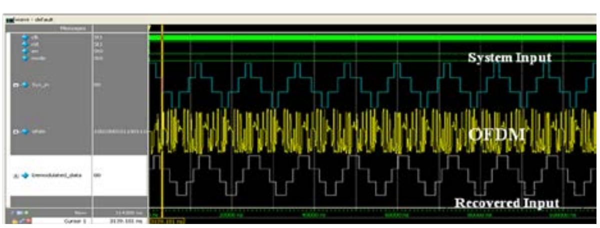


Fig 5.3 RTL Simulation of complete design in MODELSIM

System in QPSK mode switching to QAM mode

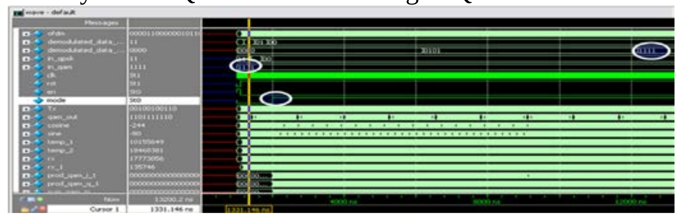


Fig 5.4 POST-SYNTHESIS Simulation of complete design in MODELSIM

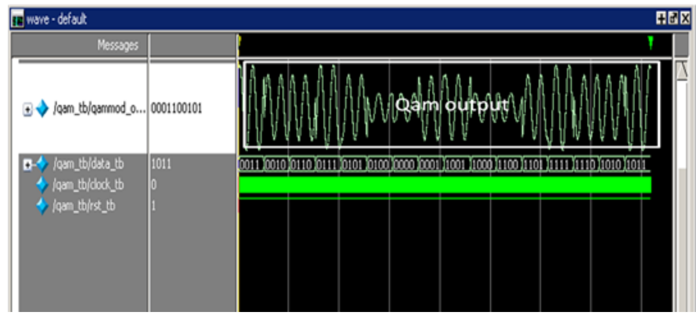


Fig 5.5 RTL Simulation of QAM in MODELSIM

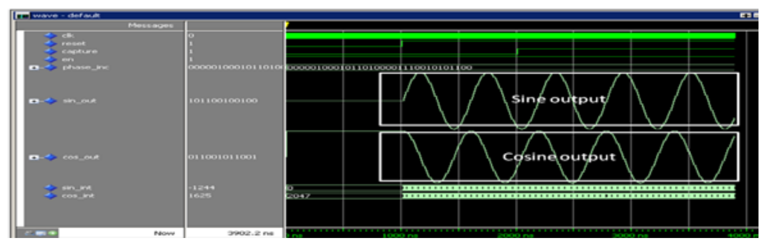


Fig 5.6 RTL Simulation of NCO in MODELSIM

System in QAM mode

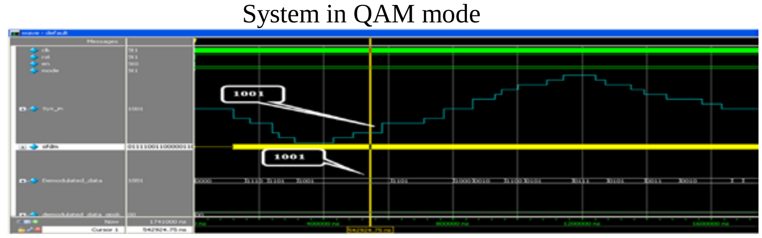


Fig 5.7 RTL Simulation of complete design in MODELSIM

6. CONCLUSIONS

The relevance of the MIMO-OFDM based systems is based on the fact that, these systems are immune to RF interferences, have high spectral efficiency, lower multipath distortion and support various modulation schemes for high data systems. The aforementioned facts suggest the Preference of FPGAs over ASIC implementation of the MIMO-OFDM system. The FPGA allows the rapid prototyping and implementation of the MIMO-OFDM system for its operational use. As the system evolves, the FPGA implementation provides the flexibility of reprogramability

The switching between the two modulation schemes, QPSK and 16-QAM, aids in re-configurability of the system, based on the channel characteristics ensuring least possible transmission errors. The NCO based design utilized just 30% of the available slice on SPARTAN 3 FPGA board, whereas the IFFT based design requires 103% of it. This is a 70% reduction in resource utilization. The maximum frequency achieved with NCO based was 25% higher than IFFT based design. The dynamic power of the NCO based architecture is 258.48 µW and leakage power at 20 µW, which is 10% lower than the IFFT, based design.

REFERENCES

[1] S.S.Riaz Ahamed, Performance analysis of OFDM, Journal of Theoretical and Applied Information Technology, 2008

[2] University of Alberta, MIMO History. Retrieved on September28,2009 from http://www.ece.ualberta.ca/~HCDC/mimohistory.html

[3] Jeffrey G. Andrews, Arunabha Ghosh and Rias Muhamed, Fundamentals of WiMax: Understanding Broadband Wireless Networking. Retrieved on October 1, 2009 from http://www.wimax.com/commentary/wimax_weekly/sidebar-1-1-a-brief-history-of-ofdm

[4] OFDM Techniques. Retrieved on October 1, 2009 from http://www.wireless-center.net/Next-Generation-Wireless/OFDM-Techniques.html

[5] T. Kaiser, A. Wilzeck, M. Berentsen, and M. Rupp, Prototyping for MIMO systems- an overview, Proceedings of 12th European Signal Processing Conference (EUSIPCO '04), pp. 681 to 688, Vienna, Austria, September 2004

[6] XILINX, (2009) Getting started with FPGAs. Retrieved on March08,2010 http://www.xilinx.com/company/gettingstarted/index.htm

[7] Oscar Robles Palacios and Carlos Silva Cardenas, Design and implementation of a reconfigurable OFDM modulator for software- defined radios, IEEE Transactions on Communications, 2008

[8] Qingbo Wang, Ling Zhuo, Viktor K. Prasanna and John Leon, A multi-mode reconfigurable OFDM communication system on FPGA, IEEE Transactions on Communications, 2008

[9] Ebrahim Saberinia, Ahmed H. Tewfik and Keshab K. Parhi, Pulsed-OFDM modulation for Ultrawideband communication, IEEE Transactions on Vehicular Technology, 2009

[10]Alfred Grau, Hamid Jafarkhani and Franco De Flaviis, A Reconfigurable Multiple-Input Multiple-Output Communication System, IEEE Transactions on Wireless Communications, vol. 7, No. 5, May 2008

[11] Bruno Bougard, Gregory Lenoir, Antoine Dejonghe, Liesbet Van der Perre, Francky Catthoor and Wim Dehaene, SmartMIMO: An Energy-Aware AdaptiveMIMO-OFDM Radio Link Control for Next-GenerationWireless Local Area Networks, EURASIP Journal on Wireless Communications and Networking, 2007

Acknowledgement

Authors would like to thank the Management Vemana Institute of Technology, KRJS, Bangalore, for funding and providing necessary help to complete this paper.

Vitae:

Veena M.B.: received B.E.,& M.E.,degree in Electronics & communication from Mysore & Bangalore university respectively. Currently pursuing Ph.D from V.T.U., Belgaum, at research center SJCE, Mysore. India. Presently working as a Assistant professor, Dept. of Telecommunication Engineering, Vemana Institute of Technology, Bangalore, India. & research interest in the field of MIMO wireless communication, circuits & systems, DSP, VLSI design & VLSI architectures for wireless systems.

M.N.Shanmukha swamy : received the B.E. & M.E.,degree in Electronics & communication from Mysore university. Ph.D from IISC, Bangalore, India. Presently working as a professor, Dept of Electronics & communication Engineering, SJCE, Mysore, India. & research interest in the field of wireless communication system, circuits & systems, VLSI. Adhoc networks & Image processing.

Development of CAD System Based on Enhanced Clustering Based Segmentation Algorithm for Detection of Masses in Breast DCE-MRI

D. Janaki Sathya[1], K. Geetha[2]

**[1] EEE Department, Karpagam University,
Coimbatore, Tamilnadu, India**

**[2] EEE Department, Karpagam Institute of Technology,
Coimbatore, Tamilnadu, India**

Abstract

Breast cancer continues to be a significant public health problem in the world. Early detection is the key for improving breast cancer prognosis. Mammography is currently the primary method of early detection. But recent research has shown that many cases missed by mammography can be detected in Breast DCE-MRI. Magnetic Resonance (MR) imaging is emerging as the most sensitive modality that is currently available for the detection of primary or recurrent breast cancer. Breast DCE-MRI is more effective than mammography, because it generates much more data. Magnetic resonance imaging (MRI) is emerging as a powerful tool for the diagnosis of breast abnormalities. Computer Aided Detection (CAD) is of great help to this situation and image segmentation is most important process of computer Aided Detection, Magnetic Resonance Imaging data are a major challenge to any image processing software because of the huge amount of image voxels. Automatic approaches to breast cancer detection can help radiologists in this hard task and speed up the inspection process. To segment the mass of the breast region from 3D MRI set, a multistage image processing procedure was proposed. Data acquisition, processing and visualization techniques facilitate diagnosis. Image segmentation is an established necessity for an improved analysis of Magnetic Resonance (MR) images. Segmentation from MR images may aid in tumor treatment by tracking the progress of tumor growth and shrinkage. The advantages of Magnetic Resonance Imaging are that the spatial resolution is high and provides detailed images. The tumor segmentation in Breast MRI image is difficult due to the complicated galactophore structure. The work in this paper attempts to accurately segment the abnormal breast mass in DCE-MRI Images. The mass is segmented using a novel clustering algorithm based on unsupervised segmentation, through neural network techniques, of an optimized space in which to perform clustering. The effectiveness of the proposed technique is determined by the extent to which potential abnormalities can be extracted from corresponding breast MRI based on its analysis, this algorithm also proposes changes that could reduce this error, and help to give good results all around. Tests performed on both real and simulated MR images shows good result.

Keywords: *Image Voxels, Neural Networks, Clustering, Thresholding, Image Segmentation, Mass Detection.*

1. Introduction

Breast cancer happens to over 8% women during their lifetime, and is the leading cause of death of women [6]. Currently the most effective method for early detection and screening of breast cancers is MRI [7]. Microcalcifications and masses are two important early signs of the diseases [15]. It is more difficult to detect masses than microcalcifications because their features can be obscured or similar to normal breast parenchyma. Masses are quite subtle and often occurred in the dense areas of the breast tissue, have smoother boundaries than microcalcifications and have many shapes such as circumscribed, speculated (or stellate), lobulated or ill-defined. The circumscribed ones usually have a distinct boundaries, 2–30mm in diameters and are high-density radiopaque; the speculated ones have rough, star-shaped boundaries; and the lobulated ones have irregular shapes [8]. Masses must be classified as benign and malignant in order to improve the biopsy yield ratio. Generally speaking, masses with radiopaque and more irregular shapes are usually malignant, and those combined with radiolucent shapes are benign [14]. A DCE-MRI is basically distinct with four levels of the intensities: background, fat tissue, breast parenchyma and calcifications with increasing intensity. Masses develop from the epithelial and connective tissues of breasts and their densities on MRI's blend with parenchyma patterns. Several studies have revealed a positive association of tissue type with breast cancer risks [9, 10]. Women who have breast cancers can easily get ontralateral cancers in the other side breast [11, 12]. Distinguishing a new primary from metastasis was not always possible due to their similar features. Asymmetry of breast parenchyma between the two sides has been one of the most useful signs for detecting primary breast cancer [13].

Segmentation of medical images is a challenging and complex task. Medical image segmentation has been an active research area for a long time. There are many segmentation algorithms [28, 29, 30] but there is not a generic algorithm for totally successful segmentation of medical images. The Segmentation of images holds an important position in the area of image processing. It becomes more important while typically dealing with medical images where pre-surgery and post-surgery decisions are required for the purpose of initiating and speeding up the recovery process. Computer aided detection of abnormal growth of tissues is primarily motivated by the necessity of achieving maximum possible accuracy [5]. Manual segmentation of these abnormal tissues cannot be compared with modern day's high speed computing machines which enable us to visually observe the volume and location of unwanted tissues. In the proposed method the acquired DCE-MRI Image of the breast will be processed in 4 different stages,

1. Pre processing
2. Clustering
3. Edge enhancement
4. Extraction of mass.

The proposed algorithm effectively segments the mass region more accurately with high efficiency. The paper is organized as follows: Section 2, materials and methods are described; the proposed methodology explained in section 3 while results are reported in Section 4. Conclusions and possible further developments are illustrated in Section 5

2. Materials and methods

This paper presents the visual and statistical results of applying CAD System Based on Enhanced Clustering Based Segmentation Algorithm for Detection of Masses in Breast DCE-MRI over the simulated and real images of breast DCE-MRI images. Images used in this study were acquired with patients prone to 1.5T scanner with use of a dedicated surface breast coil array. The imaging protocol included bilateral fat suppressed T weighted images in the sagittal plane of 1mm slice thickness and a slab interleaved 3D fat suppressed spoiled gradient echo after the injection of contrast. One slice can contain 512×512 pixels.

2.1 Medical Imaging

The medical research has been quite receptive of image processing in applications like X-ray, Computer Aided Tomography, Ultrasound and Magnetic Resonance. The output of these techniques, an image of the patient's body, allows the physician to examine and diagnose without the need of surgery. Traditional imaging modalities, like X-Ray mammography, do not provide certain and reliable results on young women or in women who underwent surgical interventions. Three-dimensional breast MRI has proven to be a valuable tool for disambiguate uncertain mammographic findings and for the pre-operative planning. DCE-MRI has a statistically significant ability to diagnose malignancy in suspicious breast lesions detected using other diagnostic modalities. The sensitivity of breast MRI is relatively high (false negative rate is low). The reported specificities of MR have been variable ranging from 50% to 80.8%, and the sensitivity of MR will be 94.2% for even dense breast [35].

In recent years, Magnetic resonance imaging (MRI) of the breast has been increasingly used in the diagnosis of radiographically dense breasts, assessment of disease extent in the patient with newly diagnosed breast cancer, problem-solving applications for difficult diagnostic evaluations, for difficult diagnostic evaluations, for screening high risk women for early cancer detection, and on monitoring response to therapy. Magnetic Resonance Imaging is the most attractive alternative to Mammography. In order to be effective, breast MRI requires the use of a paramagnetic contrast agent. When used in combination with the contrast agent, the examination is called Dynamic-contrast-enhanced - MRI (DCE-MRI). The contrast-agent has no contraindications and is tolerated better than the radiations absorbed during a single mammography [42].

DCE-MRI is sensitive for detecting some cancers which could be missed by mammography. In addition, DCE-MRI can help radiologists and other specialists determine how to treat breast cancer patients by identifying the stage of the disease. It is highly effective to image breast after breast surgery or radiation therapy. DCE-MRI forms 3D uncompressed image. It can perform with all women including who are not suitable for mammography, such as young women with dense breast and women with silicone-filled breast implants. Since it uses magnetic fields, DCE-MRI has no harmful effects on human bodies. However, DCE- MRI takes rather long time to perform and has high cost which is more than ten times greater than mammography. Its low resolution limits its application to very small lesions or micro calcifications.

The promising potential of MRI in diagnosis of breast cancer, as a complementary modality to X-ray mammography, has been well recognized [1, 2, 3]. Despite its well-recognized utilities, however, the technique has not been introduced to routine clinical breast imaging. One of the most important obstacles has been the lack of standardization in terms of interpretation guidelines [2, 4]. The reproducibility, effectiveness and relative significance of interpretation criteria in the literature are far from being well evaluated. The purpose of the proposed research is to develop computerized methods to take full advantage of the wealth information that MRI offers to improve methods for the diagnosis and interpretation of breast MR images. The research involved investigation of automatic methods for image artifacts correction and tumor segmentation. Our hypothesis was that investigation of advanced image analysis algorithms would improve the performance of existing conventional methods in the task of detecting the tumors in the breast MR Images.

2.2 Image processing tools

Segmentation is a fundamental tool which aids in identification and quantitative evaluation. It conditions the quality of analysis. Computer based segmentation has reminded largely an experimental work many efforts have exploited MRI's multi-dimensional data capability through multi-spectral analysis. Segmentation as defined by Kapur [16] is "a labeling problem in which the goal is to assign to each voxel in an input gray-level image, a unique label that represents an anatomical structure". Many approaches to MRI segmentation both supervised and unsupervised have been proposed in literature [17, 22, 24, 25, 29]. Among the unsupervised segmentation techniques, the K-means algorithm is applied. Self-organizing feature maps (SOM) in a hierarchical manner is developed, with this approach using a certain degree of supervision. An acceptable classification is obtained when applied to test images.

In particular, Neural Networks try to simulate a structure similar to the one that is believed the human brain has. Two dimensional layers of cellular modules that are densely interconnected between them model most neural networks in the brain, especially in the cortex [28]. This area of the brain is organized into several sensory modalities such as speech or hearing. The engineering approach of neural networks develops hardware or software inspired by the brain's structure [27].

Neural network attracted more and more researchers for its abilities of parallel operation, self-learning, fault tolerance, associative memory, multifactorial optimization and extensibility [26]. Neural network based clustering has yielded good results [18, 19], yet the possibility of transforming the input space in order to facilitate segmentation has been largely unexplored [34]. This paper proposes a new unsupervised algorithm for MR image segmentation is implemented. In this method, classical Kohonen map-based clustering is enhanced through the search of an optimized space in which to operate the clustering [32, 33]. It allows for the ability to make the clustering methods able to retain more information from the original image than the crisp or hard segmentation methods [31].

3. Implementation of the proposed algorithm

The framework of the present work is the development of a novel algorithm acting as a support to the diagnosis process for those affections that require medical imaging. Such tools present to the clinician both a qualitative and a quantitative description of the disease. In this proposed algorithm each input is the image dataset, which undergoes a number of sequential processing steps: preprocessing, clustering, error back propagation, classification, enhancing edges of classification output and segmenting region of interest as shown in figure 1. Magnetic resonance imaging is a tomography technique, i.e. each image comprises of a number of slices, each corresponding to a given slice of tissue; following the pulse repetition period (TR) and parameters related to the applied radio-frequency magnetic field. It is possible to obtain images with different contrast, each reflecting a different parameter regulating the relaxation of the excited tissues. After the clustering process, each cluster is manually interpreted and assigned to a proper tissue class.

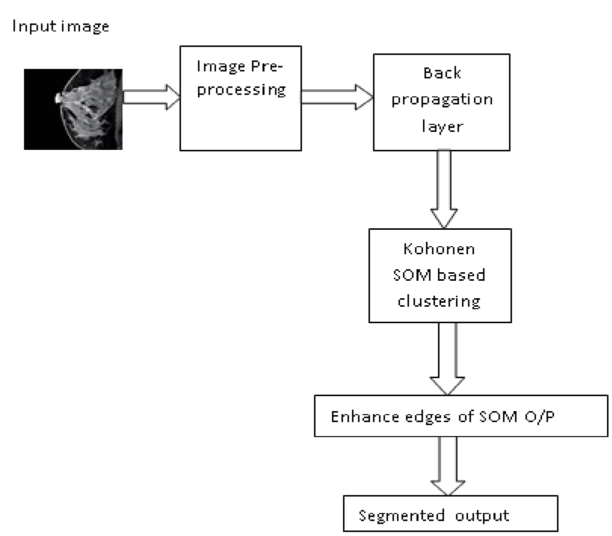


Fig. 1 Segmentation Process

3.1 Preprocessing

In segmenting DCE-MRI data, three main difficulties are considered namely: noise, partial volume effects (where more than one tissue is inside a voxel volume) and intensity in-homogeneity [30]. The majority of intensity in-homogeneities are caused by the irregularities of the scanner magnetic fields–static (B0), radio-frequency (B1) and gradient fields, which produce spatial changes in tissue static. Partial volume effects occur where multiple tissues contribute to a single voxel, making the distinction between tissues along boundaries more difficult. Noise in MR images can induce segmentation regions to become disconnection.
An important part of any image processing system is represented by the pre-processing phase. This phase could imply contrast enhancement techniques or methods for removing the noise. Preprocessing aims at improving the quality of each input image and reducing the computational burden for subsequent analysis steps, Each input voxel is formed as a feature vector as described by the preprocessing technique proposed in [30, 37]. This aims at compensating the effects of random noise, while minimizing the loss of resolution.

All feature vectors are normalized prior to segmentation by subtracting the mean and dividing by the standard deviation, where the mean and standard deviation are estimated independently for each slice.

3.2 Clustering

Clustering is a technique for finding similarity groups in data, called clusters. i.e., it groups data instances that are similar to (near) each other in one cluster and data instances that are very different (far away) from each other into different clusters. Clustering is often called an unsupervised learning task as no class values denoting an apriori grouping of the data instances are given, which is the case in supervised learning [33]. Unsupervised methods, on the other hand, do not require any human interference and can segment the breast with high precision. For this reason, unsupervised methods are preferred over conventional methods. Many unsupervised methods such as Fuzzy c-means, Self-Organizing map, etc. exist but Kohonen's Competitive Learning Algorithms yields good results [32, 34, 37].

The proposed network architecture consists of two fully interconnected layers; the first layer, composed of computing elements of order zero with linear activation function, followed by a second layer of computing elements of order two, with gaussian activation function. Let X be the input pattern, H the output of the hidden layer and Y the output of the network. W and Z are the weight vectors of the first and second layer, respectively. In order to jointly optimize both layers, training is carried out in two steps. In the first step, the second layer is trained using the standard Kohonen rule for unsupervised learning at each iteration, the winning neuron's centers are adjusted according to equation 1Footnotes should be typed in singled-line spacing at the bottom of the page and column where it is cited. Footnotes should be rare

$$\Delta Z_{ji} = \eta_z.(H_i - Z_{ji}) \qquad (1)$$

Where

ΔZ_{ji} = Change in weight vector

η_z = learning rate of the Kohonen layer

H_i = Out put of the ith neuron of the hidden layer

Z_{ji} = The weight vector of the Winning neuron

The weights of the neighboring neurons are updated according to equation 2

$$\Delta Z_{ji} = \eta_z.f_{neigh}(H_i - Z_{ji}) \qquad (2)$$

Where

f_{neigh} = Gaussian activation function

Contrarily to the second layer, the first layer is trained using Enhanced version of error back-propagation with the linear activation function, search of feature space. In supervised learning schemes, the error is given by.

$$E = \sum_P \left\| Y^P - T^P \right\|^2 \qquad (3)$$

Where $\mathbf{T^P}$ is the user-supplied target associated to the $\mathbf{P^{th}}$ training pattern. Here the target is determined by associating each input pattern with the winning neuron. Intuitively, this corresponds to searching a linear transformation of the feature space, requiring that input patterns be as close as possible to the associated centroids. The hidden layer is then trained using the classical delta rule for training and is derived from equation (3)

$$\frac{\partial \mathrm{E}}{\partial \mathrm{W}_{\mathrm{li}}} = \sum_{p} \eta_{p} \sum_{j} (\delta_{j}^{lp} . x_{i}^{p}) \quad (4)$$

Where p denotes the pth input pattern and

$$\delta_{j}^{lp} = y_{j}^{p} - t_{j}^{p}$$

.

The weights of the first layer are then updated according to equation 5

$$\Delta W_{ij}(t+1) = -\eta_{w} . \frac{\partial E}{\partial W_{ij}} + \mu \Delta W_{ij}(t) \quad (5)$$

$$\Delta W_{ij}(t) = \eta_{w} \delta_{p} H_{j}$$

Where

μ = momentum factor

η_{w} = learning rate of the Back propagation layer

The momentum term introduces the old weight change as a parameter for the computation of the new weight change. This avoids oscillation problems common with the regular back propagation algorithm when the error surface has a very narrow minimum area. Momentum allows the net to make reasonably large weight adjustments as long as the corrections are in the same general direction for several patterns. Using smaller learning rate prevents a large response to the error from any training pattern.

The first layer consists of 5computing elements with linear activation function. Thus, not only the hidden layer performs a linear transformation of the input space, but it also reduces the dimensionality of the feature space. This allows obtaining, in average, better experimental results than when all features are retained in the clustering step. The second layer has 5 computing elements. Five clusters are sufficient to discriminate between adipose tissue, glandular tissue, ducts, benign and malignant masses.

The network is separately trained for each image to account for inhomogeneities in intensity across different images by randomly selecting pixels per image as the training set. A Gaussian neighborhood function f_{neigh} is used for unsupervised training. An adaptive learning coefficient is initially selected for the first layer as η_w and for the second one as η_z. If the error increases, η is decreased and weight values are set to those of the previous iteration, whereas if the error decreases below a predefined threshold, η is increased. Finally, training is stopped when a predetermined level of error is reached.

3.3 Edge Enhancement

Digital image enhancement techniques are concerned with improving the quality of the digital image. The principal objective of enhancement techniques is to produce an image which is better and more suitable than the original image for a specific application. This process detects boundaries between objects and background in the image.

Many characteristics are used to segment an image into regions e.g. colour, brightness, texture and edge detection. Usually, the obtained edges need some additional improvement for the satisfactory segmentation. Linear filters have been used to solve many image enhancement problems. The unsharp filter is a simple sharpening operator which derives its name from the fact that it enhances edges through a procedure which subtracts an unsharp, or smoothed, version of an image from the original image. The unsharp filtering technique is commonly used in the photographic and printing industries for crispening edges.

Unsharp masking produces an edge image $g(x, y)$ from an input image $f(x, y)$ via

$$g(x, y) = f(x, y) - f_{smooth}(x, y) \quad (6)$$

where $f_{smooth}(x, y)$ is a smoothed version of $f(x, y)$ as illustrated in Figure 2.

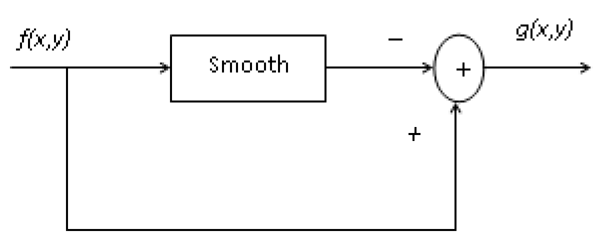


Fig. 2 Unsharp filter

Then the enhanced SOM based k-means clustered output image is then edge enhanced by unsharp filter as illustrated in [43] which is used to extract the edges of the tumour very efficiently in MRI images. The unsharp filter can be implemented using an appropriately defined lowpass filter to produce the smoothed version of an image which is then pixel subtracted from the original image in order to produce a description of image edges, i.e. a high passed image.

3.4 Segmentation

The segmentation used for extraction of masses in MR images is the region based segmentation. In region-based techniques, segmentation of an object is achieved by identifying all pixels that belong to the object based on the intensity of pixels. They are looking for the regions satisfying a given homogeneity criterion. Since in MR images masses mostly have high contrast and ill-defined edges, it is difficult to determine their boundary with edge-based techniques. Region-based techniques are more suitable for MR images since suspicious regions are brighter than the surrounding tissues. There are many region-based techniques such as Region growing [38], Watershed algorithm [39], and Thresholding [40].

In this work utilizes region segmentation based-thresholding. Thresholding is based on the image histogram; or local statistics such as mean value and standard deviation, or the local gradient. When only one threshold is selected for the entire image, based on the image histogram, it is called global thresholding. If the threshold depends on local properties, it called local thresholding. If the thresholds are selected independently for each pixel, it called dynamic or adaptive thresholding. The selection of threshold value is determined by experimenting with various threshold values and the best threshold value has been selected for breast DCE-MRI images. The thresholded image g (x, y) is defined as,

$$g(x,y) = \begin{cases} 1 if(x,y) > T \\ 0 if(x,y) \leq T \end{cases}$$

4. Experimental Results and Discussion

In this section, the results obtained using real and simulated DCE- MR Images are illustrated. Ultimately, the effectiveness of the proposed technique is determined by the extent to which potential abnormalities can be extracted from corresponding breast MRI based on its analysis.

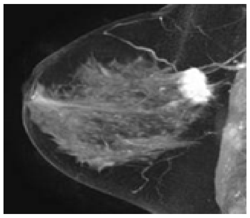
4.1 Testing on Breast DCE- MR Images

The use of simulated images simplifies the task of validating a segmentation method as a reproducible. Moreover, it allows to separately testing the proposed segmentation method stability against intensity inhomogeneities and random noise [31]. But in this paper both real and simulated images are used for segmentation procedure, Since the real images obviously contains the noise induced during the imaging process, when the real image is used for validating the algorithm then we can find how efficiently the algorithm works for the image with noise, so no need to introduce or add noise to simulated image for testing the algorithm efficiency. The reference image is selected. With each cluster is associated with the most probable tissue class using maximum likelihood estimation.

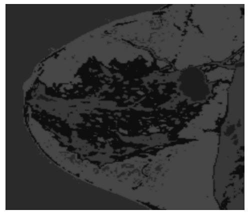
A set of original and simulated breast DCE-MR images representative slice are considered for estimating the validation of the algorithm is shown in figure 3 to figure 7. To evaluate the results, trainings for each reference slice were performed with different random initial conditions for the centers of the neurons in the second layer. It is well-known that the training speed depends on the choice of the learning rate. If the learning rate is small, the learning process is stable but at the expense of computation time [26]. If the learning rate is too large, the estimation of the weights may diverge. Because of fast convergence in using SOM with adaptive learning rate, it can be applied in online applications. The lower learning rate provides better convergence and better quality than higher learning rates.

The abnormal image pairs were used to measure performance. The true positive detection rate and the number of false positive detection rate at various thresholds of the images are used to measure the algorithm's performance. These rates are represented using Receiver Operating Characteristic (ROC) curves. True Positive (TP) and False Positive (FP) rates are calculated at different thresholds selected on image pixels to generate an ROC curve and the best solution has been plotted as illustrated in figure 9. A region extracted in the asymmetry image, which overlaps with a true abnormality as provided in the ground truth of the image, is called a true positive detection. An overlap means that at least 80% of the region extracted lies within the circle indicating a true abnormality as determined using SOM with adaptive learning rate; it can be applied in online applications. The lower learning rate provides better convergence and better quality than higher learning rates as illustrated in [37].

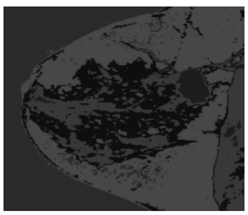
The preliminary results of the proposed algorithm illustrated in figures 3and figure 4 shows a reasonable match between manual division and that of automatic method. In the best case, there is an average difference of only 20-30 pixels. This was not always seen, as in the worst case the difference was as great as 150 pixels. This algorithm also proposes changes that could reduce this error, and help to give good results all around.



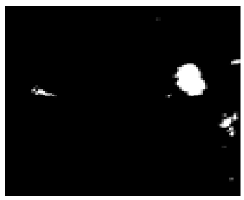
(a) Original image



(b) clustering output

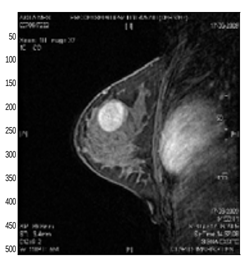


(c) Edge enhanced output

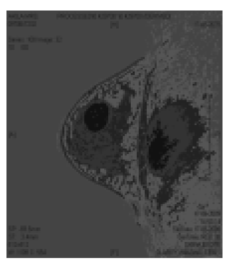


(d) mass segmented separately

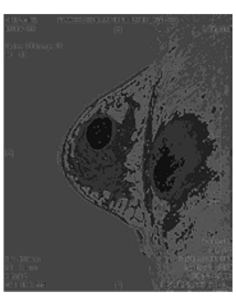
Fig. 3 A representative simulated breast image with mass and the corresponding segmentation of the mass separately



(a) Original image



(b) clustering output



(c) Edge enhanced output

(d) mass segmented separately

Fig. 4 A representative original breast image with tumor obtained from 1.5Telsa MRI and the corresponding segmentation of the mass separately

4.2 Performance evaluation of the algorithm

An objective method is needed to evaluate the performance of the new proposed image segmentation algorithm. The most important performance criterion is accuracy that is the degree to which an algorithm's segmentation matches some reference standard segmentation [36]. A number of similarity coefficients are used to specify how well a given segmentation matches a reference and the performance of the segmentation depends upon the learning rate factor. The performance of clustering output for different learning rate is illustrated in figure 5. In general, it is expected that the true positive detection rate in an ROC curve will continue to increase or remain constant as the number of false positives increase. The figure 6 illustrates the ROC curve performance of the algorithm for good values of the learning rate and the threshold value; if the threshold value is low true detections may become merged with false positive regions. The Table 1 shows the comparison of sensitivity and specificity rates of Breast Cancer using Clinical examination, mammography and MRI for disease diagnosis [35]. Evaluating the results obtained, it's found that the best results obtained when using K-NN classifier especially with using feature vector and enhanced images.

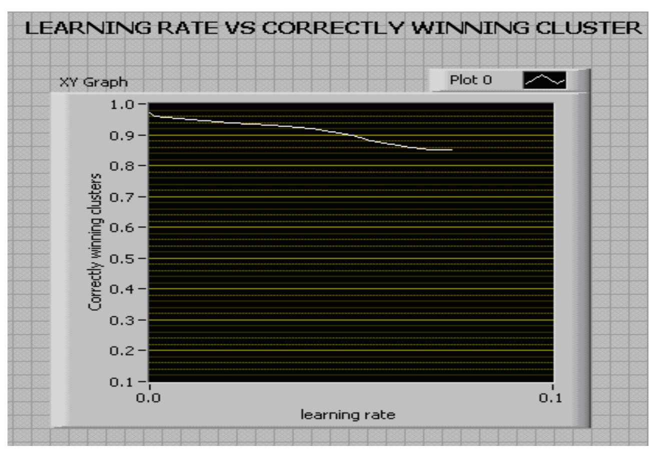


Figure 5 The performance of Clustering results with varying levels of learning rate

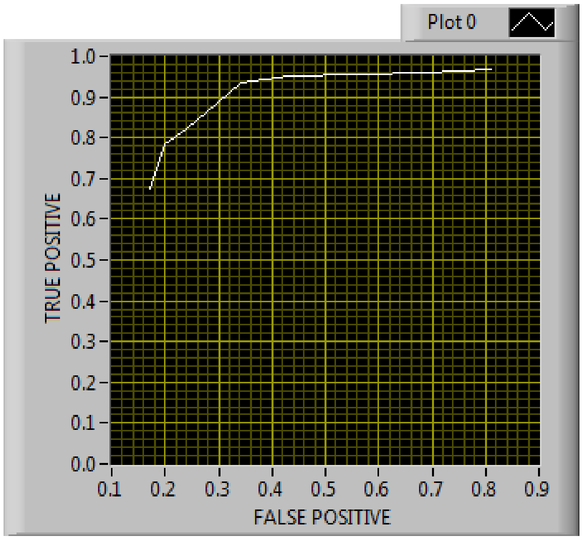


Figure 6 ROC Curve of Proposed Algorithm

5. Conclusions

Breast cancer is one of the major causes of death among women. So early diagnosis through regular screening and timely treatment has been shown to prevent cancer. In this paper we have presented a novel segmentation approach to identify the presence of breast cancer mass in breast DCE-MRI images. The proposed work utilizes the unsupervised clustering capabilities of a Kohonen self-organizing map with a linear transformation of the input space. Enhanced back propagation based Neural network is employed for clear identification of clusters.

The accurate and effective algorithm for segmenting image is very useful in many fields, especially in medical image. The training set originally had large dimension of data matrix, so the program used to reduce the dimension of training set and the new algorithm method is applied to show the ability of the method. A Self-Organizing Map network was programmed to receive images, as input signal regions. For this work, breast MRI images were used and segmentation of tumor is obtained. The proposed technique was evaluated on real and simulated DCE-MR images, showing promising performances from a qualitative point of view. Furthermore, being the proposed technique fully unsupervised, and the results substantially independent of the initial network conditions, Future efforts will be devoted to the further testing of the proposed technique, both from a qualitative and quantitative point of view, and to its application to the study of breast pathologies, in particular to breast tumor diagnosis and follow-up. The software's used for this algorithm development are MATLAB 7.5 and LABVIEW 10.0.

Acknowledgments

The authors would like to thank Dr.K.S.Murugan MD, DNB (Rad), Fellow in MRI (Ger) of clarity imaging, Coimbatore, India, for providing the breast DCE-MRI images used in this paper.

References

[1] E. A. Morris, "Breast cancer imaging with MRI," Radiol. Clin. North Am. 2002, **40**, 443–466.

[2] M. D. Schnall, "Breast MR imaging," Radiol. Clin. North Am. 2003, **41**, 43–50.

[3] P. J. Kneeshaw, L. W. Turnbull, and P. J. Drew, "Current applications and future direction of MR mammography," Br. J. Cancer. 2003, **88**, 4–10.

[4] L. W. Nunes, M. D. Schnall, S. G. Orel, M. G. Hochman, C. P. Langlotz, and C. A. Reynolds, "Breast MR imaging: Interpretation model," Radiology, 1997, 202, 833–841.

[5] Kunio Doi, "Computer-Aided Diagnosis in Medical Imaging: Historical Review, Current Status and Future Potential," Computerized Medical Imaging and Graphics July 2007, vol.31, no.4-5, pp.198-211.

[6] A.S. Constantinidis, M.C. Fairhurst, F. Deravi, M. Hanson, C.P. Wells, C. Chapman-Jones, Evaluating classification strategies for detection of circumscribed masses in digital mammograms, in: Proceedings of 7th International Conference on Image Processing and its Applications, 1999, pp. 435–439.

[7] K. Bovis, S. Singh, J. Fieldsend, C. Pinder, Identification of masses in digital mammograms with MLP and RBF nets, in: Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks Com,2000, pp. 342–347.

[8] I. Christoyianni, E. Dermatas, G. Kokkinakis, Fast detection of masses in computer-aided mammography, IEEE Signal Process. Mag. 2000, 17 (1), 54–64.

[9] R.L. Egan, R.C. Mosteller, Breast cancer mammography patterns, Cancer1997, 40, 2087–2090.

[10] J.N. Wolfe, Breast patterns as an index of risk for developing breast cancer, Am. J. Roentgen.1976, 126, 1130–1139.

[11] H.H. Storm, O.M. Jensen, Risk of contralateral breast cancer in Denmark 1943-80, Br. J. Cancer 1986, 54, 483–492.

[12]G.F. Robbins, J.W. Berg, Bilateral primary breast cancers, A Prospective Clinicopathol. Study Cancer 1964, 17, 1501–1527.

[13]T.J. Rissanen, H.P. Makarainen, M.A. Apaja-Sarkkinen, E.L.Lindholm, Mammography and ultrasound in the diagnosis of contralateral breast cancer, Acta Radiol.1995, 36, 358–366.

[14] B. Zheng, Y.H. Chang, X.H. Wang, W.F. Good, Comparison of artificial neural network and Bayesian belief network in a computer assisted diagnosis scheme for mammography, IEEE International Conference on Neural Networks, 1999, pp. 4181–4185.

[15] H.D. Cheng, X.P. Cai, X.W. Chen, L.M. Hu, X.L. Lou, Computer-aided detection and classification of microcalcifications in mammograms: a survey, Pattern Recognition 2003, 36, 2967–2991.

[16] T Kapur. Model based three dimensional Medical Image Segmentation, Ph.D. Thesis, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1999.

[17] L. P. Clarke, R. P. Velthuizen, M. A. Chamaco, J. J. Heine, M. Vaidyanathan, L.O. Hall et al.: MRI Segmentation: methods and applications. Magnetic Resonance Imaging, 1995, 13, 343-368.

[18] L. Morra, F. Lamberti, C. Demartini: A neural network approach to unsupervised segmentation of single-channel MR images. Proc. of the 1st IEEE/EMBS Conference on Neural Engineering, 2003, 1, 515-518.

[19] W.E. Reddick, J. O. Glass, E. N. Cook, T. D. Elkin, R. J. Deaton: Automated segmentation and classification of multispectral magnetic resonance images of brain using artificial neural networks. IEEE Trans. on Medical Imaging, 1997, 16, 911-918.

[20] M. S. Atkins, B. T. Mackiewich: Fully automatic segmentation of the brain in MRI. IEEE Transaction on Medical Imaging, 1998, 17, 98-107.

[21] R.K-S. Kwan, A.C. Evans, G.B. Pike: MRI simulation-based evaluation of image processing and classification methods. IEEE Transaction on Medical Imaging,1999,18, 1085-1097.

[22] Constantino Carlos Reyes-Aldasoro, "Image Segmentation with Kohonen Neural Network Self- Organising Maps". International Conference on Telecommunications ICT 2000 May 2000.

[23] Clustering,http://www2.cs.uregina.ca/hamilton/courses/ 831/ notes/ clustering/ clustering.html.

[24] Erhan Gokcay, A New Clustering Algorithm For Segmentation Of Magnetic Resonance Images, Thesis for Doctor of Philosophy, University of Florida, 2000.

[25] M. C. Clark., L.O. Hall., D. B. Goldof., L. P. Clarke., R. P. Velthuizen., M.S. Silbiger., "MRI segmentation using fuzzy clustering techniques", IEEE Engineering in Medicine and Biology, November 1994 pp. 730/742.

[26] Laurene Fausett "Fundamentals of Neural Networks" , Prentice – Hall, Inc. 1994

[27] Mohamad H Hassoun, "Fundamentals of Artificial Neural Networks", MIT press. 1995

[28] J. G. Taylor and Mannion C.L.T, "Theory and applications of Neural Networks", springer.1992

[29] S. Datta, B. R. Sajja, and P. A. Narayana, " Generalized fuzzy clustering for segmentation of multi- spectral magnetic resonance images ", Computerized Medical Imaging and Graphics, 2008, vol. 32, pp: 353–366.

[30] W. M. Wells, W. E. L. Grimson, and R. Kikins, "Adaptive segmentation of MRI data", IEEE Trans. Medical Imaging,1996, vol. 15, pp: 429-442.

[31] D. L. Pham., C. Y. Xu, and J. L. Prince, "A survey of current methods in medical image segmentation", Annual Review of Biomedical Engineering, 2000, vol. 2, pp. 315-337.

[32] Kong, W. Lu, J. Wang, N. Che, and Y. Lu, " A Modified Fuzzy Kohonen's Competitive Learning Algorithms Incorporating Local Information for MR Image Segmentation", IEEE International Conference on Bioinformatics and Bioengineering , Oct. 2007, pp: 647-653,.

[33] Balafar MA et al, "New multi- scale medical image segmentation based on fuzzy c-mean (FCM)". In: IEEE conference on innovative technologies in intelligent systems and industrial applications, 2008, pp 66–70.

[34] Shen, S., Sandham, W., Granat, M., Sterr, A.: MRI Fuzzy Segmentation of Brain Tissue Using Neighbourhood Attraction with Neural-Network Optimization. IEEE Trans. Inform. Tech. Biomedicine. 2005, Vol. 9, no. 3. 459 – 467.

[35] Mitchell M; "MRI for diagnosis of breast cancer." International Society of Technology Assessment in Health Care. Meeting. Annu Meet Int Soc Technol Assess Health Care Int Soc Technol Assess Health Care Meet; 18: abstract no. 213, 2002.

[36] Petrou Maria, "Image Processing: The Fundamentals". , John Wiley.1999

[37] Janaki Sathya D, Geetha K, "A Novel Clustering Based Segmentation of Multispectral Magnetic Resonance Images",

International journal of advanced research in Computer science,2010, vol.1, no.4.

[38] I.N. Bankman, T. Nizialek, I. Simon, O.B Gatewood, I.N. Weinberg, W.R. Brody, " Segmentation algorithms for detecting microcalcifications in mammograms, " IEEE Trans. Inform. Techn. Biomed., 1993, Vol. 1, no.2, pp. 161-173.

[39] S.Bernsen, "Segmentation tools in mathematical morphology, "SPIE, Image algebra and morphology image processing, 1990, vol. 1350, pp. 70-84.

[40] J.S.Weszka,“A Survey of threshold selection techniques, Computer Graphics and Image Proc., 1978, vol. 7, pp. 259-265.

[41] Cancer facts and figures 2010.www.cancer.org/research/cancer facts figures/index.

[42] Franco Alberto Cardillo1 and Francesco Masulli, “Image Analysis Methods in MRI Examinations of the Breast” Technical Report: TR-09-16, September 2009.

[43] Janaki Sathya D, Geetha K, “Comparative study of Different Edge Enhancement Filters in Spatial Domain for Magnetic Resonance Images”. Selected for publication in AMSE Journal.

D. Janaki Sathya has received B.E (ECE) degree from Bharathiar University Coimbatore, Tamil Nadu, India in 2003. M.E (Applied Electronics) degree from Anna University Chennai, Tamil Nadu, India in 2008.she is currently pursuing Ph.D. from Karpagam University, Coimbatore. She has worked as a lecturer for more than 6 years in computer science and engineering department in P.S.G College of Technology, Coimbatore. She is currently working as design engineer in selka electronics, Coimbatore. She has 4 papers in national/ international journals / conferences to her credit. Her areas of interest include medical image processing, artificial intelligence, neural networks, swarm intelligence.

Dr. K. Geetha has received B.E (EEE) degree from Bharathiar University Coimbatore, Tamil Nadu, India in 1999. M.E (Applied Electronics) degree from Bharathiar University Coimbatore, Tamil Nadu, India in 2003. Received the Ph.D. degree in EEE from Anna University Chennai, Tamil Nadu, India in 2009. She has more than 12 years of teaching experience. She is currently working as Professor and HOD EEE department of Karpagam institute of technology, Coimbatore, India. She has been the chief editor of international journal of research and reviews in electrical and computer science engineering and editor of Elixir journal. She has more than 20 papers in national/international journals/ conferences to her credit.

Design of a new model of miniature antenna quasi-isotropic coverage

Abdellatif Berkat[1], Noureddine Boukli-Hacene[2] and F. Tarik Bendimerad[3]

[1] **Telecommunication Laboratory, Faculty of Technology, Abou-Bekr Belkaïd University Tlemcen, 13000, Algeria**

[2] **Telecommunication Laboratory, Faculty of Technology, Abou-Bekr Belkaïd University Tlemcen, 13000, Algeria**

[3] **Telecommunication Laboratory, Faculty of Technology, Abou-Bekr Belkaïd University Tlemcen, 13000, Algeria**

Abstract

In this paper we propose a new model for antenna design with a new geometry, in order to generate a quasi-isotropic radiation pattern in the far-field region. The main feature of the proposed antenna is the capability to generate a quasi-isotropic radiation pattern. The design details of the conceived antenna are presented and discussed. Simulations of the reflection coefficient and radiation pattern are presented. These were carried out using *CST Microwave Studio* .This antenna is easy to make and has got numerous applications in wireless systems, network sensors, field measurements and electromagnetic compatibility.

Keywords: *miniature antenna, quasi-isotropic, circular polarization, isotropic coverage.*

1. Introduction

Wireless sensor networks for home, industrial or environmental monitoring, personal body area networks, motion capture systems based on body sensors as well as satellite positioning devices are typical upcoming applications demanding for reliable wireless transmissions with constant link budget between devices, even if randomly oriented or quickly rotated with respect to each other. In such systems, various phenomena can deteriorate the transmission at the physical layer level such as an obstruction between devices, a multi-path fading or the presence of interferers [2]. These phenomena are particularly sensitive to the motion of devices or objects in the environment.

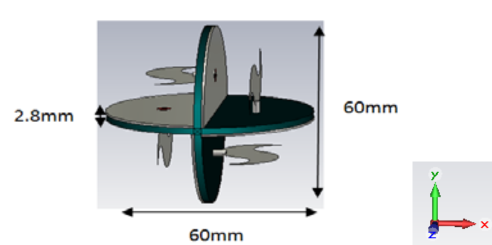
Concerning the impact of an arbitrary device orientation, the two main phenomena appear to be the anisotropy of the radiating pattern as well as the polarization mismatch between antennas. Directions of departure and arrival of a beam can change rapidly while in use and fall into the antenna radiating holes. Tilt between polarization states of antennas causes attenuation in the transmitted power. These effects can be greatly mitigated by a proper design of the radiation pattern properties of the antenna. Nowadays, new wireless applications come to market. They involve small telecommunication devices where orientation between emitter and receiver can change randomly while in use [2].

However, the transmitted signal is expected to be as stable as possible, whatever the orientation of the communicating objects is. For short distance, low cost, low data rate and low consumption applications, that is to say, when an adaptive solution cannot be envisaged, the most straightforward strategy is to search for an antenna radiating uniformly in all directions, knowing that an isotropic antenna doesn't exist [11].

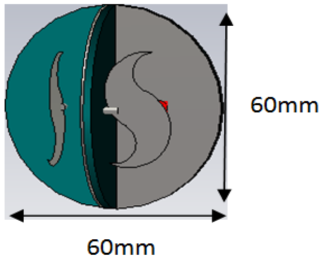
The antenna structure, the feeding network as well as the far-field pattern results are successively presented below.

2. Antenna configuration

The antenna structure is depicted in Fig1. Four patches are located along the sides of two intersected cylinders. patches are fed through a ground plane by a microstrip network, etched on the bottom side of the PCB.



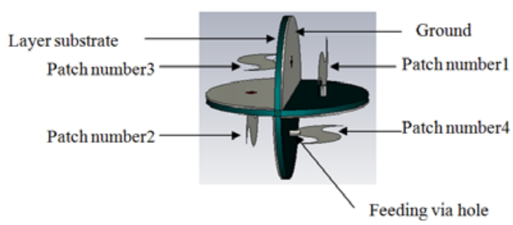
(a)



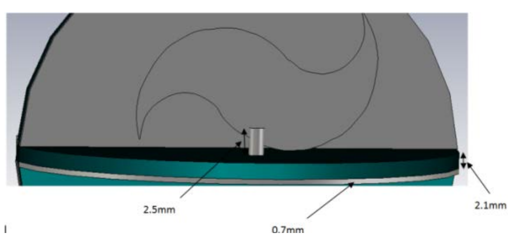
(b)

Fig1-(a) Horizontal view, (b) Vertical view of the new antenna model.

Fig 2 illustrates the antenna structure. The green top layer is made of a low permittivity and low-loss substrate, in order to optimize the antenna efficiency and bandwidth, where ε_r = 2.33, μ = 1 and thickness = 2.1 mm. A 0.7 mm thick copper layer is used as a ground plane for the antenna structure.



(a)



(b)

Fig 2 - (a) Structure of the antenna with four patches, (b) Layer and arrangement of the patch antenna

Fig 3 shows that the elementary patch is 32.64 mm long, 19.46 mm wide and are made of copper. It is fed via holes that are 1 mm across and 2.5 mm long. These holes are connected to the feeding network.

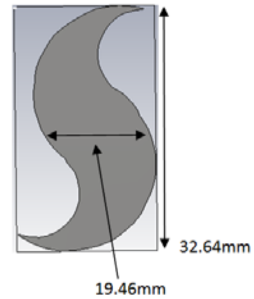


Fig3: Patch used

The four patches are fed with equal amplitudes. S2 and S3 are fed with the same phase of 90°. There is a phase difference of 90° between S1, (S2,S3) and S4. This feeding scheme leads to the targeted particular quasi-isotropic radiation pattern. It presents the advantage of greatly reducing the mutual coupling between patches.

Patch number	1	2	3	4
Amplitude relative to patch	1	1	1	1
Phase delay relative to patch	0°	90°	90°	180°

Table : Amplitude and phase constraints of the antenna.

3. Feeding network

The feedind network aims at feeding each Patch with the required amplitudes and phase. A microstrip network with two 90° hybrid couplers and one 180° hybrid coupler are located in the bottom side of the PCB. The network architecture is shown in Fig 4.

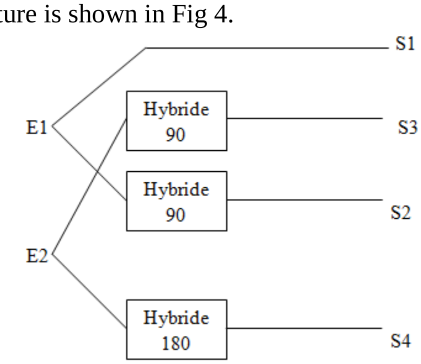


Fig4: Schematic of the feeding network [9]

The circuit layout, as illustrated in Fig 5, was designed using ISIS Proteus. The components are ultra small SMT. The input network is connected through two U-fl coaxial connectors (E1 and E2).

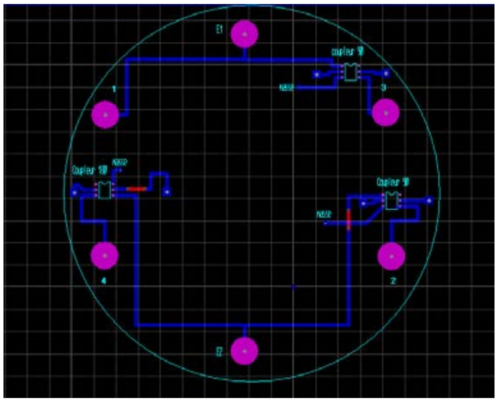


Fig 5: Layout of the microstrip network [14].

The antenna will be positioned in a vertical mode on the PCB.

4. Radiation properties of the antenna

In Fig 6, at the frequency of 3.8 GHz, a resonant mode and a good adaptation are observed. A peak appears at -35 dB .

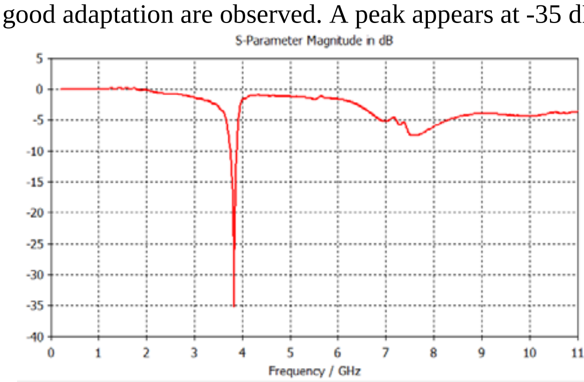


Fig 6: Computed return loss of the antenna

The main purpose of the antenna is to obtain a quasi-isotropic radiation pattern which allows the communication performances to be uniform between devices whatever their orientations are. The antenna radiation pattern is nearly isotropic. Fig 7, Fig 8, Fig 9 and Fig 10 exhibit the antenna radiation patterns at 3.8 GHz.

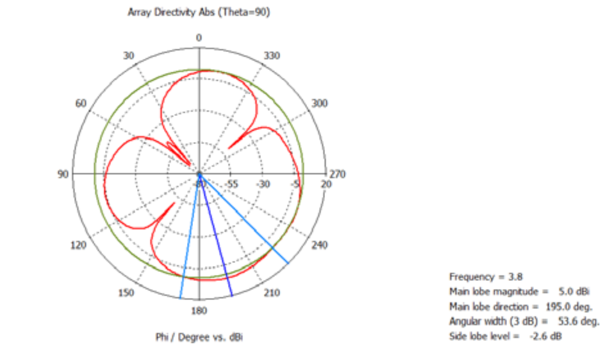


Fig7 : Polar diagrams (Theta=90°) at frequency = 3.8 Ghz

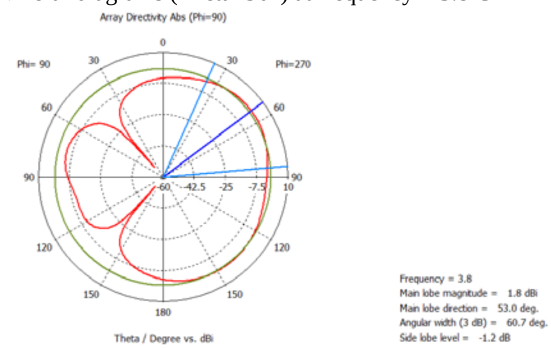


Fig8 : Polar diagrams (Phi=90°) at frequency = 3.8 Ghz

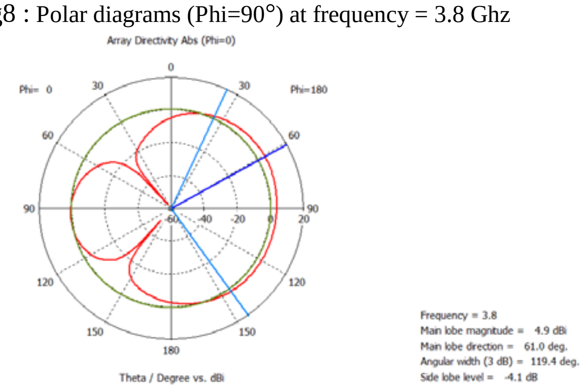


Fig9 : Polar diagrams (Phi=0°) at frequency = 3.8 Ghz

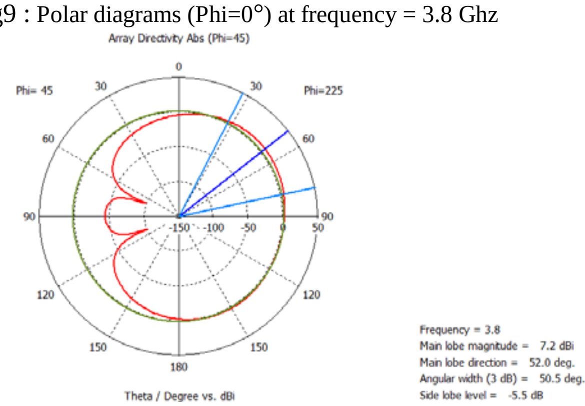


Fig10 : Polar diagrams (Phi=45°) at frequency = 3.8 Ghz

5. Conclusion

It was shown that this new miniature antenna that involves four novel patches presents a quasi-isotropic radiation pattern. The simulated results were conducted using the CST Microwave Studio. This new quasi-isotropic radiator has great applications in wireless systems and network sensors. Finally, it is also worth mentioning that similar quasi-isotropic antennas can be designed for other frequency bands, by a small change in the antenna geometry.

References

[1] E. da S. Pires(LEMA, Av. Aprígio Veloso, 882, UFCG/CCT/DEE, Campina Grande, PB, 58109-970, Brazil), G. Fontgalland(LEMA, Av. Aprígio Veloso, 882, UFCG/CCT/DEE, Campina Grande, PB, 58109-970, Brazil), R. M. do Valle(LEMA, Av. Aprígio Veloso, 882, UFCG/CCT/DEE, Campina Grande, PB, 58109-970, Brazil), G. F. Aragão(LEMA, Av. Aprígio Veloso, 882, UFCG/CCT/DEE, Campina Grande, PB, 58109-970, Brazil), W. R. N. Santos(LEIAM, Av. Aprígio Veloso, 882, UFCG/CCT/DEE, Campina Grande, PB, 58109-970, Brazil), T. P. Vuong(2LCIS/ESISAR-INPG, 50, rue Barthélémy de Laffemas, Valence, 26902, France) ,'' Proposal of a New Compact Isotropic Antenna'',IEEE 2006 .

[2] Huchard, M., Delaveaud, C.,Tedjini, S,''Miniature Antenna for Circularly Polarized Quasi Isotropic Coverage'', Antennas and Propagation, 2007. IEEE 2007 .

[3] Lev Pazin, Aleksey Dyskin, and Yehuda Leviatan,'' Quasi-Isotropic X-Band Inverted-F Antenna for Active RFID Tags'', IEEE ANTENNAS AND WIRELESS PROPAGATION LETTERS, VOL. 8, 2009 .

[4] M. HUCHARD, C. DELAVEAUD, S. TEDJINI, '' Critère d'évaluation de la couverture d'une antenne à partir de son champ lointain'' , 14èmes Journées Nationales Microondes 11-12-13 Mai 2005 Nantes .

[5] Sami HEBIB,''Nouvelle topologie d'antennes multi-bandes pour applications spatiales'', Délivré par l'Université Toulouse III - Paul Sabatier .

[6] Zhi Ning Chen, Kazuhiro Hirasawa(*Member, IEEE),* Kwok-Wa Leung,(*Member, IEEE),* and Kwai-Man Luk, (*Senior Member, IEEE),* ''A New Inverted F Antenna with a Ring Dielectric Resonator'', IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, VOL. 48, NO. 4, JULY 1999 .

[7] M. Fawzi Bendahmane, Mehadji Abri , F. Tarik Bendimerad and Noureddine Boukli-Hacene, '' A Simple Modified Transmission Line Model for Inset Fed Antenna Design'' , IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 5, September 2010 .

[8] Haim Matzner(Holon Academic Institute of Technology, Holon, Israel) , Kirk T. McDonald(Joseph Henry Laboratories, Princeton University, Princeton, NJ 08544) ,'' Isotropic Radiators'', April 8, 2003 .

[9] Mathieu HUCHARD ,'' Caractérisation et Conception d'Antennes Isotropes Miniatures pour Objets Communicants'', préparée au Laboratoire d'Electronique et de Technologie de l'Information du CEA Grenoble dans le cadre de l'Ecole Doctorale .

[10] Yue Gao, ''Characterisation of Multiple Antennas and Channel for Small Mobile Terminals'', Department of Electronic Engineering Queen Mary, University of London,United Kingdom,June 2007 .

[11] Mathieu Huchard(CEA-LETI, Grenoble, France) , Christophe Delaveaud(CEA-LETI, Grenoble, France) and Smail Tedjini(LCIS/ESISAR-INPG, Valence, France) , ''Characterization of the Coverage Uniformity of an Antenna based on its Far-Field'',IEEE 2005 .

[12] Kirk T. McDonald ,'' Radiation of Turnstile Antennas Above a Conducting Ground Plane'', Joseph Henry Laboratories, Princeton University, Princeton, NJ 08544 (September 18, 2008) .

[13] Huan-Chu Huang , Xiaojing Xu, and Yuanxun Ethan Wang ''Dual-Band Isotropic Radiation Patterns from a Printed Electrically Small Loop-Loaded Dipole Antenna'', Department of Electrical Engineering, University of California at Los Angeles, 405 Hilgard Ave., Los Angeles, CA90095-1594, U.S.

[14] Jean-Christophe MICHEL, ''www.gecif.net'' , ISIS Proteus ,July 2011 .

Abdellatif BERKAT was born in Algeria in 1987. He obtained his Master's Degree in Telecommunications, from Abou Bekr Belkaid University, Tlemcen, Algeria, in 2010. Abdellatif BERKAT is interested in the following topics: antenna design, algorithmic and programming theories, optimization algorithms, development of artificial intelligence methods. Abdellatif BERKAT is a doctorate student in the same university working on antenna design.

Noureddine Boukli-Hacene was born in 1959 in Tlemcen, Algeria. He received his Diplome d'Etudes Approfondies in microwave engineering (DEA Communications, Optiques et Microondes) and his Doctorate Degree in electrical engineering from Limoges University, France and from the National Center of Spatial Studies (Centre National d'Etudes Spatiales) in Toulouse, France, in 1982 and 1985 respectively. Recently, he was appointed as a lecturer at the University of Tlemcen. His research interests include, among others, microstrip antennas and microwave circuits.

F. Tarik Bendimerad was born in 1959 in Sidi Bel Abbès, Algeria. He received his Engineering Degree in 1983 from the Science and Technology University in Oran, Algeria and his Doctorate Degree from Nice Sophia Antiplois University, France, in 1989. He is a professor and the Director of the Telecommunications Laboratory. His main area of research is the microwave techniques and radiation and he is responsible of the antenna section.

Object-Oriented Software Methodologies: Roadmap to the Future

Usman Ali Khan[1], I. A. Al-Bidewi[2] and Kunal Gupta[3]

**[1] Department of IS, FCIT, King Abdul Aziz University, Jeddah
Kingdom of Saudi Arabia**

**[2] Department of IS, FCIT, King Abdul Aziz University, Jeddah
Kingdom of Saudi Arabia**

**[3] Amity Institute of Information Technology, Amity University, Lucknow,
Uttar Pradesh, India**

Abstract
Software Development Methodologies have survived a never ending evolving era, ever since it first came in horizon. Amongst the several methodologies, only Object–Oriented Methodology has been able to see the dawn of the day. Object–Oriented Methodology survived all the critics as well as the rapid changes in the software development industry. But is it braced for the future yet? The paper is oriented towards the existing Object–Oriented Software Development Methodologies. A brief discussion involving their origin and focus of the methodologies is given followed by a review of UML. We discuss the various approaches taken up by various methodologies. The discussion facilitates for the key notes for the survival of Object Oriented Methodologies.

Keywords: *Object Oriented Methodology, Software Crisis, UML, Integrated Methodology, Agile Methodologies.*

1. Introduction

The Software Crisis was identified four decades ago. Various methodologies and models came forward to address the issue, but almost all perished and became extinct as they were unable to transform according to the much necessary change. Object Oriented Methodology emerged as a revolution some two decades ago. It displayed its versatile nature and adopted to encounter the rapidly changing Software Industry. Object Oriented Methodology evolved from Semi- Structured, Partly Object oriented to the Unified Model to the Integrated and Agile Methodology. Its true that Object Oriented Methodology have survived the harsh wrath of change, but on the other hand, it has been unable to provide the complete solution for the Software Crisis.
Software Crisis is real and it still exists. One reason may be because the way people have treated the methodologies, to use them for their own purpose. Some treat them as a mean, others as ends. A majority have been treating them as products, which makes them easier to sell. But on the downside, if methodologies are treated as products, they seem to clutter. They start having advertisement like descriptions, obscuring wrappings, which is inefficient while explaining its underlying process.
A much better way of viewing the Methodologies would have been to view them with the perspective of process, rather than with the perspective of modeling languages, as it can provide help to the user of the methodologies by providing them more information with respect of their context. The description of such a methodology might include the details of activities performed in sub process, and the order in which they are performed; as well as the concise description of the underlying modeling language used in sub process definition.
The discussion that follows will make a comparative study of the object-oriented software methodologies, along with other methodologies and then provides a detailed overview of process pattern, and process meta models, and the future of next generation methodologies.

2. Methodologies – Framework for the Development of the Modern Era Software:

The Software Development Methodologies have been viewed as the means for organizing the various methods of software development in a timely and orderly execution manner. Informally, Software Development Methodology has been termed as a collection of phases, procedures, rules, techniques, tools, documentation, management, and training that can be utilized for the development of a system. It basically comprises of a set of modeling conventions, comprising of a modeling language, and a process, which can provide guidance as to the order of the activities, and offers criteria for monitoring and measuring a project’s activities. The modeling language aids in modeling the different aspects of the system, and the process determines what activities should be carried out in order to develop the system.

3. Object Oriented Methodologies: A brief insight

By 1990's, there were various methodologies in the Software Industry to design software products; but each and every one of them exhibited several limitations. In order to overcome these limitations, Object–oriented Methodologies were developed. Object–oriented methodologies for software development were specifically aimed at viewing, modeling and implementing the system as a collection of interacting objects, using the specialized modeling languages, activities and techniques needed to address the specific issues of the object-oriented paradigm. Several Object–oriented Software Development were developed in the evolution of the Object–oriented Paradigms. A brief overview of some of them is given below:

3.1 First and Second Generations of Object-Oriented Software Methodologies:

The first software development methodologies termed as object-oriented were in fact hybrid: partly structured and partly object-oriented. The analysis phase was typically done using structured analysis (SA) techniques, producing data flow diagrams, entity relationship diagrams, and state transition diagrams, whereas the design phase was mainly concerned with mapping analysis results to an object-oriented blueprint of the software. These methodologies were hence categorized as *transformative*. The second generation of object-oriented methodologies evolved from the first generation and appeared between 1992 and 1996. First and second-generation methodologies are collectively referred to as *seminal* methodologies, in that they pioneered the unexplored field of pure object-oriented analysis and design, and in doing so laid the groundwork for further evolution.

3.2 The Unified Modeling Language (UML):

Grady Booch, Ivar Jacobson, and James Rumbaugh. All three had developed their own methods, but collaborated to combine them into the Unified Method, the OMG announced plans for a standard OO notation, and in June of 1996 UML version 0.9 was released. UML version 1.1 was adopted by the OMG in November of 1997. With these initial UML releases, dozens of competing methodologies were replaced by the language- and method independent UML. Several factors contributed greatly to the widespread adoption of UML. First, UML is language independent. Second, it does not advocate nor require a particular method. Third, it is readily accessible as UML specifications are free for download and any company may join the OMG. The Object Management Group (OMG) is the body responsible for creating and maintaining the language specifications. They define UML as, "a graphical language for visualizing, specifying, constructing, and documenting the artifacts of object oriented software systems. The UML represents a collection of best engineering practices that have proven successful in the modeling of large and complex systems. The UML uses mostly graphical notations to express the design of software projects. Using the UML helps project teams communicate, explore potential designs, and validate the architectural design of the software. The primary goals in the design of the UML were:

- To provide users with a ready-to-use, expressive visual modeling language so they can develop and exchange meaningful models. Provide extensibility and specialization mechanisms to extend the core concepts.
- Be independent of particular programming languages and development processes.
- Provide a formal basis for understanding the modeling language. Encourage the growth of the Object Oriented tools market.
- Support higher-level development concepts such as collaborations, frameworks, patterns and components.

UML was developed as a language that can be utilized for Modeling Object Oriented Systems and Applications, and provide them with more clarity by making them readable, and thus more understandable. This essentially means that UML provides the ability to capture the characteristics of a system by using notions. UML provides a wide array of simple, easy to understand notions for documenting systems based on the Object Oriented Design Principles. These notions are called the diagrams of UML. These diagrams provide the user with the means of visualizing and manipulating model elements. The underlying premise of UML is that no one diagram can capture the different elements of a system in it's entirely. The UML is just that. It "unifies" the design principles of each of these methodologies into a single standard language that can be easily applied across the board for all Object Oriented Systems. UML does not have any dependencies with respect to any technologies or languages.

3.3 Integrated Methodologies: Third Generation:

Methodologies in this category are results of integrating seminal methodologies and are characterized by their process-centered attitude towards software development, typically targeting a vast variety of applications. Integrations have resulted in huge methodologies, difficult to manage and enact. In trying to achieve manageability, some of them have gone to extreme measures to ensure customizability (RUP), others have turned into generic process frameworks that should be instantiated to yield a process (OPEN), and yet others have resorted to process patterns for customizability (Catalysis). It was frustration with these methodologies that ultimately caused the agile movement. Although unwieldy and complex, integrated methodologies have a lot to offer in terms of process components, patterns, and management and measurement issues. Furthermore, some of them propose useful ideas on seamless development, complexity management, and modeling approaches.

3.4 Agile Methodologies:

Agile methodologies first appeared in 1995. The once common perception that agile methodologies are nothing but controlled code-and-fix approaches, with little or no sign of a clear-cut process, is only true of a small—albeit influential—minority of these methodologies, which are essentially based on practices of program design, coding, and testing that are believed to enhance software development flexibility and productivity. Most agile methodologies incorporate explicit processes, although striving to keep them as lightweight as possible.

4. Object oriented Methodologies: The big leap on to the next generation

Object Oriented Methodologies have come a long way, and is still standing tall. In order to take big strides in the future, Object oriented Methodologies should incorporate the following within their structure:

- The advocates of Object oriented Methodology assumes that software should be developed according to a mental model of the actual or imagined objects it represent, i.e. it should focus on the real world. It should however provide for the uninterrupted exposition of logic in a more easy representation, more of a ordinary human language. This would lead to even poorly thought-out designs and decisions to be more subtle and obvious. The resultant model would bridge the gap between intuitive and formal models.
- Object-Oriented Methodology still provides a way or method of solving a problem. If it is to stay for longer duration, it needs to evolve itself as a technology which can address various issues in the modern era software development.
- In several cases, Object-Oriented methodologies have proved to be less fruitful as compared to Procedural Language. Amongst them are Economy of Execution, Economy of Small Scale Development, and Economy of Compilation. Object Oriented Methodology should scale themselves up for such shortcomings. A Significant difference in productivity between OOP and procedural development has to be achieved in the next era of Object Orientation.
- Object Oriented Technology has been poor in modeling time in a coherent manner. With the growing advent of real-time systems, it is imperative that Object-Orientation should evolve to model time and real time objects in a more decisive manner and thus aiding in the accurate design of real time systems.
- By its very nature, Object Orientation is anti-modular and anti-parallel. Which in the current scenario is a serious issue. With the growing popularity of parallel systems, it is of essence that Object Orientation should incorporate features which support parallel computing and applications.
- Currently, Object Orientation is unable to provide for interface specifications that are rich enough to cover all the phases of the design cycle. In component based development especially, non functional characteristics can be incorporated as a part of interface specification to overcome this limitation.
- Object Orientation provides for standardizing notations, which is often not sufficient to achieve effective methods and unambiguous communication amongst designers. With the advent of Opinion mining and profiling, these modeling notations maybe freely reinterpreted, which in turn weakens the value of notation as an effective communication vehicle and designing tool. This implies that a model written in one formalism could be ill formed in another formalism.
- It becomes difficult for Object Orientation to combine Heterogeneous system and depict their composite behavior. Object oriented methodology has to provide for embedding the

detailed models in question into a framework that can understand the model being composed.

- Object oriented should target towards being a model driven, context management, aspect oriented, service driven architecture. And should take the role of whichever is required at respective place.
- Although Object oriented Storage technology is being used in Cloud Computing, but the methodology is not sufficient for providing an impact structure for Web enabled Context aware systems and services. The main issue comes in the handling and distribution of context information efficiently. Furthermore, challenges like aggregation of context information in a structured format, discovery and selection of appropriate context services are key area, where Object oriented paradigm has to evolve and adopt in order to provide for a better framework and design for such system.

5. Conclusion:

The above review has resulted in a number of conclusions, which can be elaborated as follows. UML was developed in an attempt to standardize and integrate the methodologies into a single, comprehensible unit. Yet some of the limitations were still at large. The complexity and inconsistency was still there, which gave rise for the development of some agile and lightweight methodologies which actually followed a different path from modeling. But even they were not entirely successful and we witnessed the comeback of the old methodologies as well as new developments in methodologies emerged which did not adhere to UML conventions. The evolution suggested that in order to develop new methodologies and technologies, not only the capabilities of the old methodologies should be considered; but also the fact that they have to be developed with a more systematic approach in mind. Despite of the entire enhancement in the development of methodologies, a number of problem areas have been observed. The new integrated methodologies are more complex, to be efficiently being brought into the practice. They have lack of scalability, and lack of a specific, unambiguous process. Object Oriented Methodologies have evolved over a period of time, and despite of all the limitations they are still considered to be the pioneer when it comes to software development. Ongoing researches are aimed to further bring around an improved version, which can provide for compactness, extensibility, consistency, visible rationality, and traceability to requirement. Considering the motivations and the special circumstances surrounding methodologies mergers and development, engineering a methodology through integration can be one of the most appealing one. Disciplined Engineering and a systematic approach is desired for the extraction of prosperous potential of Object-Oriented Software Development.

References

[1] Aoyama, M. (1998a, April 19 - 25). Agile software process and its experience. Paper presented at the International Conference on Software Engineering, Kyoto Japan.

[2] Aoyama, M. (1998b). Web-based agile software development. IEEE Software, 15(6), 56-65.

[3] Bloomberg, J. (1999, October). Software Methodologies on Internet Time. Developer.com. Retrieved March 11, 2001, from the World Wide Web : http:// softwaredev.earthweb.com/java/sdjjavaee/article/0, ,12396_616711,00.html.

[4] Bloomberg, J. (2001, January). Using the RUP for Enterprise e-business Transformation. WaveBend Solutions. Retrieved April 7, 2001, from the World Wide Web: http://www.therationaledge.com/content/jan_01/f_rupent_jb.html.

[5] Coad, P. (1999). Feature-Driven Development. Object International. Retrieved April 8, 2001, from the World Wide Web: http://www.togethersoft.com/jmcu/ chapter6.PDF.

[6] Cockburn, A. (2000, July/August). Selecting a project's methodology. IEEE Software, 17(4), 64-71.

[7] Cockburn, A. (2000, September). Balancing Lightness with Sufficiency. American Programmer. Retrieved March 11, 2001, from the World Wide Web:http://www.crystalmethodologies.org/articles/blws/balancinglightnesswithsufficiency.html.

[8] Cockburn, A. (2001). Just-In-Time Methodology Construction. Humans and Technology. Retrieved March 11, 2001, from the World Wide Web:http://www.crystalmethodologies.org/articles/jmc/justintimemethodologyconstruction.html.

[9] Cutter. (2000, October). Light Methodologies Best for E-business Projects. Cutter Consortium. Retrieved March 11, 2001, from the World Wide Web:http://cutter.com/consortium/research/2000/crb001003.html.

[10] DOJ (2000, March). The Department of Justice Systems Development Life Cycle Guidance Document. United States Department of Justice. Retrieved April 1, 2001, from the World Wide Web: http://www.usdoj.gov/jmd/irm/lifecycle/table.htm.

[11] LexiBot (2001). Our Technology - Results: The LexiBot Expression. BrightPlanet. Retrieved April 1, 2001, from the World Wide Web:http://www.brightplanet.com/technology/results2.asp.
[12] Lindvall, M., & Rus, I. (2000, July/August). Process diversity in software development. IEEE Software, 17(4), 14-18.
[13] Ubiquitous Cloud: Managing Service Resources for Adaptive Ubiquitous Computing
Koichi Egami, Shinsuke Matsumoto and Masahide Nakamura
[14] 1st IEEE PerCom Workshop on Pervasive Communities and Service Clouds
[15] Component-Based Design for the Future
Edward A. Lee and Alberto L. Sangiovanni-Vincentelli
IEEE

Dr. Usman Ali Khan is an Associate Professor at IS Department, Faculty of Computing and Information Technology, King Abdul Aziz University, Jeddah, Saudi Arabia. He has done his Ph.D. IT (Object Oriented Software Design) in 2007. He is in the field of teaching and research since 1995. He has published fourteen research papers at national and international forums. He is teaching Software Engineering domain courses at graduate and undergraduate level from more than thirteen years.

Dr. Ibrahim A. Al-Bedewi is a Dean and Associate Professor at IS Department, Faculty of Computing and Information Technology, King Abdul Aziz University, Jeddah, Saudi Arabia. He has done his Ph.D. CS in 2000. He is in the field of teaching, research and administration since 2000. He has published many research papers at national and international forums.

Kunal Gupta is a Lecturer in Department of Information Technology, Amity University, India. He has done his M.Tech(IT) and MCA from India and is currently pursuing his PhD. He has published four International Papers. His teaching domain is Software Programming Languages. He has been in academics for more than seven years.

Dynamic unlicensed bands access based on the decentralized resource distribution approach.

Souleymane Oumtanaga[1], Zamble Raoul[2], Michel Babri[3]

[1]Institut National Polytechnique Houphouet Boigny INPHB-LARIT
08 BP 08 BP 475 Abidjan 08, Côte d'Ivoire.

[2]Institut National Polytechnique Houphouet Boigny INPHB-LARIT
08 BP 08 BP 475 Abidjan 08, Côte d'Ivoire.

[3]Institut National Polytechnique Houphouet BoignyINPHB-LARIT
08 BP 08 BP 475 Abidjan 08, Côte d'Ivoire.

.

Abstract

Dynamic spectrum access is a promising technique to use spectrum efficiently. Without being restricted to any prefixed spectrum bands, nodes choose operating spectrum on demand. Such flexibility, however, makes efficient and fair spectrum access in large scale networks great challenge. This paper presents a dynamic spectrum access decentralized approach. This approach is base on the game theory, mainly on the principle of Homo Egualis. There, it is assumed that operators are averse to unequal payoffs and act unselfishly, enabling a stable and sustainable community. An algorithm is proposed to solve the problems of inefficient use of spectrum and fair dynamic access to the available resources in the unlicensed bands. Numerical results show the performance of this algorithm.

Key words : *Dynamic spectrum access, decentralized approach, Homo Egualis principle*.

1. Introduction

The frequency spectrum is a natural resource very much in demand in our society. However, this resource is limited. Today, radio communication applications are increasingly various and radio- electric frequencies consuming. Recently, the telecommunication industry kept on increasing and innovating in term of wireless technologies, using more and more spectral resources. In view of these new requirements, it has been necessary to establish efficient systems for the spectrum management in order to allocate efficiently the resources while being careful to the interference risks due to the proliferation of actors. It is also widely recognized that wireless systems of digital communication don't operate on the entire available frequency band. The coming wireless systems then will be compelled to make the most of such free frequency bands, thanks to their ability to listen and adapt to their environment. Such ability comes within the concept of "cognitive radio" introduced by Joseph Mitola in 2000 [1]. This concept paves the way to an innovating approach which enables a dynamic and opportunistic but controlled use of the radio electric spectrum in support to the current statistic approach [11]. One of the main aspects of cognitive radio is the software defined radio (SDR) which promises a great flexibility by allowing a single device to access a wide range and various technologies [1]. The concept of "cognitive radio "intends to use the potential of software defined radio to allow more efficient use of spectrum. Cognitive radio devices scan the frequency spectrum dynamically by accessing the portions of the spectrum not used by the primary systems. Access is facilitated by the ability of cognitive radio nodes to adapt the waveform to the technical specifications set by the regulatory authorities.

However, access to radio resources for a secondary use creates other problems such as interference issues and fairness of access between different radio systems involved. If for the interference problem the regulator has found a solution by adopting the limited transmission power, the fact remains that the issue of fairness of access between different radio systems remains unsolved. For that purpose, several approaches have been developed and could be classified according to the criterion of priority access to the spectrum or according to the criterion of architecture. First, following the criterion of the priority access to the spectrum, two approaches can be distinguished: the approach of the vertical sharing of the spectrum where the spectrum is shared at several levels, this is the example of the "spectrum pooling approach" [4] and the approach of horizontal sharing where the spectrum is shared with equal priority as in the case of wireless local areas network (WLAN) [5, 6]. Moreover, under the test architecture, we distinguish the approaches of centralized and decentralized sharing of the spectrum. In the centralized sharing approach, there is a central entity which controls the benefit of the spectrum access procedures. Each entity in the system sends it its status such as its traffic demand in the spectral and temporal domains. So the central entity is responsible for allocating the available spectrum to the other entities by taking into account the aspect of the spectrum efficiency and fairness among the different entities involved. As for the decentralized spectrum sharing approach, "Carrier Sense Multiple Collision Avoidance (CSMA/ CA)" protocols and game theory based on distributed decision making protocols are studied in [7,8]. In these studies, based on a local or global policy, each entity is responsible for the spectrum access and its use. The aim of these studies is to find a minimum policy sets for flexible, scalable, and sustainable for the spectrum sharing dependently from the growth of market demand and amount of traffic. This paper is exclusively focused on the decentralized sharing approach of the available radio spectrum by the high lighting of a method of inter- system cooperation based on the access scheme resulting from the principle of Homo Egualis society and by the suggestion of an algorithm allowing the implementation of this access scheme from the perspective of reaching the spectrum efficiency and the inter-system fairness.

The rest of the paper is organized as follows. In section 2, problem formulation is presented. Section 3 gives the system model, while section 4 tackles the numerical results. Finally the article ends with a conclusion in section 5.

1. Problem formulation

Two groups of networks or operators are considered in the 5 GHz unlicensed band; the primary networks or operators and the secondary or cognitive radio networks. The primary networks have an exclusive access to their dedicated spectral bands; while the cognitive radio operators or networks access to the spectrum only when this one is not used by the primary networks. Here, the primary networks choose the radio-location systems such as radar system and the secondary system networks can be represented by wireless operators whose transmission systems are based on OFDM technology (Orthogonal Frequency Division Multiplexing). To allow an opportunistic access to radio resources of primary system, we assume that the secondary networks also have agile spectrum. So the radio cognitive nodes are responsible for localizing the available resources in both spectral and temporal domains [2],[3]. Indeed we precise that the primary networks do not cooperate with other systems. We have m primary networks with m channels; each network having a dedicated channel and exclusively reach it. We also admit the existence of n operators or cognitive radio networks that access to the spectrum only when this one is not used by the primary networks. Depending on a primary network's spectrum usage pattern, the duration of a spectral opportunity can exceed hours, even days in spectral bands reserved for emergencies; or can be only few milliseconds in heavily-used spectral bands. It will be relatively easy, for the secondary network to use long-lasting opportunities. However, for short-lasting spectral opportunities, a secondary network may not be able to detect their existence and then utilize them before "expire". Therefore, our study will rule out the short-lasting spectral opportunities.

In the perspective of making the most of these spectral opportunities, a node belonging to a cognitive radio network scans first the spectrum, either periodically or randomly to discover and use of idle spectrum portions. When it will detect all the activities of the primary network, it will automatically release the channels used with the purpose of avoiding interferences. All nodes in a network of cognitive radio operator use the same spectral opportunities to maintain their inter-connectivity at all time. Therefore, the different nodes must also keep the same information on the spectral and temporal environment, so as to make the decision to transfer their traffics on the selected spectral opportunities.

The spectrum is divided into "channels" which represents small units of spectral bands. Let suppose that each cognitive radio node uses a single channel for its basic communication, but has the ability to use several adjacent channels, simultaneously available for a better quality of transmission. The software defined radio will rightly adopt the modulation schemes required for the use of a wide bandwidth. In addition, the secondary network will use these unassuming adjacent channels, as sub-carriers of a multi-carrier modulation scheme such as OFDMA. The temporal use of each channel by the primary network can be characterized by a random process. Any time the primary network will not use its dedicated frequency band, it will leave some idle channels which will be exploited by the cognitive radio networks.

In this paper, we suppose that the different spectral channels are perfect. That is to say that the channel is either busy or idle. Cognitive radio nodes of different operators are trying to access the idle channels left by the primary network nodes to satisfy their communication needs. We consider simultaneously a set of k idle channels among m channels from the spectrum, available for secondary use. In our case, n secondary networks are trying to exploit the k available spectral opportunities, with $n > 1$. Of course, if $n \le k$, each cognitive radio network will get a channel. Otherwise, these n networks compete for the access to the different channels. It is then that appear other parameters which will enable an efficient and fair access to the available spectrum; knowing that we have kept the decentralized approach for the sharing of these idle channels. In what follows, we will simply refer to secondary networks or cognitive radio by the term operators.

2. System model

2.1. Homo Egualis society scheme

For the decentralized cooperative approach it is important to design policy rules which can control the compromise between the fairness among operators and the spectrum efficiency. For instance, in [10] a punishment policy against selfish users is suggested in the case of CSMA/CA networks, which means that basically all potential operators for a secondary use of the spectrum are selfish. The basic mechanism of this punishment policy must systematically jam the packets from deviating selfish users. This mechanism then reallocates a portion of the shared spectrum with a view to signal transmission, which leads to the deterioration of the spectrum efficiency. On the other hand, another policy based on the altruistic character of the operator is suggested in [9] and known as the “inequality aversion model”. In the model, each operator is characterized by its “payoff” which represents its gain. This model is based on the aversion to inequality of payoffs of operators and, was originally established by studies of social anthropology. By using this policy, operators act as altruists and their unselfish behavior allow their community to be stable and sustainable. We focus on this model to examine decentralized available spectrum sharing among operators. We adopt this because it is natural to assume that all the users want a sustainable environment for radio communication. The utility function based on aversion to inequality can be modeled as follows :

$$u_i = \frac{x_i}{A_i} + \frac{\alpha_i}{n-1} \times \sum_{\frac{x_j}{A_j} > \frac{x_i}{A_i}} \left(\frac{x_j}{A_j} - \frac{x_i}{A_i} \right) - \frac{\beta_i}{n-1} \times \sum_{\frac{x_j}{A_j} < \frac{x_i}{A_i}} \left(\frac{x_i}{A_i} - \frac{x_j}{A_j} \right)$$

where u_i is the utility of the i^{th} operator, and x_i and x_j ($i \neq$j) respectively indicate the payoffs of the i^{th} and j^{th} operators. Payoffs are numbers which represent the motivations of the operators such as their profit, quantity, or other continuous measures. In this paper, the payoff shows the amount of spectrum used for their signal transmission. Term n represents the number of operators sharing the spectrum. A_i is the priority level of the i^{th} operator among all the operators for the payoff. When some operators have priorities over other operators in (1) are reduced in accordance with their priority level. In addition, α_i and β_i show the reaction factor of the i^{th} operator respectively against those which receives a higher payoffs and against the operators which receive lower payoffs. Knowing that the utility functions express the satisfaction level of operators, each operator adopts a behavior allowing it to maximize its own utility function independently. Based on an anthropological study [9], it is also shown that a sustainable community where each operator receives the same payoff can be established by the setting $\alpha_i > \beta_i$. This parameter setting model of the operator’ preference when his payoff is less than that of the other operator is different from that when his payoff is more than that of the other operators.
Here, we assume that each operator can receive information about the payoff value and the priority level of the others through a backbone network to which all the operators are connected. It is also possible for each operator to measure individually its payoff by monitoring all transmitted signals from users and by detecting to which operators the users belong by using signal headers which contains their affiliation identities.

2.2. Application to wireless communication systems

In this section, we describe how cognitive radio nodes access the channel and how do they collect information in order to avoid collisions and signal interferences between the various stakeholders of the system. Since the channels are in a perfect condition, signal loss occurs only when there is a collision with the primary users or equivalently when the state of the channel is busy during the signal transmission time of an operator.

The operator should then use each channel appropriately by transmitting its own signals between the busy states. In addition, we consider that cognitive radio system go through the simple principle of the protocol "Listen Before Talking (LBT)" to access the channel. In this protocol, first the node senses the selected channel to check if it is idle or busy. In practice, this can be done by the detection of energy. The sensing of the channel is only an option when the cognitive radio nodes have knowledge of the characteristics of the physical layers signals from primary users. When a cognitive radio node finished its communication, it automatically releases the radio resources used.

We also make the assumption of perfect coordination between cognitive radio nodes. In other words, if the channel is occupied by pair of transmitter and receiver, all other cognitive radio nodes in the area are aware of it so as to avoid a collision between their signals. The protocol access to the spectrum is not persistent, which means that if the channel is found busy in the sensing, the transmission cycle ends and statistic data concerning the occupation of the channel are recorded, and the node tries other attempts to use another channel. Otherwise, the node transmits its signals. Thus, from the scheme of society Homo Egualis, we define the transmit probability. This probability for each operator i at time t is given by :

$$\Delta P_i(t) = \frac{\alpha_i}{n-1} \times \sum_{\frac{x_j}{A_j} > \frac{x_i}{A_i}} \left(\frac{x_j}{A_j} - \frac{x_i}{A_i}\right) - \frac{\beta_i}{n-1} \times \sum_{x_j \, x_i} \left(\frac{x_i}{A_i}\frac{x_j}{A_j}\right) \quad (3)$$

w n
f e
i^{th} operator, and L_i is the traffic demand for the i^{th} operator.

By controlling the probability of access of each operator based on equation 2, the rules laid down on the access to the shared spectrum may reflect the behavior of each operator.

2.3. Proposed algorithm

We have below, the proposed algorithm for the transmit probability for decentralized dynamic access to the available spectral resources according to Homo Egualis model :

Function PROBA_ACCESS_HOMO_EUGALIS

```
BEGIN
    ARRAY Alpha (Number_operator2) into REAL
    ARRAY Beta (Number_operator2) into REAL
    FOR a ← 1 to Number_operator2
        READ Alpha (a)
        READ Beta (a)
    END FOR
    ARRAY Xi (Number_operator1) into REAL
ARRAY Pi (Number_operator1) into REAL
    FOR i ← 1 to Number_operator1
        [vi, xi] ← CALL PAYOFF FUNCTION
        Pix ← 1
        lxi ← LEN (xi)
        READ Xi(Xi,xi)
        h ←1
        som ← 0
        WHILE   Xi(i,h)  <  Xi(i,lxi)
            som ← som + (Xi(i,lxi) -
    Xi(i,h))/(Xi(i,lxi))
Palphai ← Alpha(1,i)/
Number_operator1)*som
        h ← h+1
        END WHILE
        WHILE   Xi(i,h)  >  Xi(i,lxi)
        som ← som + (Xi(i,h) -
        Xi(i,lxi))/(Xi(i,h))
            Pbetai ← Beta (1,i)/
Number_operator1)*som
        h ← h+1
        END WHILE
        Pi = MAXIMUM (0, (MINIMUM(1,
(Pix + Palphai - Pbetai)))
```

RETURN Pi

END FOR

FIN

We shall proceed to the evaluation of this algorithm in the following section.

3. Numeric results

For the evaluation of our algorithm giving the transmit probability for a secondary usage of the available spectrum among several radio cognitive operators, we consider that each channel alternates between the busy state and the idle state.The durations of busy and idle states are given by random distributions with unknown mean. We analyze two scenarios where the number of primary users is 12 ($m = 12$).The main reason for choosing this number is that there are 12 (non-overlapping) channels in the 5 GHz band for the IEEE 802.11a. In both scenarios, we consider the same traffic demand for all operators; with $\alpha = 1$ and $\beta = 0.01$.General parameters such as the number of idle radio resources, the number of cognitive radio operators, are given for each scenario. The reduction factor W_i is set to 1 for all operators.

Scenario 1 : The number of operators is lower or equal to the number of idle channels ($n = 5$ and $k = 6$).

In this scenario, there are five operators who wish to reach the frequency spectrum with 6 idle channels. The figure 1-a shows the payoffs of these five operators.perator number 3 presents the highest payoff, while operator number 4 has the lowest payoff.This situation not influences the spectrum access of the whole of these operators.Here, the number of radio cognitive operators being lower than the number of idle channels, as announced in the problem formulation, we observe very well that each operator reaches a channel.t justifies the transmit probability value of all the operators which is 1 as shown in the figure 1-b.

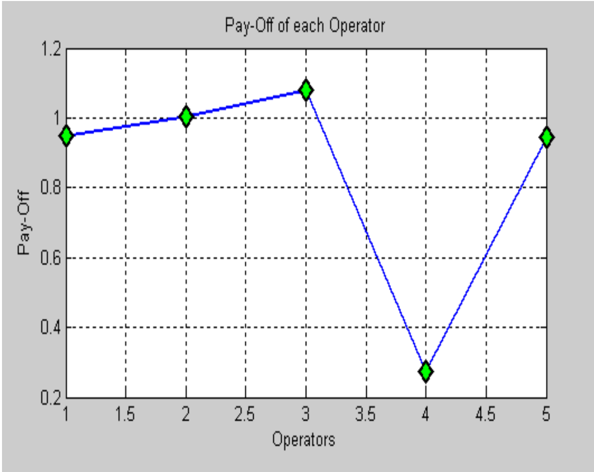


Figure 1-a : Payoffs of different operators for scenario 1

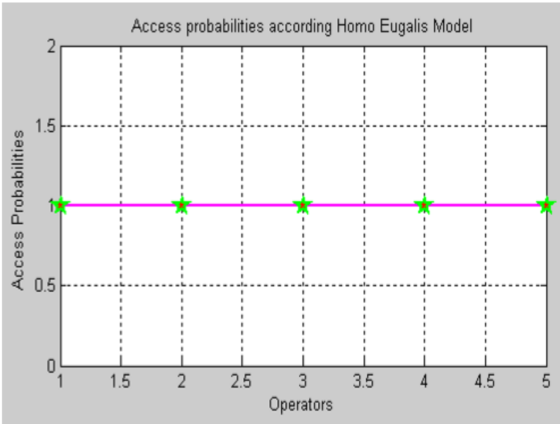


Figure 1-b : Transmit probability of different operators for scenario 1

Scenario 2 : The number of operators is higher than the number of idle channels ($n = 6$ and $k = 5$).

In this second scenario, there are 6 operators who wish to reach the frequency spectrum with 5 idle channels.The figure 2-b shows the payoffs of these 5 operators. The operator number 1 presents the highest payoff, while the operator number 5 has the lowest payoff.The number of cognitive radio operators being higher than the number of idle channels, the access to the idle channels will be given by the various values of the transmit probability of the operators. As we can see on the figure 2-b, only, the operator number 6 does not reach the spectrum.On this figure, its transmit probability is equal to zero while knowing that it does not have the lowest payoff among the six operators. That is due to the high levels of priority from the operator number 2, 3 and 5, compared to that of the operator number 6 whose payoff is the largest among the four operators.

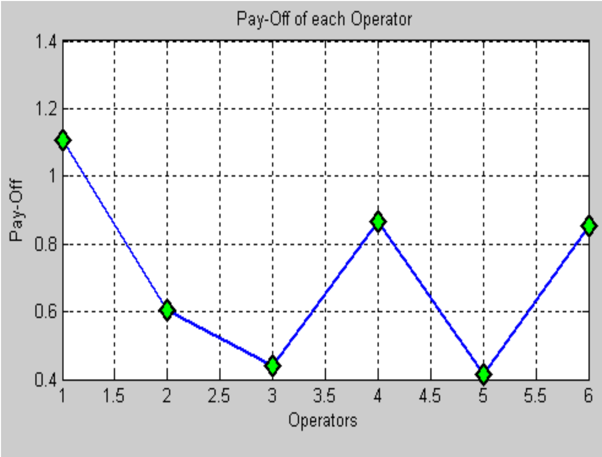


Figure 2-a : Payoffs of different operators for scenario 2

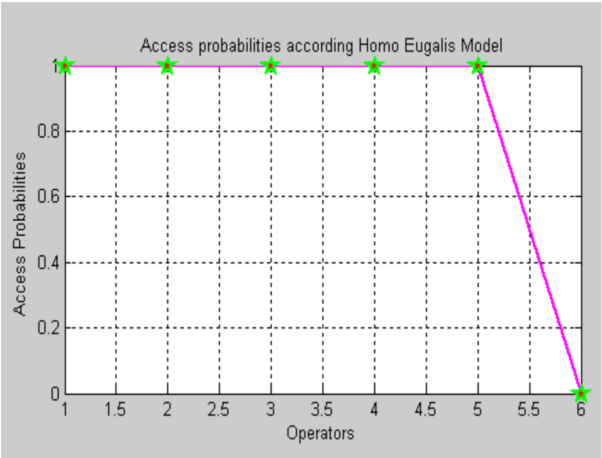


Figure 2-b : Transmit probability of different operators for scenario 2

Then, we have also to carry out the changes the number of operators and the idle channels number while remaining within the framework of this scenario and we ended to the same remark.

4. Conclusion

In this article, we exposed the model of dynamic access to the spectrum based on the principle of the game theory. In this particular case, we used Homo Egualis model. This model allows a decentralized and fair sharing of the available radio resources between the various operators. Besides, we proposed an algorithm for implementation of this dynamic spectrum access scheme. We also evaluated our algorithm through two scenarios which show the resolution of spectrum under use problem by giving additional opportunities to the operators and the resolution of fairness problem among different operators

REFERENCES

[1] J. Mitola, "The software radio architecture" IEEE Communications, Vol. 33, N°5, 1995, pp. 26-38.

[2] C.T. Chou, Hyoil, K., Sai Shankar, N., and Shin, K. G., "Spectrum agility radio for broadband applications" *EETimes In Focus Article,* Aug. 23.

[3] C.T. Chou, Hyoil, K., Sai Shankar, N., and Shin, K. G., "What and how much to gain from Spectrum Agility" *Submitted to IEEE/ACM Trans. On Networking*, June 2004

[4] T. A. Weiss and F. K. Jondral, " Spectrum Pooling: an innovative strategy for the enhancement of spectrum efficiency," *IEEE Communications Magazine*, vol. 42, no. 3, pp. S8–14, 2004.

[5] M.Wen and L. Hanwen, "Radar detection for 802.11a systems in 5 GHz band," in *Proceedings of the IEEE International Conference on Wireless Communications, Networking and Mobile Computing (WiMob '05)*, vol. 1, pp. 512–514, Montreal, Canada, September 2005.

[6] C. Cordeiro, K. Challapali, D. Birru, and N. Sai Shankar, "IEEE 802.22: the first worldwide wireless standard based on cognitive radios," in *Proceedings of the 1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '05)*, pp. 328–

337, Baltimore, Md, USA, November 2005.

[7] Z. Ji and K. J. R. Liu, “Cognitive radios for dynamic spectrum access—dynamic spectrum sharing: a game theoretical overview,” *IEEE Communications Magazine*, vol. 45, no. 5, pp. 88–94, 2007.

[8] A. B. MacKenzie and L. DaSilva, *Game Theory for Wireless Engineers*, Morgan & Claypool, San Rafael, Calif, USA, 2006.

[9] H. Gintis, *Game Theory Evolving: A Problem-Centered Introduction to Modeling Strategic Interaction*, Princeton University Press, Princeton, NJ, USA, 2000.

[10] M. Cagalj, S. Ganeriwal, I. Aad, and J.-P. Hubaux, “On selfish behavior in CSMA/CA networks,” in *Proceedings 24th IEEE Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM ’05)*, vol. 4, pp. 2513–2524, Miami, Fla, USA, March 2005.

[11] Y. Xing, R. Chandramouli, S. Mangold, and N. Sai Shankar, “Dynamic spectrum access in open spectrum wireless networks,” *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 3, pp. 626–637, 2006.

Souleymane Oumtanaga received the PhD degree in Computer Science from University Paul Sabatier of Toulouse, France in 1995. During 1999-2000, he stayed in LAboratoire de echerche en Informatique et Mathématiques Appliquées (LARIMA) at INPHB (Institut National Polytechnique Houphouet Boigny), Côte d’Ivoire. Since 2000 he has been Head of Centre de Formation des Technologies de l’Information et de la Communication (CFTIC) of INPHB. Since 1990 he has also been the Head of the Network Information Center (NIC) of Côte d’Ivoire. He has been Professor in computer science since 2007 and he currently manages the Laboratoire de Recherche en Informatique et Télécoms (LARIT) at INPHB. His research interests include IP mobility, IP Network security, IPv6, Wireless Network, Mobile Networks.

Zamblé Raoul is currently a PhD student in Computer Science at University of Cheikh Anta Diop of Dakar, Senegal. He received his BS in computer Science from University of Cocody (Côte d’Ivoire) in 2003, his M.S. in Numerical Analysis from the Department of Mathematics and Informatics at University of Cocody in Côte d’Ivoire.His research interests included dynamic unlicensed bands access.

Michel Babri received the PhD degree in Computer Science from Universite Clairmont Ferand . Since 1987, he is researcher at INPHB of Yamoussoukro in Cote d'Ivoire and also member of LARIT (Laboratoire de Recherche en Informatique et Télécoms) since 2006.

Enrich the E-publishing Community Website with Search Engine Optimization Technique

R. Vadivel[1] and Dr. K. Baskaran[2]

[1]PhD Scholar, Dept. of Computer Science,
Karpagam University, Coimbatore,
Tamilnadu India 641 021

[2]Assoc. Professor, Dept. of CSE and IT,
Govt. College of Technology, Coimbatore,
Tamilnadu India 641 013

ABSTRACT

Internet has played vital role in the online business. Every business peoples are needed to show their information clients or end user. In search engines have million indexed pages. A search engine optimization technique has to implement both web applications static and dynamic. There is no issue for create search engine optimization contents to static (web contents does not change until that web site is re host) web application and keep up the search engine optimization regulations and state of affairs. A few significant challenges to dynamic content poses. To overcome these challenges to have a fully functional dynamic site that is optimized as much as a static site can be optimized. Whatever user search and they can get information their information quickly. In that circumstance we are using few search engine optimization dynamic web application methods such as User Friendly URL's, URL Redirector and HTML Generic. Both internal and external elements of the site affect the way it's ranked in any given search engine, so all of these elements should be taken into consideration. Implement these concepts to E-publishing Community Website that web site have large amount of dynamic fields with dynamic validations with the help of XML, XSL & Java script. A database plays a major role to accomplish this functionality. We can use 3D (static, dynamic and Meta) database structures. One of the advantages of the XML/XSLT combination is the ability to separate content from presentation. A data source can return an XML document, then by using an XSLT, the data can be transformed into whatever HTML is needed, based on the data in the XML document. The flexibility of XML/XLST can be combined with the power of ASP.NET server/client controls by using an XSLT to generate the server/client controls dynamically, thus leveraging the best of both worlds.

Keywords: *Search Engine Optimization (SEO), community website, Dynamic web, Friendly URLs, ASP.Net, XML/XSLT and dynamic controls.*

1. Introduction

If we have a website, we definitely need it to be a friend of search engines. There are several ways to attract visitors to our website, but in order to make searchers know about our website, search engine is the tool where we need to prove our contents. If we are just having a static HTML content, then there is no much problem in promoting it. But where in today's world of Content Managed Websites and eCommerce Portals we need to look further and implement a few more techniques in order to make the site more prominent to robots. In this article we will discuss how we can develop a SEO Friendly website where the content is driven from the Database with a Content Management System which is developed using ASP.NET. We will learn to build a simple CMS driven site with no nonsense URL, which Search Engines invite.

Search Engine Optimization (SEO) is often considered the more technical part of Web marketing. This is true because SEO does help in the promotion of sites and at the same time it requires some technical knowledge – at least familiarity with basic HTML. SEO is sometimes also called SEO copyrighting because most of the techniques that are used to promote sites in search engines deal with text.

Generally, SEO can be defined as the activity of optimizing Web pages or whole sites in order to make them more search engine-friendly, thus getting higher positions in search results.

A Search Engine Optimization (SEO) is very popular term in web application industry. We can implement the SEO concepts to both applications static and dynamic web application. No matter for implement SEO to static web application. We have just followed up the SEO rules and conditions. We have to implement to dynamic / MVC web application it should be an insignificant complicate and use some tricky.

The specific objective is to develop the e-publishing community website with search engine optimization technique. There is no specified web technology in dynamic web applications. We can use any Microsoft or

any other corporation software. In my work .NET has played major role.
To understand dynamic content, it's important to have an idea of its opposite, static content. The term static content refers to web content that is generated without using a data source such as a database. Essentially, the site viewer sees exactly what is coded in the web page's HTML.
With dynamic pages, a site can display the same address for every visitor, and have totally unique content for each one to view. For example, when I visit the social networking site Face book (facebook.com), I see http://www.facebook.com/home.php as the address in my web browser, but I see a unique page that's different from what anyone else sees if they view that page at the same time. The site shows information about my friends in my account, and different information for each person in his account, or for someone who has no account.
Not all dynamically generated content is unique to every viewer, but all dynamic content comes from a data source, whether it's a database or another source, such as an XML file.

1.1 SEO in web application

A web application has playing most important role in the online business.

A million of static and dynamic web pages are available in the internet and million users can have used those web pages for their required information.

In this circumstances search engine optimization is play most important play between user and web applications.

In Million web pages are available the user should need their specific search criteria such as business man have search the own needs, students have search their own needs and etc.,

Our aim is whatever user search and they can get information their information quickly. In that situation we are using few search engine optimization methods and concepts such as a crawler, an index (or catalog) and a search interface, search engine algorithms and page rank algorithms.

Search engines take advantage of reverse broadcast networks to help save you time and money. Search allows you to "sell what your customers want, when they want it! “.

Search Engine Optimization is the science of customizing elements of your web site to achieve the best possible search engine ranking. That’s really all there is to search engine optimization. But as simple as it sounds, don’t let it fool you.

Both internal and external elements of the site affect the way it’s ranked in any given search engine, so all of these elements should be taken into consideration. Good Search Engine Optimization can be very difficult to achieve, and great Search Engine Optimization seems pretty well impossible at times.

Optimization involves making pages readable to search engines and emphasizing key topics related to your content. Basic optimization may involve nothing more than ensuring that a site does not unnecessarily become part of the invisible Web (the portion of the Web not accessible through Web search engines).

2. Existing System

Previously SEO have implemented in static commercial / non-commercial web sites. In this way there is no dynamically site map and have not well-defined RSS feed for those implementations and there is no specific way to find the back links.

Dirty URLs

Complex, hard-to-read URLs are often dubbed dirty URLs because they tend to be littered with punctuation and identifiers that are at best irrelevant to the ordinary user. URLs such as http://www.example.com/cgi-bin/gen.pl?id=4&view=basic are commonplace in today's dynamic web. Unfortunately, dirty URLs have a variety of troubling aspects, including:

URLs are difficult to type

The length, use of punctuation, and complexity of these URLs makes typos commonplace.

URLs do not promote usability

Because dirty URLs are long and complex, they are difficult to repeat or remember and provide few clues for average users as to what a particular resource actually contains or the function it performs.

URLs are a security risk

The query string which follows the question mark (?) in a dirty URL is often modified by hackers in an attempt to perform a front door attack into a web application. The very file extensions used in complex URLs such as .asp, .jsp, .pl, and so on also give away valuable information about the implementation of a dynamic web site that a potential hacker may utilize.

URLs impede abstraction and maintainability

Because dirty URLs generally expose the technology used (via the file extension) and the parameters used (via the query string), they do not promote abstraction. Instead of hiding such implementation details, dirty URLs expose the underlying "wiring" of a site. As a result, changing from one technology to another is a difficult and painful process filled with the potential for broken links and numerous required redirects.

3. Related Works

There is a three technologies have been used that is 1. User Friendly URL's, 2. URL Redirector and 3. HTML Generic.

An e-publishing community website has been used Microsoft .NET web application with ASP.NET and C#. In this application has used data model and business layer in separate module and its like a DLL (Dynamic Link Library) and we have started to created and converted dynamic URL's into Static URLs.

The URLs converting code first we must grab the incoming URL and split the extension of the page. Which pages have ".html" extension we should redirect that page to related ".aspx" page on code behind they have executed business logic or data manipulation or whatever functionality need, and display to the end user exact content for that particular page with proper Meta description and keywords. In this time of period user can only view ".html" page but all other logics will execute the code behind.

3.1 Dynamic Content and SEO

SEO for dynamic content poses a few significant challenges. Luckily, you have ways to overcome these challenges to have a fully functional dynamic site that is optimized as much as a static site can be optimized. This section discusses the pitfalls of dynamic sites, and how to overcome them to create fully optimized dynamic sites.

3.2 Challenges for Optimizing Dynamic Content

Here are some common areas of dynamic sites that provide setbacks for humans as well as search engine spiders.

3.2.1 Dynamic URL

A Dynamic URL is an address of a dynamic web page, as opposed to a Static URL, which is the address of a static web page. Dynamic URLs are typically fairly cryptic in their appearance. Here's an example from http:// www.meetingtechonline.com/ article for a product called Kindle:

http://www.meetingtechonline.com/article/product/B000FI73MA/ref=amb_link_7646122_1?pf_rd_m=ATVPDKIKX0DER&pf_rd_s=center-1&pf_rd_r=1FYB35NGH8MSMESECBX7&pf_rd_t=101&pf_rd_p=450995701&pf_rd_i=507846

Notice that the URL doesn't contain any information about the item's product type, or anything about the item's name. For a well-trusted site like Amazon, this is not a problem at all. But for a new site, or for a site that's gaining credibility and popularity, a better solution can help search results by showing a searcher some relevant keywords in the page's URL. Here's an example of something a little more effective:

http://www.meetingtechonline.com/article/products/electronics/kindle/

While search engines may not have problems indexing URLs with variables, it's important to note that highly descriptive URLs like the one just shown can get more clicks in searches than cryptic URLs.if searchers can clearly see keywords that have to do with the content they're looking for in your page's URL.

3.2.2 Logins and other forms

Login forms can restrict access to pages not only to users, but also search engines. In some cases, you want pages behind logins made searchable. In those cases, you can place code in those pages that determines whether the person visiting has access to view that content, and determine what to do from there.

Other web forms, referring to content in <FORM> tags, can restrict access to pages as well. While Google has revealed that googlebot can go through simple HTML forms (see http://googlewebmastercentral.blogspot.com/2008/04/crawling-through-html-forms.html), not all search engines follow this same process, which means content hidden behind forms may or may not be indexed.

3.2.3 Cookies

Web cookies are small bits of data that are stored in a user's web browser. Cookies are used frequently on the Web for storing temporary data like shopping cart information or user preferences. Pages that require cookies can block spiders because spiders don't store cookies as web browsers do.

3.2.4 Session IDs

Session IDs are similar to cookies in that if you need them to view pages, then spiders don't index those pages.

3.2.5 Hidden pages

Sometimes, pages on a website are hidden from search engines because they're buried too deep in a site's architecture. For example, a page more than three clicks deep from the home page of a website may not be crawled without an XML sitemap. Other pages that may be hidden include pages only visible via a site search.

3.2.6 JavaScript

Search engines don't index content that requires full-featured JavaScript. Remember that spiders view content in much the same way as you would if you were using a browser with JavaScript disabled. Text that is created using JavaScript, and therefore only accessible with JavaScript enabled, will not be indexed.

3.3 Ways to Optimize Dynamic Content

Dynamic content is often necessary in websites. In addition, content that is easily changed through an outside data source helps keep a site's content fresh and relevant. This increases its value to search engines. You don't need to worry that because your site is dynamic, your content won't be indexed. You just need to make sure you're following the appropriate guidelines when using dynamic content in order to keep your site optimized. Here are some things you can do to optimize your sites that contain dynamic content.

3.3.1 Creating static URLs

Dynamic URLs, especially dynamic URLs with vague names, can be a turnoff to searchers. In order to have friendly URLs, you want to rewrite your dynamic URLs as static URLs.

Blogs powered by wordpress or Blogger make it easy to convert dynamic links to static links. Blogger automatically creates static URLs, and with wordpress you need only a simple change in your settings. For wordpress, log in to your administrator account, and then, under Settings, click the Permalink button. From there, you simply select a static URL publishing method or create a custom one and save the changes. Nice!

If your site isn't powered by a blogging application, you need to rewrite the URLs manually. The process is somewhat complex, and it requires modifying your .htm access file. Because modifying your .htm access file can cause permanent changes to your website, you want to either practice on a testing server or know exactly what you're doing before using these techniques on a production server. To test this process on a testing server, you can download and install a testing server (discussed in Chapter 4), and then download all or part of your website to your computer. That way, changes you make on your local computer don't affect your live site.

3.3.2 Optimizing content hidden by forms

The fact that web forms can hide content can be a good thing, but sometimes forms hide content you may not want hidden. Login forms (forms that require a user name and password) can potentially block search engines if a login form is the only way to access that information. Of course, sometimes this feature is intentional, like for protecting bank account information on a banking site. For non-login forms, assuming that search engines index content that's accessible only by filling out text fields or other form elements is dangerous. Further, it's equally dangerous to assume that search engines don't index content that's accessible only via non-login forms. If you want your form's hidden content to be indexed, make sure to give access to it in ways other than through a form alone. If you don't want the content to be indexed, make sure to hide it from search engines via robots.txt, or some other method.

Typically, content that's viewable only after a user is logged into an account isn't necessary to index. If you have content that you want indexed hidden in a login-only area, consider taking that content out of the restricted area so it can be indexed.

4. Results

Successfully implemented search optimization e-publishing community web application with help of those technologies. Here show the few mock-up screen shots.

Fig – 1 and 2 have found relevant search information on google search engine.

Fig – 3 has displayed detail information about that article which found in google search results.

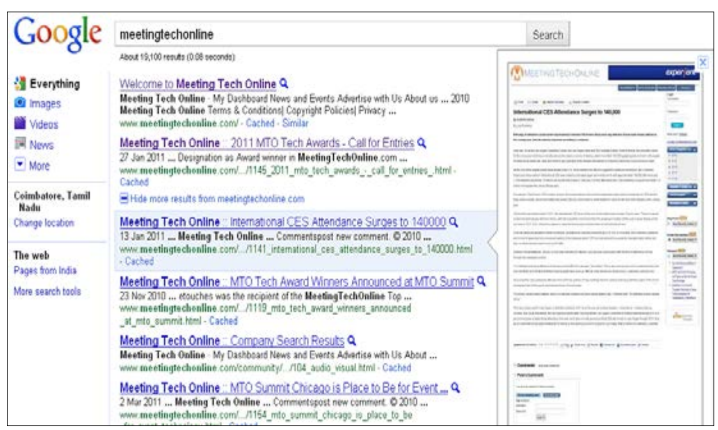


Fig 1: Getting relevant search information from google search engine– I.

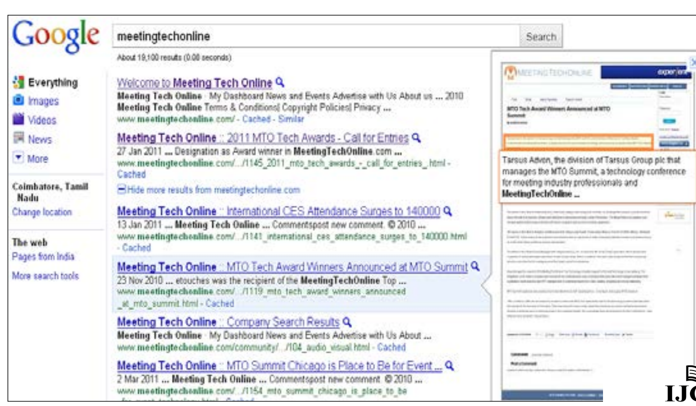


Fig 2: Getting relevant search information from google search engine– II.

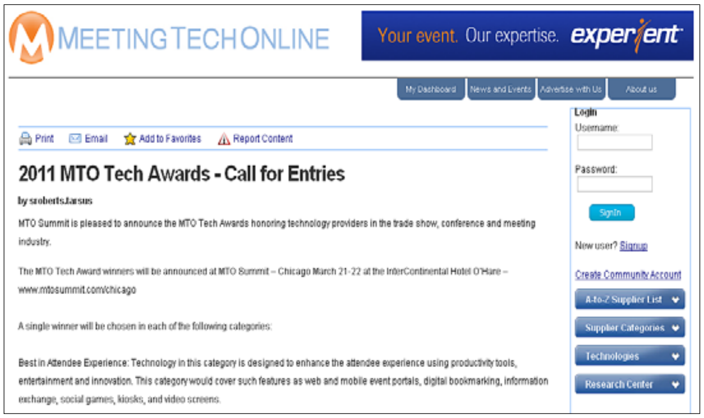


Fig 3: Detail information about that article which found in google results.

Hence we have successfully implemented Search Engine Optimization technique for e-publishing community web application.

5. Conclusions

A Search Engine Optimization has been implemented in E-Publishing community website. It has been used model view controller patterns to develop the web application and their techniques such as URL Redirector, HTML Generic, .NET security tools. The proposed is implementing the multiple query searches and personalized concept based clustering.

Most of the tips presented here are fairly straightforward, with the partial exception of URL cleaning and rewriting. All of them can be accomplished with a reasonable amount of effort. The result of this effort should be cleaned URLs that are short, understandable, permanent, and devoid of implementation details. This should significantly improve the usability, maintainability and security of a web site. The potential objections that developers and administrators might have against next generation URLs will probably have to do with any performance problems they might encounter using server filters to implement them or issues involving search engine compatibility. As to the former, many of the required technologies are quite mature in the Apache world, and their newer IIS equivalents are usually explicitly modelled on the Apache exemplars, so that bodes well.

As to the search engine concerns, fortunately, Google so far has not shown any issue at all with cleaned URLs. At this point, the main thing standing in the way of the adoption of next generation URLs is the simple fact that so few developers know they are possible, while some who do are too comfortable with the status quo to explore them in earnest. This is a pity, because while these improved URLs may not be the mythical URN-style keyword always promised to be just around the corner, they can substantially improve the web experience for both users and developers alike in the long run.

Reference

[1] Kenneth Wai-Ting Leung, Wilfred Ng, and Dik Lun Lee, “Personalized Concept-Based Clustering of Search Engine Queries”, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 20, NO. 11, NOVEMBER 2008

[2] AARON MATTHEW WALL,”Search Engine Optimization”, JUNE 2008

[3] Ernest Ackermann & Karen Lartman, “The information Specialist Guide to Searching & researching on the Internet and World Wide Web”, Fitzroy Dearbon Publishers, 1999.

[4] R.Elmasri and S.B. Navathe, “Fundamentals of Database Systems”, 2nd Edition, Menlo Park, CA: Addison- Wesley 1994.

[5] Jeff Ferguson, Brian Patterson, Pierre Boutquin “ C# Bible”, John Wiley and Sons, June 2002

[6] Wei Meng Lee, “C#.net Web Developer's Guide”, Syngress, January 1970

[7] Jose Mojica, C# Web Development for ASP.NET, Peachpit Press, March 2003

[8] Ke Yi , Feifei Li , Graham Cormode , Marios Hadjieleftheriou , George Kollios , Divesh Srivastava, Small synopses for group-by query verification on outsourced data streams, ACM Transactions on Database Systems (TODS), v.34 n.3, p.1-42, August 2009

[9] HweeHwa Pang , Jilian Zhang , Kyriakos Mouratidis, Scalable verification for outsourced dynamic databases, Proceedings of the VLDB Endowment, v.2 n.1, August 2009

[10] Alberto Trombetta, Danilo Montesi, "Equivalences and Optimizations in an Expressive XSLT Fragment" IEEE

TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 20, NO. 11, JULY 2009

[11] Kyriakos Mouratidis , Dimitris Sacharidis , Hweehwa Pang, Partially materialized digest scheme: an efficient verification method for outsourced databases, The VLDB Journal — The International Journal on Very Large Data Bases, v.18 n.1, p.363-381, January 2009

[12] HweeHwa Pang , Kyriakos Mouratidis, Authenticating the query results of text search engines, Proceedings of the VLDB Endowment, v.1 n.1, August 2008

[13] http://www.macronimous.com/ 2009

[14] http://www.webtop.com.au/seo 2009

[15] http://www.seocompany.ca/seo/seo-techniques.html

[16] http://searchengineland.com/21-essential-seo-tips-techniques-11580

[17] http://msdn.microsoft.com/en-us/library/ms972974.aspx

Fuzzy Kernel and Fuzzy Subsemiautomata with Thresholds

M.Basheer Ahamed[1], and J.Michael Anna Spinneli[2]

[1] Department of Mathematics, University of Tabuk,

Tabuk-71491, Kingdom of Saudi Arabia.

[2] Department of Mathematics, Karunya University,

Coimbatore-641114, Tamil Nadu, India.

Abstract

In 1967, Wee introduced the concept of fuzzy automata, using Zadeh's concept of fuzzy sets. A group semiautomaton has been extensively studied by Fong and Clay .This group semiautomaton was fuzzified by Das and he introduced fuzzy semiautomaton, fuzzy kernel and fuzzy subsemiautomaton over finite group. Fuzzy subgroup with thresholds was defined by Yuan et. al. In this paper, we introduce the idea of fuzzy kernel and fuzzy subsemiautomaton with thresholds. Further, we discuss some results concerning them.

Mathematics subject Classification 2010: 03E72, 20M35.

Keywords: *Fuzzy kernel with thresholds, Fuzzy subsemiautomaton with thresholds, Fuzzy subgroups with thresholds*.

1. Introduction

Lofti A. Zadeh introduced fuzzy sets in 1965. Rosenfeld defined fuzzy subgroups in 1971. Anthony and Sherwood replaced "min" in Rosenfeld axiom by t-norm and introduced T-Fuzzy subgroup. Bhakat and Das introduced ($\in$, $\in \vee q$)-fuzzy normal, quasinormal and maximal subgroup in 1992.Also in 1997 they introduced fuzzy kernel and fuzzy subsemiautomaton of a fuzzy semiautomaton over a finite group using the notions of a fuzzy normal subgroup and a fuzzy subgroup of a group. In the year 1999 Sung-jin Cho et al. introduced the notion of T-fuzzy semiautomata, T-fuzzy kernel, and T-fuzzy subsemiautomata over a finite group. In 2003, Yuan et. al defined fuzzy subgroup with thresholds which is a generalization of Rosenfeld's fuzzy subgroup and Bhakat and Das's fuzzy subgroup. This paper defines fuzzy kernel with thresholds, fuzzy subsemiautomaton with thresholds and discusses some results concerning them.

2. Preliminaries

In this section we summarize some preliminary definitions and results which are required for developing main results.

Let $(G,*)$ denote a group. We sometimes write G for $(G,*)$ when the operation $*$ is understood.

2.1 Definition [7]

Let X be a nonempty set. A fuzzy set A in X is characterized by its membership function $A: X \to [0,1]$ and $A(x)$ is interpreted as the degree of membership of element x in fuzzy set A for each $x \in X$.

2.2 Definition [4]

A fuzzy subset λ of a group G is a fuzzy subgroup of G if for all $x, y \in G$

(i) $\lambda(x*y) \geq \lambda(x) \wedge \lambda(y)$

(ii) $\lambda(x^{-1}) \geq \lambda(x)$

2.3 Definition [4]

A fuzzy subgroup λ of G is called a fuzzy normal subgroup of G if $\lambda(x*y*x^{-1}) \geq \lambda(y)$ for all $x, y \in G$.

2.4 Definition [6]

Let $\lambda, \mu \in [0,1]$ and $\lambda < \mu$ .Let A be a fuzzy subset of a group G . Then A is called a fuzzy subgroup with thresholds of G if for all $x, y \in G$

(i) $A(x*y) \vee \lambda \geq A(x) \wedge A(y) \wedge \mu$

(ii) $A(x^{-1}) \vee \lambda \geq A(x) \wedge \mu$

2.5 Definition [2]

Let $\alpha, \beta \in [0,1]$ and $\alpha < \beta$. Let μ be a fuzzy subset of a group G. Then μ is called a fuzzy normal subgroup with thresholds of G if $\mu(y^{-1}*x*y) \vee \alpha > \mu(x) \wedge \beta, \forall x, y \in G.$

2.6 Definition [4]

A fuzzy semiautomaton over a finite group $(Q,*)$ is a triple (Q, X, μ) where X is a finite set and μ is a fuzzy subset of $Q \times X \times Q$.

2.7 Definition [4]

Let $S = (Q, X, \mu)$ be a fuzzy semiautomaton over a finite group G. A fuzzy subset λ of Q is called fuzzy kernel of S if the following conditions hold. For all $p, q, r, k \in Q, x \in X$

(i) λ is a fuzzy normal subgroup of Q

(ii) $\lambda(p*r^{-1}) \geq \mu(q*k, x, p) \wedge \mu(q, x, r) \wedge \lambda(k)$

2.8 Definition [4]

Let $S = (Q, X, \mu)$ be a fuzzy semiautomaton over a finite group G. A fuzzy subset λ of Q is called fuzzy subsemiautomaton of S if the following conditions hold:

(i) λ is a fuzzy subgroup of Q

(ii) $\lambda(p) \geq \mu(q, x, p) \wedge \lambda(q)$ $\mathit{for\ all}\ p, q \in Q, x \in X.$

2.9 Definition [4]

Let $\lambda\ \mathit{and}\ \mu$ be fuzzy subsets of G. The product $\lambda * \mu$ of λ and μ is defined by $(\lambda*\mu)(x) = \vee\{\lambda(y) \wedge \mu(z) / y, z \in G\ \mathit{such\ that}\ x = y*z\}$

3. Main Results

3.1 Definition

Let $S = (Q, X, \mu)$ be a fuzzy semiautomaton over a finite Group. A fuzzy subset λ of Q is called a fuzzy kernel of S with thresholds if

(i) λ is a fuzzy normal subgroup of Q with thresholds

(ii) $(\lambda(p*r^{-1}) \vee \alpha) \geq \mu(q*k, x, p) \wedge \mu(q, x, r) \wedge \lambda(k) \wedge \beta$ for all $p, q, r, k \in Q$

3.2 Definition

Let $S = (Q, X, \mu)$ be a fuzzy semiautomaton over a finite Group. A fuzzy subset λ of Q is called a fuzzy subsemiautomaton of S with thresholds if the following conditions hold

(i) λ is a fuzzy subgroup of Q with thresholds

(ii) $(\lambda(p)\vee\alpha)\ge\mu(q,x,p)\wedge\lambda(q)\wedge\beta$ for all $p,q\in Q\, and\, x\in X$.

3.3 Definition

Let λ and μ be fuzzy subsets of G with thresholds α,β . The product $\lambda*\mu$ is defined by

$$((\lambda*\mu)(x)\vee\alpha)=\vee\begin{Bmatrix}\lambda(y)\wedge\mu(z)\wedge\beta/\\ y,z\in G\, such\, that\, x=y*z\end{Bmatrix}$$

Note

Let $S=(Q,X,\mu)$ be a fuzzy semiautomaton over a finite group in the remaining of the results. The element $'e'$ be the identity of $(Q,*)$.

3.4 Proposition

Let $\alpha,\beta\in[0,1]$ and $\alpha<\beta$. Let λ be a fuzzy kernel of $S=(Q,X,\mu)$ with thresholds α,β . Then λ is a fuzzy subsemiautomaton of S with thresholds α,β if and only if $(\lambda(p)\vee\alpha)\ge\mu(e,x,p)\wedge\lambda(e)\wedge\beta$ for all $p\in Q,x\in X$

Proof: We have $(\lambda(p)\vee\alpha)=\lambda(p*r^{-1}*r)\vee\alpha$

$\ge\lambda(p*r^{-1})\wedge\lambda(r)\wedge\beta$ (By definition 2.2)

$\ge(\lambda(p*r^{-1})\wedge\lambda(r)\wedge\beta)\vee\alpha$

$=(\mu(p*r^{-1})\vee\alpha)\wedge\lambda(r)\wedge\beta$

$\ge(\mu(q,x,p),\wedge\mu(e,x,r)\wedge\lambda(q)\wedge\beta)\wedge\lambda(r)\wedge\beta$
(By definition 3.1)

$\ge(\mu(q,x,p)\wedge\mu(e,x,r)\wedge\lambda(e)\wedge\lambda(q)\wedge\beta)$

Since $\lambda(r)\ge\mu(e,x,r)\wedge\lambda(e)$ by given condition

$\ge\mu(q,x,p)\wedge\mu(e,x,r)\wedge\lambda(q)\wedge\beta$

Since $\lambda(e)\ge\lambda(q)$

$\ge\mu(q,x,p)\wedge\lambda(q)\wedge\beta$

3.5 Proposition

Let $\alpha,\beta\in[0,1]$ and $\alpha<\beta$. Let λ be a fuzzy kernel of $S=(Q,X,\mu)$ with thresholds α,β and ν be a fuzzy subsemiautomaton of S with thresholds α,β . Then $\lambda*\nu$ is a fuzzy subsemiautomaton of S with thresholds α,β .

Proof: Since λ is a fuzzy normal sub group with thresholds and ν is a fuzzy sub group with thresholds of Q, it follows that $\lambda*\nu$ is a fuzzy sub group of Q with thresholds α,β

$((\lambda*\nu)(p)\vee\alpha)=((\lambda*\nu)(p*r*r^{-1}))\vee\alpha$,
By definition 3.3

$\ge(\lambda(p*r^{-1})\wedge\nu(r))\wedge\beta$

$$\geq\left(\left(\lambda\left(p*r^{-1}\right)\wedge\nu\left(r\right)\right)\vee\alpha\right)\wedge\beta$$

Using Lemma 2.1 in [2]

$$=\left(\lambda\left(p*r^{-1}\right)\vee\alpha\right)\wedge\left(\nu\left(r\right)\vee\alpha\right)\wedge\beta,$$ since $\vee$ is distributive.

$$\geq\left(\mu\left(a*b,x,p\right)\wedge\mu\left(a,x,r\right)\wedge\lambda\left(b\right)\wedge\beta\right)\wedge \left(\mu\left(a,x,r\right)\wedge\nu\left(a\right)\wedge\beta\right)\wedge\beta$$

$$=\mu\left(a*b,x,p\right)\wedge\lambda\left(b\right)\wedge\nu\left(a\right)\wedge\beta$$

$$\left(\left(\lambda*\nu\right)\left(p\right)\vee\alpha\right)\geq\vee\begin{Bmatrix}\mu\left(a*b,x,p\right)\wedge\lambda\left(b\right)\wedge\nu\left(a\right)\wedge\beta\,/\\ \forall a,b\in Q,a*b=q\end{Bmatrix}$$

$$=\mu\left(q,x,p\right)\wedge\left(\vee\begin{Bmatrix}\lambda\left(b\right)\wedge\nu\left(a\right)\wedge\beta\,/\\ a*b=q,\forall a,b\in Q\end{Bmatrix}\right)$$

$$=\mu\left(q,x,p\right)\wedge\left(\lambda*\nu\right)\left(q\right)\vee\alpha$$

$$>\mu\left(q,x,p\right)\wedge\left(\lambda*\nu\right)\left(q\right)\wedge\beta$$

$$\geq\mu\left(q,x,p\right)\wedge\left(\lambda*\nu\right)\left(q\right)\wedge\beta$$

3.6 Proposition

Let $\alpha,\beta\in\left[0,1\right]$ and $\alpha<\beta$ .If λ and ν are fuzzy kernels of S with thresholds α,β then $\lambda*\nu$ is a fuzzy kernel of S with thresholds α,β .

Proof: Since λ and ν are fuzzy normal subgroups of Q with thresholds then $\lambda*\nu$ is also fuzzy normal sub group of Q with thresholds

$$\left(\left(\lambda*\nu\right)\left(p*r^{-1}\right)\vee\alpha\right)\geq\left(\lambda*\nu\right)\left(p*q^{-1}*q*r^{-1}\right)\vee\alpha$$

$$\geq\lambda\left(p*q^{-1}\right)\wedge\nu\left(q*r^{-1}\right)\wedge\beta$$ (By normal subgroup definition)

$$=\left(\lambda\left(p*q^{-1}\right)\vee\alpha\right)\wedge\left(\nu\left(q*r^{-1}\right)\vee\alpha\right)\wedge\beta$$

Using Lemma 2.1 in[2]

$$\geq\left(\mu\left(a*b*c,x,p\right)\wedge\mu\left(a*b,x,q\right)\wedge\lambda\left(c\right)\wedge\beta\right)\wedge \left(\mu\left(a*b,x,q\right)\wedge\mu\left(a,x,r\right)\wedge\nu\left(b\right)\wedge\beta\right)\wedge\beta$$

(By definition 3.1)

$$=\mu\left(a*b*c,x,p\right)\wedge\lambda\left(c\right)\wedge\nu\left(b\right)\wedge\beta$$

$$\left(\text{Since }\mu\left(a*b*c,x,p\right)\leq\mu\left(a*b*,x,p\right)\right)$$

$$\left(\lambda*\nu\right)\left(p*r^{-1}\right)\vee\alpha\geq\vee\begin{Bmatrix}\mu\left(q*b*c,x,p\right)\wedge\mu\left(q,x,r\right)\\ \wedge\lambda\left(c\right)\wedge\nu\left(b\right)\wedge\beta\,/\\ b,c\in Q,b*c=k\end{Bmatrix}$$

$$=\mu\left(q*k,x,p\right)\wedge\mu\left(q,x,r\right)\wedge\left(\vee\begin{Bmatrix}\lambda\left(c\right)\wedge\nu\left(b\right)\wedge\beta\,/\\ b*c=k,b,c\in Q\end{Bmatrix}\right)$$

$$= \mu(q * k, x, p) \wedge \mu(q, x, r) \wedge (\lambda * \mu)(k) \vee \alpha$$

$$> \mu(q * k, x, p) \wedge \mu(q, x, r) \wedge (\lambda * \mu)(k) \wedge \beta$$

References

[1] Anthony J. M. and Sherwood, H., A characterization of fuzzy groups, Fuzzy Sets and System, 7, (1982) pp 297-305.

[2] Bao Qing Hu, Fuzzy groups and T- fuzzy Groups with Thresholds, Advances in fuzzy mathematics Vol. 5, No.1, (2010) pp. 17-29.

[3] Bhakat S. K., (∈, ∈Vq) -fuzzy normal, quasinormal and maximal Subgroups, Fuzzy Sets and Systems, 112, (2000) pp. 299-312.

[4] P. Das, On some properties of fuzzy semiautomaton over a finite group, Information Sciences 101, (1997) pp.71-84

[5] Rosenfeld, A., Fuzzy groups, Journal of Mathematical Analysis and Applications, 35, (1971) pp. 512-517.

[6] Yuan, X., Zhang, C., and Ren, Y., Generalized fuzzy groups and many valued implications, Fuzzy Sets and Systems, 138, (2003) pp. 205-211.

[7] Zadeh, L. A., Fuzzy sets, Information and Control,8,(1965)pp.338-353

Basheer Ahamed M received the Ph,D.,degree in Mathematics from Bharathidasan University, Tiruchirappalli, Tamilnadu, India in 2005. He is working inUniversity of Tabuk, Kingdom of Saudi Arabia . His research areas are FuzzyMathematics and Discrete Mathematics.
J. Michael Anna Spinneli received her M.Sc and M.Phil degrees in Mathematics From Manonmanium Sundaranar.University.Tirunelveli, Tamilnadu. Now she is working as an Assistant Professor of mathematics inKarunya University, Coimbatore, India. She is doing research on Fuzzy Automata.

Power Optimization in Wireless Sensor Networks

Debmalya Bhattacharya[1] and R.Krishnamoorthy[2]

[1] School of Electrical
VelTech Dr.RR & Dr.SR Technical University
Chennai,Tamilnadu,India

[2] School of Electrical
VelTech Dr.RR & Dr.SR Technical University
Chennai, Tamilnadu, India

Abstract—
Wireless Sensor Networks (WSNs) consist of a network of wireless nodes that have the capability to sense a parameter of interest. Sensors of various types are deployed ubiquitously and pervasively in varied environments such as office buildings, wildlife reserves, battle fields, mobile networks, etc The sensed parameter is relayed to a base station through the network formed amongst these nodes. The devices used are typically characterized by low cost, low power and are rugged in operation. The node integrates programming, computation, communication, and sensing onto a single system and provides an easy user interface for operating and deploying it. The paper presents such a design which minimizes cost and power consumption, thus enhancing the life time of the node.

Keywords: node, Wireless Sensor network, ZigBee

I. INTRODUCTION

The advances in the hardware and wireless technologies have resulted in inexpensive low power communication devices that can be deployed throughout a physical space, providing dense sensing close to physical phenomena, processing and communicating this information, and coordinating actions with other nodes. Such a deployment can be termed as a Wireless Sensor Network (WSN). To realize such a network, we must address a new collection of challenges. The individual devices in a WSN are inherently resource constrained: they have limited processing speed, storage capacity, battery capacity, and communication bandwidth. These devices have substantial processing capability in the aggregate, but not individually. These devices are called as nodes.

A “node” in a wireless sensor network is capable of gathering information, processing and communicating with other connected nodes in the network. Typically the node may contain one or more sensors that can monitor the surroundings for specific parameters. Some of these sensors commonly used are to sense temperature, light, sound, position, acceleration, vibration, stress, weight, pressure, humidity, etc. The microcontroller performs all the data processing tasks and controls the functionality of other components in the sensor node. The sensors measure data of the area to be monitored. The continual analog signal sensed by the sensors is digitized by an Analog-to-digital converter and sent to controllers for further processing. The nodes also contain the communication module which provides communication over wireless medium using transceivers. The nodes can be powered by using batteries. A large number of nodes hence communicate over wireless channel form an ad-hoc network. All the information can eventually be transmitted to a gateway node.

By forming an ad-hoc network they can function for a long time without any human intervention since they consume limited power.

The paper is organized as follows: Section 2 deals with the architectural issues in wireless sensor network. Section 3, presents the overview of Wireless Sensor Network. Section 4 deals with power constraints in WSN. Section 5 deals with the various aspects of ZigBee. The paper is concluded in section 6.

II.ARCHITECTURAL ISSUES IN WSN

Wireless Sensor Networks (WSNs) are an important new class of networked system. Simultaneously presenting intellectually deep CISE research challenges and promising tremendous societal impact through scientific progress, better engineering, improved productivity, and enhanced security, research in this area has progressed substantially. Dealing with both scale and density is hard enough in ideal environments. Unfortunately, we don’t have the luxury of ideal environments with sensor networks. Because sensor networks are intended to monitor the physical world, they must often be deployed in natural and uncontrolled environments. No longer can we assume the carefully controlled temperature, abundant power, and human monitoring of server rooms and data centers. Instead, wireless sensor networks must be designed to operate while unthread (no external power), unattended (no manual configuration or management), intermittently connected (radios may be turned off for substantial periods of time to conserve power), and uncontrolled environment.[1]. The following architectural issues must be considered while designing power optimized WSN.

A. *Implications on WSN*

1. *Network Dynamics*

There are three main components in a sensor network. These are the sensor nodes, sink and monitored events. Supporting the mobility of sink nodes is an important issue in WSN design. Routing plays important role as routing messages from or to moving nodes is more challenging since route stability becomes an important optimization factor, in addition to energy, bandwidth etc [2].

2. *Node Deployment:*

Another consideration is the topological deployment of nodes. This is application dependent and affects the performance of the routing protocol. The deployment is either deterministic or self-organizing

3. *Energy Considerations:*

During the creation of an infrastructure, the process of setting up the routes is greatly influenced by energy considerations. Since the transmission power of a wireless radio is proportional to distance squared or even higher order in the presence of obstacles, multi hop routing will consume less energy than direct communication.

4. *Node Capabilities:*

In a WSN, different functionalities can be associated with the sensor nodes. According to previous studies, all sensor nodes are assumed to be homogenous, having equal capacity in terms of computation, communication and power. However, depending on the application a node can be dedicated to a particular special function such as relaying, sensing and aggregation since engaging the three functionalities at the same time on a node might quickly drain the energy of that node. Inclusion of heterogeneous set of sensors raises multiple technical issues related to data routing. The results generated from these sensors can be at different rates, subject to diverse quality of service constraints and following multiple data delivery models. Therefore, such a heterogeneous environment makes data routing more challenging.

III. OVERVIEW OF WIRELESS SENSOR NETWORK

Sensor networks are applied to various fields ranging from special application fields such as wild environment monitoring, industrial machine measurement and military purpose measurement to daily application fields such as fire monitoring and pollution monitoring.[3]. A wireless sensor network is a wire and wireless network, which consists of several sensor nodes deployed in a certain field. A sensor node should have computation, sensing and wireless communication functions.

The numerous sensors are used for delivering crucial information in real-time from environments and processes, where data collection is impossible previously with wired sensors [4]. Typically, wireless sensor networks are composed of low power sensor nodes and integrate general-purpose computing with heterogeneous sensing and wireless communication. Their emergence has enabled observation of the physical world at an unprecedented level of granularity. One of the most important components of a sensor node is the power unit [5].A wireless sensor network limits the radio frequency channel, due to, that is to say, unstable links, limit of physical protection of each sensor node, actual of each nodes connection, variation topology in addition dangerousness about routing security is high by activity spite nodes. In addition, restrictions of the hardware of the sensor node itself makes it difficult guarantee the maintenance of security because of vulnerability.[6][7].A wireless sensor network consists of spatially distributed autonomous sensors to cooperatively monitor physical or environmental conditions, such as temperature, sound, vibration, pressure, motion or pollutants.

A sensor network normally constitutes a wireless ad-hoc network, meaning that each sensor supports a multi-hop routing algorithm. The network does not rely on a pre-existing infrastructure, such as routers in wired networks or access points in managed (infrastructure) wireless networks. Instead, each node participates in routing by forwarding data for other nodes, and so the determination of which nodes forward data is made dynamically based on the network connectivity.

The sensor nodes must
(i) Consume extremely low power
(ii) Operate in high volumetric densities
(iii) Have low production cost and be dispensable and
(iv) Be adaptive to the environment.

The base stations are one or more distinguished components of the WSN with much more computational, energy and communication resources. They act as a gateway between sensor nodes and the end user. Fig1 shows the basic components of a sensor node.

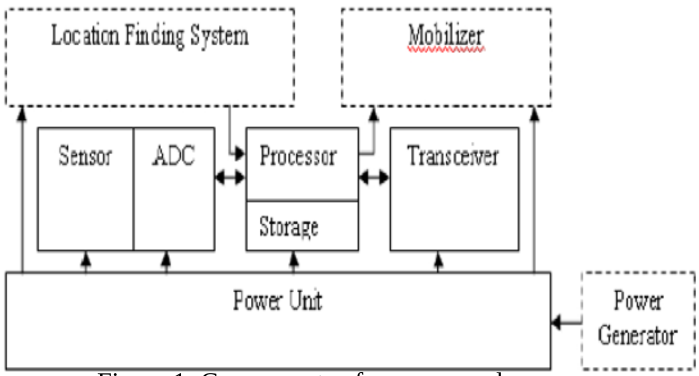


Figure 1: Components of a sensor node

IV. POWER CONSTRAINTS IN WSN

Wireless sensor networks typically have power constraints. The absence of wires implies the lack of an external power supply such as battery packs. Although photovoltaic's or other passive energy gathering techniques are possible, these approaches typically provide only a modest amount of operating pow
Therefore it is necessary to extend the battery life

individual sensors so that the network can remain functional as long as possible. Moreover, for biomedical sensors, power usage results in heat dissipation that may further require minimizing the total power consumed by the wireless sensor network.[9].

The handling of the wireless transceiver contributes significantly to the node's overall energy consumption [10]. In order to extend the working time of individual devices, it is frequent practice that some node elements are deactivated, including the radio transceiver. They remain inactive for most time and are activated only to transmit or receive messages from other nodes. Radio transceiver in WSN network node can operate in one out of four modes, which differ in the consumption of power necessary for proper operation: transmission – signal is transmitted to other nodes (greatest power consumption), receiving – message from other node is received (medium power consumption), stand-by (idle) – transceiver inactive, turned on and ready to change to data transmission or receiving (low power consumption), sleep – radio transceiver off [11].

A typical wireless sensor network consists of sensors powered by small batteries that are difficult to replace if not impossible. Hence, the sensor nodes can only transmit a finite number of bits before they run out of energy. Thus, reducing the energy consumption per bit for end-to-end data transmission is an important design consideration for such networks. We assume that each information bit collected by a sensor is useful for a finite amount of time; after this time the information may become irrelevant. Hence all the bits collected by the sensors need to be communicated to a hub node before a certain deadline. Therefore, the maximum end to-end transmission delay for each bit must be controlled to meet a given deadline under the hard energy constraint. Since all layers of the protocol stack affect the energy consumption and delay for the end-to-end transmission of each bit, an efficient system requires a joint design across all these layers as well as the underlying hardware where the energy is actually expended. [12].

For many applications, wireless sensor networks (WSNs) are required to be unobtrusive, with numerous nodes that are dependent on a battery power source. These nodes are typically very simple, small, and inexpensive modules that are equipped with a sensor to measure a phenomenon. A simple transceiver is used to transmit and receive the measured observations to and from neighboring nodes.

Figure 2: A typical sensor node

Figure 2 depicts a typical sensor node with these simple components. Ultimately, the sensor nodes cooperate in transmitting their observations to a data sink where they can be processed as shown in Figure 3. Since nodes must be as small, inexpensive, and as efficient as possible, there are stringent constraints on their computational and energy resources. On the other hand, the data sink is assumed to have access to substantial energy and computational resources, within the limits of reasonable expense and contemporary technology. Therefore one of the main challenges of designing a successful WSN is in minimizing the probability of error in transmitting data, subject to constraints in available power and computational resources [13].

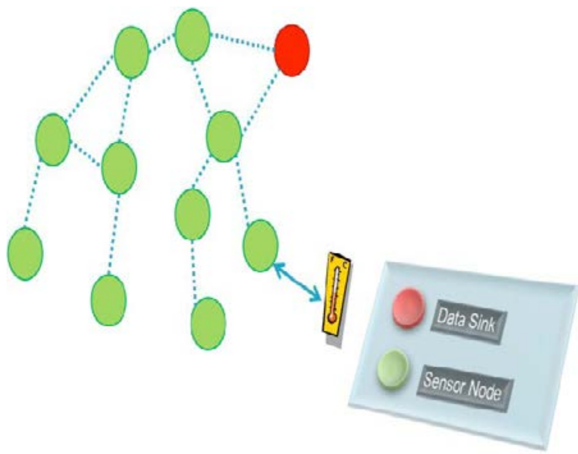


Figure 3: A Typical Sensor Network

V. WHY ZIGBEE FOR WIRELESS SENSOR NETWORK?

The name "ZigBee" is derived from the erratic zigzag patterns many bees make between flowers when collecting pollen. This is suggestive of the invisible webs of connections existing in a fully wireless environment, similar to the way packets would move through a mesh network.

During the last few years, the ZigBee Alliance has made significant modification and improvement on the ZigBee standard IEEE 804.15.4, making it more applicable to the increasing demand on Personal Area Network (PAN) service. However, the ZigBee standard specified that the maximum data rate of a ZigBee link be 250kbps. This data rate faces many difficulties dealing with the increasing data transmission pressure in many applications [14]. ZigBee is a low-cost, low-power wireless mesh networking proprietary standard. The low cost allows the technology to be widely deployed in wireless control and monitoring applications, the low power-usage allows longer life with smaller batteries and the mesh networking provides high reliability and larger range. The technology defined by the ZigBee is intended to be simpler and less expensive than other WPANs, such as Bluetooth. ZigBee wireless sensor network has great advantages in terms of low power consumption, high fault tolerance, flexibility, and autonomy.

ZigBee operates in the industrial, scientific and medical (ISM) radio bands; 868 MHz in Europe, 915 MHz in the USA and Australia, and 2.4 GHz in most jurisdictions worldwide. The ZigBee specifications are available free of cost for all non-commercial purposes. ZigBee can go from sleep to active mode in 15 msec or less, thus the latency can be very low and devices can be very responsive particularly compared to Bluetooth wake-up delays, which are typically around three seconds. Because ZigBee can sleep most of the time, average power consumption can be very low, resulting in long battery life. It is a typical wireless communication technology, which is widely used in wireless sensor network. ZigBee wireless sensor network has great advantages in terms of low power consumption, high fault tolerance, flexibility, and autonomy [15]. Using Zigbee techniques as a back bone to develop ubiquitous applications has been warming up while current information technology evolution moving from electrification to mobilization. However, most successful business cases still rely on mobile tools, such as PDA, WIFI, RFID, and GPS, to realize the concept of ubiquitous. The main challenges ahead for making real U-applications on the market are not only the definitions of ZigBee specifications and protocols, but the related optimal database build-up and interface design methods [16].

The block diagram of a node is shown in the fig4 [17].

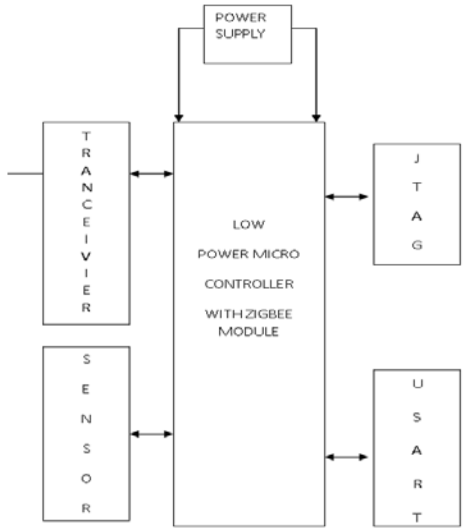


Figure4. Block Diagram of node

As the main controller of the whole system, the microcontroller's major responsibilities are initializing the system, receiving and executing the orders and memorizing these. The flow of the main programs is illustrated in figure 4. [Q].

VI. CONCLUSION

In this paper we have presented the power related issues of WSN and the role of ZigBee in designing the node. In a wireless sensor node, the radio consumes a vast majority of the system energy. This power consumption can be reduced through decreasing the transmission output power or through decreasing the radio duty cycle. Both of these alternatives involve sacrificing other system metrics.

REFERENCES

[1] http://snap.cs.berkeley.edu/documents/architecture.pdf

[2] Kemal Akkaya and Mohamed Younis , "A Survey on Routing Protocols for Wireless Sensor Networks" http://www.cs.umbc.edu/~kemal1/mypapers/Akkaya_Younis_JoAdH ocRevised.pdf

[3] Yong-Sik Choi, Young-Jun Jeon, Sang-Hyun Park, "A study on sensor nodes attestation protocol in a Wireless Sensor Network", ICACT 2010.

[4] Warneke, B., Last, M., Liebowitz, B., and Pister, K. S. J. Smartdust: Communicating with a cubic-millimeter computer. Computer 34, 1(2001), 44–51.

[5] M.Ismail, M.Y.Sanavullah, "Security Topology in Wireless Sensor Networks With Routing Optimisation"IEEE 2008.

[6] G.M. Ben Ezovski, S.E. Watkins, "The Electronic Sensor Node and the Future of Government-Issued RFID-Based Identification", RFID 2007. IEE

International Conference, pp 15-22, 2007
[7] I.F. Akyiliz, W. Su, Y.Sankarasubramaniam and E. Cayirci, "A Survey on Sensor Network", IEEE Communication Magazine, pp 102-114, 2002.
[8]. F. Akyilidz, W.Su, Y. Sankarasubramaniam and E.Cayirici, "A survey on sensor networks", IEEE communication magazine, Aug 2002: 102-114.
[9] Ayad Salhieh Jennifer Weinmann _ Manish Kochhal _ Loren Schwiebert, "Power Efficient Topologies forWireless Sensor Networks" reference.kfupm.edu.sa/.../power_efficient_topologies_for_wireless_529104.pdf
[10] Ewa Niewiadomska-Szynkiewicz, Piotr Kwaśniewski, Izabela Windyga, "Comparative Study of Wireless Sensor Networks Energy-Efficient Topologies and Power Save Protocols"
[11] Shuguang Cui and Andrea J. Goldsmith, "Cross-layer Optimization in Sensor Networks with Energy Constraints", http://wsl.stanford.edu/projects/energy_eff.pdf
[12] Nariman Farsad, "Low-Complexity Energy Optimization of Wireless Sensor Networks", Thesis, Master of Science, Toranto 2009.
[13] Shaosheng Li, Wenze Li, Jianxun Zhu, "A Novel Zigbee Based High Speed Ad Hoc Communication Network", Proceedings of IC-NIDC2009,IEEE 2009.
[14] Liu Yanfei, Wang Cheng, Qiao Xiaojun, Zhang Yunhe, Yu chengbo,Liu Yanfei, "An Improved Design of ZigBee Wireless Sensor Network", IEEE 2009.
[15] Chia-Ping Huang, "Zigbee Wireless Network Application Research Case Study Within Taiwan University Campus", Proceedings of the Eighth International Conference on Machine Learning and Cybernetics, Baoding, 12-15 July 2009.
[16] Nikhil S Kulkarni, Rakesh R, Sarthak Bhargava, Shantanu Singh Bundela, Rajeshwari Hegde, "Zigbee Based Low Power Wireless Sensor Network Motes", IET International conference on Next Generation Network", Sept 24-25th, 2010, Mumbai.
[17] Peng Jiang , Hongbo Xia, Zhiye He and Zheming Wang. "Design of a Water Environment Monitoring System based on Wireless Sensor Networks" , Sensors 2009, ISSN 1424-8220

Dr. Debmalya Bhattacharya India. He is working as Associate Professor in Vel Tech Technical University Chennai. He has also published papers on wireless communication in National and International Conferences. His current research interests are Robotics, wireless Communications.

R.Krishnamoorthy, Finished his B.E in Sree Sastha Institute of Science and Technology and perusing M.Tech in VelTech Technical University. Working as Junior Research Scholar (TCOE), He presented papers in several national and international conferences.

Towards Maximum Spanning Tree Model in Web 3.0 Design and Development for Students using Discriminant Analysis

S.Padma [1], Dr.Ananthi Seshasaayee [2]

[1] Research Scholar, Bharathiar University, Coimbatore, Assistant Professor, School of Computing Sciences, Vels University, Chennai , India.

[2]Dr.Ananthi Seshasaayee, Associate Professor and Head, Department of Computer Science Quaid-e-Millath Government College for women, Chennai , India.

Abstract

Web 3.0 is an evolving extension of the web 2.0 scenario. The perceptions regarding web 3.0 is different from person to person . Web 3.0 Architecture supports ubiquitous connectivity, network computing, open identity, intelligent web, distributed databases and intelligent applications . Some of the technologies which lead to the design and development of web 3.0 applications are Artificial intelligence, Automated reasoning, Cognitive architecture, Semantic web . An attempt is made to capture the requirements of Students inline with web 3.0 so as to bridge the gap between the design and development of web 3.0 applications and requirements among Students. Maximum Spanning Tree modeling of the requirements facilitate the identification of key areas and key attributes in the design and development of software products for Students in Web 3.0 using Discriminant analysis.

***Keywords** : Web 3.0, Discriminant analysis , Design and Development ,Model, Maximum Spanning Tree*

1. Introduction

Web 3.0 is an extension of www, in which the information can be shared and interpreted by other software agent to find and integrate applications to different domains. Web 3.0 provides integrated real time application environment to the user. The applications are involved in searching using semantic web, 3D web and are media centric. Web 3.0 supports pervasive components. Each component and its relations are represented below.

In web 3.0, web is transformed into database or Data Web wherein the data which are published in the web is reusable and can be queried. This enables a new level of data integration and application interoperability between platforms. It also makes the data openly accessible from anywhere and linkable as web pages do with hyperlinks. Data web phase is to make available structured data using RDF[1]. The scope of both structured and unstructured content would be covered in the full semantic web stage. Attempts will be to make it widely available in RDF and OWL semantic formats.

The driving force for web 3.0 will be artificial intelligence. Web 3.0 will be intelligent systems or will depend on emergence of intelligence in a more organic fashion and how people will cope with it. It will make applications perform logical reasoning operations through using sets of rules expressing logical relationships between concepts and data on the web. With the realization of the semantic web and its concepts web 3.0 will move into Service Oriented Architecture.

The evolution of 3D technology is also being connected to web 3.0 as web 3.0 may be used on massive scale due to its characteristics.

Web 3.0 is media centric where users can locate the searched media in similar graphics and sound of other media formats.

The pervasive nature of web 3.0 makes the users of web in wide range of area be reached not only in computers and cell phones but also through clothing, appliances, and automobiles.

2. Review of Literature

Claudio Baccigalupo and Enric Plaza discussed in the paper poolcasting : a social web radio architecture for Group Customization about Pool casting a social web radio architecture in which groups of listeners influence in real time the music played on each channel. Pool casting users contribute to the radio with songs they own, create radio channels and evaluate the proposed music, while an automatic intelligent technique schedules each channel with a group customized sequence of musically associated songs[2] . M.T.Carrasco Benitez discussed in the paper Open architecture for multilingual social networking about an open architecture for all the multilingual aspects of social networking. This architecture should be comprehensive and address well-trodden fields such as localization, and more advanced multilingual techniquesto facilitate the communication among users[3] .

Autona Gerber, Alta van der Merwe, and Andries Barnard discussed in the paper A functional Semantic web architecture about the CFL architecture which depicts a simplification of the original architecture versions proposed by Bernes-Lee as a result of the abstraction of required functionality of language layers. Gerber argues that an abstracted layered architecture for the semantic web with well defined functionalities will assist with the resolution of several of the current semantic web research debates such as the layering of language technologies [4].

Ferda Tartanoglu val'erie Issarny, Alexander Romanovsky and Nicole Levy discussed in the paper Dependability in the web services architecture which discusses about how to build dependable systems based on the web services architecture. It surveys base fault tolerance mechanisms and shows how they are adapted to deal with the specifics of the web in the light of ongoing work in the area[5]. Barry Norton, Sam Chapman and Fabio Ciravegna discussed in the paper developing a Service- Oriented Architecture to Harvest information for the Semantic web which discusses about the Armadillo architecture, how it is reinterpreted as workow templates that compose semantic web services and show how the porting of Armadillo to new domains, and the application of new tools, has been simplified[6].

3. Problem Definition

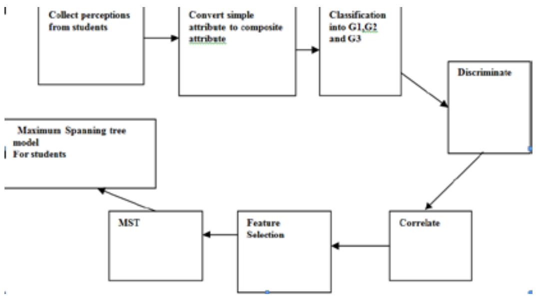
The Design and Development of web 3.0 products are on the course. Due to the existence of the ambiguity in the requirements of Students for structuring the web 3.0 products , bridging the gap between web 3.0 design and development fraternity and Students becomes need of the hour. The key factors for students are to be identified and their preference order is to be extracted.

Let G1, G2, G3 denote the three groups for design and development in web 3.0 . The problem is to find the order of preferences in the three groups and their attributes for Students by identifying the attributes v1 , v2 , ….. vn included and hence design Maximum Spanning Tree based model.

3. Materials and Methods

We collected the perceptions of students inline with web 3.0 attributes. A five point scale was followed which ranges from very low satisfaction , low satisfaction, Medium satisfaction, high satisfaction to very high satisfaction.

1. Block diagram of Web 3.0 Maximum Spanning Tree Model



4. Algorithm in Web 3.0 Spanning Tree modeling

a. Start
b. Collect the perceptions regarding the attributes of web 3.0 for students.
c. Combining simple attributes v1,v2…vn into composite attribute p1,p2,….pk.
d. Classification of composite attributes p1,p2,…pk into three groups G1, G2 and G3.
e. Discriminant Modeling for the Students
f. Correlation Coefficient among the composite attributes p1,p2….pn in groups G1, G2
g. Feature selection on composite attributes
h. Maximum Spanning Tree Algorithm.
i. Design maximum spanning tree model for students

Preprocessing

The data collected are verified for completeness. The missing values are replaced with the mean value.

Transforming from Simple attributes to composite attributes

Data's are collected based on the attributes. Some attributes are composite which are tabulated below.

Table 1 :Simple to Composite Attributes in Media

S.No	***Simple Attribute***	***Composite Attribute***
1.	2D Videos , Screen Partial 2D and Partial 3D	2D
2.	3D text, 3D screen, 3D demo and 3D tutorial website	3D

Table 2 : Simple to Composite Attributes in Output

S.No	***Simple Attribute***	***Composite Attribute***
1.	Mash up, Mash up of results, Mash up of tutorial websites, Mash up of Social networking sites	Result as Mash up

5.3 Classification

The data's are collected from the Students based on the attributes 2D, 3D, Audio, Custom mash up, E decisions, Multilingual, Result as Mash up, Semantic Maps, Semantic Wiki, Software Agents, Speech recognition. Based on the functionality, the attributes they are grouped into G1 , G2 and G3. G1 comprises of Multilingual, Semantic maps, Edecisions, Semantic wiki and Software agents . G1 is termed as Applications . G2 comprises of 3D, Audio, 2D and Speech recognition. G2 is termed as Media. G3 comprises of Custom Mash up, Result as Mash up . G3 is termed as Output.

5.4 Discriminant modeling on groups:

Table 3 : Classification Function Coefficients for Students

Group	*Students*
Applications	14.048
Media	9.374
Output	8.074
Constant	-46.475

The order of preferences for the three groups are given below based on the above Classification Function Coefficients.

Order of Preferences for Students

STUDENTS	APPLICATIONS	Multilingual (P1) Semantic map(P2) Semantic Wiki (P3) Edecisions (P4) Software agents (P5)
	MEDIA	2D (P6) 3D (P7) Speech recognition (P8) Audio (P9)
	OUTPUT	Custom mash up (P10) Result as mash up (P11)

From the above table the design and development of web 3.0 products specifically related to Students, can ensue the preference orders and attributes . The products can be designed with the maximum attributes in the first group preference followed by lesser attributes in the second and third group.

5.5 Correlation Coefficient between all pairs of composite attributes

The correlation coefficient for all pairs among the Groups are calculated using the following formula.[7]

Correlation(r) =[NΣXY - (ΣX)(ΣY) / Sqrt([NΣX2 - (ΣX)2][NΣY2 - (ΣY)2])]
where

N = Number of values or elements
X = First Score
Y = Second Score
ΣXY = Sum of the product of first and Second Scores
ΣX = Sum of First Scores
ΣY = Sum of Second Scores
ΣX2 = Sum of square First Scores
ΣY2 = Sum of square Second Scores

sample correlation coefficient of G3 (output) for Students

Table 4 : Correlation Coefficient

S.No	*Source*	*Destination*	*Correlation Coefficient*
1.	p10	p11	.139

5.6 Feature Selection on Composite Attributes:

The Attribute pairs which have positive correlation are selected. The Attribute pairs which have negative correlation are removed.

5.7 Maximum Spanning Tree Algorithm

A spanning tree of an undirected graph of n nodes is a set of n − 1 edges that connects all nodes. This note develops two algorithms for finding the minimum spanning tree. Properties of spanning trees In a spanning tree:

• There is no cycle: a cycle needs n edges.

• There is exactly one path between any two nodes: there is at least onepath between any two nodes because all nodes are connected. Further,there is not more than one path between a pair of nodes because thenthere would be a cycle that includs both nodes.

• Adding a non-tree edge creates a cycle: Suppose a non-tree edge (x, y) is added to a spanning tree. Now there are two distinct paths between(x, y), the added edge and the path in the tree. Hence there is a cycle.

• Removing an edge from a cycle as above creates a spanning tree: after removal of the edge there are (n − 1) edges. All nodes of the graph are connected: suppose edge (x, y) is removed that belonged to the original graph. The nodes x, y are still connected because x, y were on a cycle. For other node pairs, in the path in the original graph replace the edge (x, y) by the path between x, y.

Kruskal's Algorithm

Let $G = (V, E)$ be the given graph, with $|V| = n$

{

Start with a graph $T = (V,\phi)$ consisting of only the vertices of G and no edges;

/* This can be viewed as n connected components, each vertex being one connected component */

Arrange E in the order of increasing costs;

for ($i = 1, i <= n - 1, i + +$)

{

Select the next biggest cost edge;

if (the edge connects two different connected components)

add the edge to T;

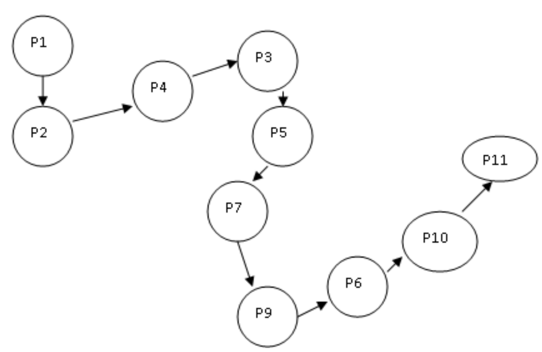
}

}

Order of preferences for Groups and Attributes for students :

Students	1. Applications	1.1 Multilingual (p1)	
		1.2 Semantic Maps (p2)	0.297
		1.3 E decisions (p4)	0.273
		1.4 Semantic Wiki (p3)	0.268
		1.5 Software agents (p5)	0.206
	2. Media	2.1 3D (p7)	
		2.2 Audio (p9)	0.49
		2.3 2D (p6)	0.143735
		2.4 Speech recognition (p8)	
	3. Output	3.1 Custom Mash up (p10)	
		3.2 Result as Mash up (p11)	0.139
		Total cost	1.816735

Maximum Spanning Tree model for Students



7.0 Conclusion

The perceptions inline with web 3.0 products are collected from students. The data's are preprocessed , classified, Mean, Standard deviation and correlation coefficient are computed to understand the descriptive and Discriminant modeled. An model for Students based on Maximum Spanning Tree is designed . At the outset of evolving growth in Web 3.0 this model is an initiative for the of web 3.0 product design for Students .

References

[1]. *e*n.wikipedia.org/wiki/Resource_Description_Framework

[2] Baccigalupo C.; Plaza E.; Poolcasting: a Social Web Radio Architecture for Group Customisation.Computer.org 2007.http://portal.acm.org/citation.cfm?id=1332471.1332755.

[3].]. Carrasco Benitez M.T;, Open architecture for multilingual social networking. Position paper for the W3C workshop on the Future of Social Networking [FSN], 15-16 January 2009.

[4]. . Aurona Gerber; Alta van der Merwe ; Andries Barnard; A functional semantic architecture.

http://www.eswc2008.org/final-pdfs -for-web-site/fisr-1.pdf.

[5]. Ferda Tartanoglu val'erie Issarny; Alexander Romanovsky and Nicole Levy er; Dependability in the web services architecture. 2007 .http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.14.771&rep=rep1&type=pdf.

[6]. Barry Norton; Sam Chapman ; Fabio Ciravegna; Developing a Service- Oriented Architecture to Harvest information for the Semantic web .2005.

[7]. http://easycalculation.com/statistics/learn-correlation.php

[8]. http://www.statpower.net/Content/312/Lecture%20Slides/Discriminant%20Analysis.pdf

[9]. Production of Cross Media Content for Multi-Channel Distribution, 2007, AXMEDIS apos;07 Third International Conference Volume ,Issue , 28-30 Nov . 2007 Page(s):115-122.

[10] Mantovani.F; VR Learning : Potential and Challenges for the Use of 3D Environments in Education and Traning . 2001.

[11]. Andr.P.Freire; Renata; P.M.Fortes ;An Evaluation of Web Accessibility Metrics based on their Attributes .2008.

[12]. Nuno Laranjeiro; Towards Fault Tolerance in Web Services Compositions .2007.

[13]. Dhiraj Joshi ; Ritendra Datta ; Ziming Zhuang ;PARAgrab : A Comprehensive Architecture for Web Image .2006.

[14]. Antonio Tapiador; Antonio Fumero, Joaqu'm Salvach'ua; Sandra Aguirre; A Web Collaboration Architecture .2006.

[15]. http://www.cs.utexas.edu/users/misra/Notes.dir/MinSpanningTree.pdf

AUTHORS PROFILE

S.Padma , is a research scholar in Bharathiar university ,Coimbatore. She has published 2 international journals . Her area of interest is web mining.

Dr. Ananthi Seshasaayee received her Ph.D in Computer Science from Madras University. At present she is working as Associate professor and Head, Department of computer science, Quaid-e-Millath Government College for Women, chennai. She has published 16 international journals. Her area of interest involve the fields of Computer Applications and Educational technology.

ZigBee Based Electric Meter Reading System

[1]S.Arun , [2]R.Krishnamoorthy and [3] Dr.VenuGopala Rao.M

[1]Research Scholar,Singhania University & Asst.Professor
Veltech High Engg College,Chennai

[2]Veltech Dr.RR & Dr.SR Technical University
Chennai

[3]Professor,EEE Dept KL University,
Vijayawada, A.P ,India

Abstract

According to the market requirements of Electric Meter. Nowadays, the system will use ZigBee and GSM system for communication protocol. The ZigBee is used since the application don't need high speed data rate, need to be low powered and low cost. Presenting the remote wireless Electric Meter Reading System, this aims at resolving the shortcomings of the technology of the traditional Electric Meter Reading, combining the characteristics of the ZigBee technology and IEEE802.15.4 standard. The hardware implementation was designed, and then analyzed the use cases for Electric Meter.

Keywords-ZigBee; IEEE802.15.4; Market Requirements

1. INTRODUCTION

Automatic Electric Meter reading is one method reading and processing data automatically with computer and communication. It is the need of improving the automatic level of energy consumption and the necessity of rapid development of computer and communication technology too. It not only may relieve reading person's labor intensity, reduce the reading mistake, but also has the advantage of high speed and good real-time. With the project of the wireless Electric Meter reading for wireless communication technology, complete the design of automatic Electric Meter reading system. Through researching the characteristic of main wireless communication protocol, ZigBee is chosen as lower layer communication protocol. With these applications, the standard is optimized for low data rate, low power consumption, security and reliability. Here describes the functional requirements to solve the technical issues related to the market applications.

2. IEEE802.15.4 AND ZIGBEE

2.1. IEEE 802.15.4

The ZIgBee protocol stack is described in Figure 1 .As we can see, IEEE 802.15.4 and the ZigBee network are tightly coupled to provide the consumer standardization for low-power and low-rate wireless communication devices.IEEE802.15.4 PHY layer provides 16 channels for ISM 2.4GHz,10 channels for ISM 900 MHz, and 1 channel for 868 MHz IEEE 802.15.4 PHY provides LOI(Link Quality Indicator) in order to characterize the quality of links between nodes, as well as data transmission and reception, IEEE 802.15.4 MAC uses the Carrier Sense Multiple Access with Collision Avoidance(CSMA/CA) mechanism for accessing the channel, like other wireless networks such as IEEE802.11 and IEEE 802.15.3 . There are two variations: Beacon Enabled Network which uses the CSMA/CA. Moreover, it provides the GTS (Guaranteed Time Slots) allocation method in order to provide real time data communication.

2.2 ZigBee

Based on IEEE 802.15.4 PHY/MAC , the ZigBee network layer provides functionality such as dynamic network formation, addressing, routing, and discovering 1 hop neighbors . The size of the network address is 16 bits , so ZigBee is capable to accept about 65535 devices in a network , an the network adderss is assigned in a hierarchical tree structure . ZigBee provides not only star topology, but last mesh topology. Since any device can communicate with other devices except the PAN Coordinator, the network has high scalability and flexibility. Besides, the self-formation and self-healing features makes ZigBee more attractive , The deployed ZigBee devices automatically construct the network, and then changes such as joining/leaving of devices are automatically reflected in the network configuration.

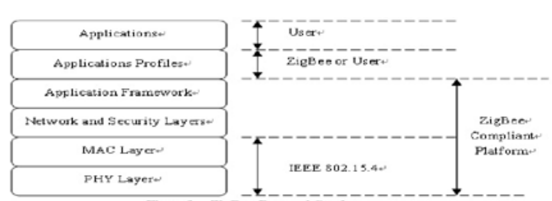


Figure1 ZigBee Protocol Stack

3. THE MARKET REQURIREMENTS

3.1 Market Needs

The utilities and Electric Metering companies continually look for improved methods to support their day to operations, which include: Providing flexible billing dates for customers, Performing Monthly/Cycle billing reads, Implementing Time-of-Use billing, Capturing Peak Demand, Supporting Critical Peak Pricing events, Forecasting energy usage, Positive outage and restoration detection and notification., Theft detection, Remote connect and validation, Market advanced Electric Metering and billing programs

3.2 Market Analyses

Within the typical ZigBee network there is a single “owner” or “stakeholder.” This owner can determine which devices are allowed on the PAN by only sharing network keys with trusted devices. There may be two stakeholders for a single network: the utility and the end customer. Neither of these stakeholders necessarily trusts the other. The utility wants to be sure that the end customer cannot use ZigBee to inappropriately manipulate a load control or demand response system, or attack an energy service portal. The customers want to be sure that the energy service portal does not allow the utility to take liberties with their equipment or compromise their privacy. This results in four primary network ownership / deployment scenarios: utility-private, customer-private, shared, and bridged. Each of these scenarios has different implications. All of these scenarios are valid for EMI deployments, though their use may be specific to particular use cases or markets .

3.2.1 Utlity-Private

Utility Private HAN might include an in-home display, or a load control device working in conjunction with energy service portal, but it would not include any customer controlled devices.

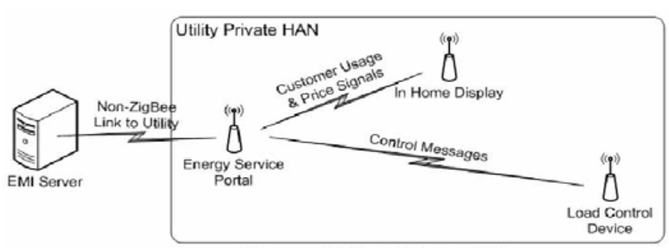


Figure 2 Utility Private HAN

3.2.2 Customer-Private

In the most extreme form, a customer private network might not even include an ESP on the ZigBee network, instead relying on some sort of customer provided device with non-ZigBee access to usage, consumption, and price data. Control messages in these examples would be one determined by the end customer, not the utility, and programmed into a home energy management console.

3.2.3 Customer and Utility Shared

The shared HAN represents the worst security scenario for an EMI deployment. Devices are on a network they cannot trust, with other devices they cannot trust. Application level authentication and authorization are required to support a shared network environment.

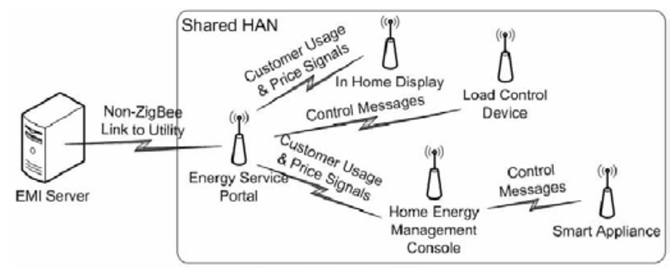


Figure 6 Shared HAN

3.2.4 Application-Linked

As an example, in the scenario below, the Utility HAN is made available strictly to utility controlled devices. The Home Energy Management Console is a utility approved device that also lives on a customer provided HAN. It can respond to EMI commands, as well as sending out HA commands to devices within the home.

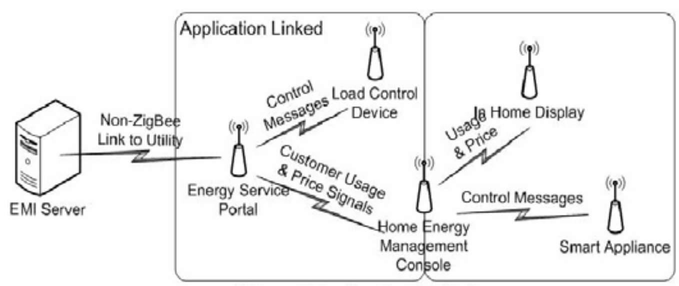


Figure 7 Application-Linked

4. DESIGN OF EMI

According the design for this system, the hardware design of EMI is divided to two parts: The Electric- Meter End Devices and The Data Acquisition Device. The former is to acquire the data of the Electric-Electric Meter, then transmit the data to the Data Acquisition Device through ZigBee network. Meanwhile display the energy and system time on the Electric Meter for customer. The latter functioned as a coordinator in the whole ZigBee network. Its function is to obtain all the information of the Electric Meters. And then transmit them to the energy management center through the parallel port

4.1Design for End Electric Meter Device

The CC2430 comes in three different versions: CC2430-F32/64/128, with 32/64/128 KB of flash memory respectively.

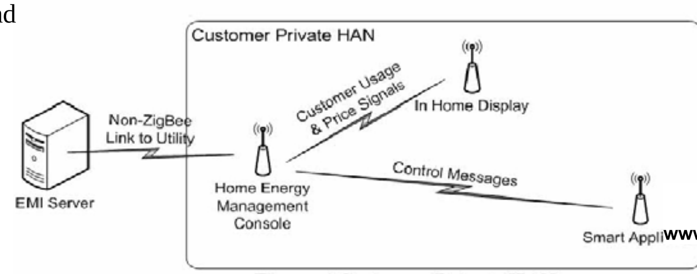


Figure 4 Customer Private HAN

The CC2430 is a true System-on-Chip (SoC) solution specifically tailored for IEEE 802.15.4 and ZigBee applications. It enables ZigBee nodes to be built with very low total bill-of-material costs. The CC2430 combines the excellent performance of the leading CC2420 RF transceiver with an industry-standard enhanced 8051 MCU, 32/64/128 KB flash memory, 8 KB RAM and many other powerful features. Combined with the industry leading ZigBee protocol stack (Z-Stack) from Figure 8 Wireless / Chipcon, the CC2430 provides the market most competitive ZigBee solution.

4.2 Designs for Data Acquisition Device

As the ZigBee node has to be used with Electric Meter Module, and it should be powered by battery, so the size of the ZigBee node is small, low-rate, and high-stability. Choose the small encapsulation designation circuit, and use the PCB as the wireless antenna. Make the bulk of the module to be minimized. Use the PIC18F4620 as the MCU, at the idle and sleeping state, it can minimize the power consumption of the system, choose the Chip on CC2430 which conforms to ZigBee protocol stack standard, and it needs a few external equipments, stable performance and the power consumption is low.

The interface circuit between PIC18F4620 and CC2430 is simple and the external equipments is fewer, simplify the difficulty of the debug, improve the stability of the system. In the addition of using the PCB , this system can communicate 60 miles.

The Data Acquisition Device reads the data from the Electric Meter timely, which read the impulse of the sensor, sending the data to the gather through the ZigBee communication module, till the Electric-Electric Meter data transport module read the data of this area.

5. SOLUTION USE CASES

The following sections describe the predominant areas of use cases for the EMI/EMR Market space, they are:_

1. Mobile EMR: Describes the market needs and the utilization of ZigBee to facilitate Electric Meter reading using mobile reading devices.

2. Energy Management: Provides the use cases that utilize the ZigBee based devices that support or enable EMS programs within premises.

5.1 Mobile EMR

Mobile EMR solutions consist of two scenarios, a Walk-By solution where Hand Held Computers are typically used to gather Electric Meter information, and a Drive-By solution where Computers used in conjunction with dedicated radios are installed in vehicles to remotely read Electric Meter information. Below are examples of both scenarios .

As depicted in the above diagram, a ZigBee based profile is used to transport the Electric Meter information to both the Walk-by and the Drive-by solutions. The types of Electric Metered information collected on a monthly basis ranges from simple Consumption to very complex Electric Metering including TOU (Time of Use), Load Profile (profile of consumption), Peak Demand.
The steps to accomplish this use case are:

- The CIS/MDMS requests the EMR solution to collect a series of Electric Meter reads. This may be for all Electric Meters or may be only for ones needed for that day business needs.

- The Electric Meter Reading Host Processor breaks the read requests into the appropriate routes for the individual Hand Held Computers or Vehicle based equipments.

- The Electric Meter Readers Proceed to collect the Electric Meter information along their designated routes.
- The Electric Meter information id uploaded to the Electric Meter Reading Host Processor and then forwarded upstream to the CIS/MDMS.

5.2 Energy Management Use Cases

ZigBee is to be utilized as the communication medium between home and building automation devices, Electric Metering devices, in-home displays, and fixed network devices such a gateways, bridges or access points. ZigBee based solutions for Energy Management should be capable of operating independently but in conjunction with current and future EMI solutions.

1) Utility Customer Reduces Load Voluntarily In Response to CPP (Critical Peak Pricing)

When the utility determines that the next day will be a Critical Peak Pricing (CPP) day and needs to invoke a voluntary load reduction program, it will notify its customer base of the impending event. The notification can occur using a variety of methods such as newspaper, TV, website, email, etc but may also include providing notice through the EMI solution via the Electric Metering device or customer display.

2) Utility *Customer Accesses Pricing Information*

Customers are becoming aware of the importance of understanding how much energy they are using and when it is being used. Customers want to understand how their energy consumption habits affect their monthly energy bills and to find ways to reduce their monthly energy costs. The utility and regulatory agencies also want customers to be aware of the energy they are consuming and associated costs. By providing customers better visibility to their energy usage and cost at their site, they can make more educated energy related decisions regarding participation in load reduction programs, be more inclined to install energy efficient equipment and potentially to change their energy consumption habits. EMI solutions will enable improved communications between the utility and its customers by making if possible to remotely transmit energy usage, cost and other related utility messages to the EMI solution and down to the customer display device within the home or business.

3) Utility *Customer Uses Prepayment Services*

Most utility customers pay for usage after the fact. The utilities would also like to provide customers the ability to prepay for their electric quantities. This would apply to purchasing power for a residence or commercial site.

6. CONCLUSION

ZigBee technology is a new wireless protocol that widely used various areas for it excellent performance in reliability, capability, flexibility and cost, ZigBee corresponds to a large market. This paper provides an application in the field of automatic Electric Meter Reading System. With the developments of the ZigBee technology and the communication network technology of computer, wireless Electric Meter Reading System will grow up and practical mostly.

7. REFERENCES

[1] ZigBee Alliance, ZigBee Specification Version 1.0, http:// www.ZigBee.org, December 14th, 2004

[2] Microchip Technology Inc. Microchip Stack for the ZigBee™

[3]

[4] Version 1.0,2005.

[5] ZigBee Document 075356r08ZB, Advanced Metering Initiative Profile Protocol Specification Version 08, November 13th, 2007.

[6] ZigBee Document 074994r08, ZigBee AMI Technical Requirements Document Version 08, september19th, 2007.

AUTHORS PROFILE

S.Arun was born on 18th october 1979 Tiruvanamalai**,** India. He is working as Assistant Professor in Vel Tech High Engneering College Chennai and pursuing Ph.D. in the School of Electronics and Electrical Engineering, Singhania University, Rajasthan, India. He obtained M.E. degree in Communication systems 2004, Anna University and B.E. degree in Electronics and Communications Engineering from University of Madras, Chennai, India in the year 2001. . He has also published papers on image processing, wireless communication in National and International Conferences. His current research interests are Robotics, wireless Communications.

R.Krishnamoorthy, Finished his B.E in Sree Sastha Institute of Science and Technology , Affiliated to Anna university in 2009, He presented papers in several national and international conferences .

Dr.Venu Gopala Rao finished his Phd in JNTU, Hyderabad in 2009, M.Tech in Electrical MSUniversity, Baroda, 1999, M.Tech in Computer Science in JNTUCE, Kakinada in 2004 and B.E in EEE in 1996 Distinction. Presently working as professor in KL University, Vijayawada, A.P , IndiainHe has published several papers in international journals High Voltage.

Entropy of Data Compression Using Object Oriented Data Warehousing

Dr. (Mrs Pushpa Suri)[1] and Mrs Meenakshi Sharma[2]

[1] Associate professor, Department of Computer Science and Applications, Kurukshetra University, Kurukshetra, India

[2] Assistant Professor, Department of Computer Science and Engineering, Haryana College of Technology and Management Kaithal, India

ABSTRACT

A data warehouse is a repository of subjectively selected and adapted operational data which can successfully answer any ad hoc, statistical, complex or analytical queries. Data warehousing technology is becoming essential for effective business intelligence, business strategy formulation and implementation in a globally competitive environment where in larger and larger amounts of data are required to be processed faster and faster for comprehension of its real meaning and impact [4].Data warehousing has been around in some form or another since the inception of data storage, people were never able to exploit the information that was wastefully sitting on a tape somewhere in a back room. Organizations across the country and around the world are seeking expertise in this exploding field of data organization and manipulation. It is not a surprise, really, that business users want to get a better look at their data. Today, business opportunities measure in days, instead of months or years, and the more information empowering an entrepreneur or other business person, the better the chances of beating a competitor to the punch with a new product or service. In this paper we have discussed the efficiency of the data warehousing packages so that less CPU time and less Memory is consumed.

Keywords: *Entropy, Data warehousing, object oriented, Data compression*

1. INTRODUCTION

A data warehouse is a mechanism for data storage and data retrieval. Data can be stored and retrieved with a multidimensional structure--hypercube or relational, a star schema structure or several other data storage techniques. The task of transitioning from a procedural mindset to an object-oriented paradigm can seem overwhelming; however, the transition does not require developers to step into another dimension or go to Mars in order to grasp a new way of doing things. In many ways, the object-oriented approach to development more closely mirrors the world we've been living in all along: We each know quite a bit about objects already. It is that knowledge we must discover and leverage in transitioning to object-oriented tools and methodologies. Our research has been from a different point of view – our primary motivating factor is to show how existing applications can be enhanced using object –oriented Technology. Like Many new ideas, object –oriented programming does not have a universally accepted definition [1, 2]. Ideas on the subject do, however, seem to be converging the "best" definition that we have seen to date is "object-oriented = object + classes + inheritance" [3]. OOP can also be defined as an extension of the idea of abstract data type. The task of transitioning from a procedural mindset to an object – oriented paradigm can seem overwhelming: however, the transition does not require developers to step into another dimension or go to grasp a new way of doing thing. In many ways, the object oriented approach to development more closely mirrors the world we've been living in all along. [8]. we each know quite a bit about objects already. It is that knowledge we must discover and leverage in transitioning to object- oriented tools and methodology.

2. ENTROPY IN DATA COMPRESSION

Data compression is of interest in business data warehousing, both because of the cost savings it offers and because of the large volume of data manipulated in many business applications. The types of local redundancy present in business data files include runs of zeros in numeric fields, sequences of blanks in alphanumeric fields, and fields which are present in some records and null in others.[1],[2] Run length encoding can be used to compress sequences of zeros or blanks. Null suppression may be accomplished through the use of presence bits. Another class of methods exploits cases in which only a limited set of attribute values exist. Dictionary substitution entails replacing alphanumeric representations of information such as bank account type, insurance policy type, sex, month, etc. by the few bits necessary to represent the limited number of possible attribute values. The problem of compressing digital data can be decoupled into two subproblems: modeling and entropy coding. Whatever the given data may represent in the real world, in digital form it exists as a sequence of symbols, such as bits. The modeling problem is to choose a suitable symbolic representation for the data and to predict for each symbol of the representation the probability that it takes each of the allowable values for that symbol. The entropy-coding problem is to code each symbol as compactly as possible, given this knowledge of probabilities. (In the realm of lossy compression, there is a third subproblem: evaluating the relative importance of various kinds of errors.)

For example, suppose if it is required to transmit messages composed of the four letters a, b, c, and d. A straightforward scheme for coding these messages in bits would be to represent a by \00", b by \01", c by \10" and d by \11". However, suppose if it is known that for any letter of the message (independent of all other letters), a occurs with probability .5, b occurs with probability .25, and c or d occur with probability .125 each. Then a shorter representation might be chosen for a, at the necessary cost of accepting longer representations for the other letters. a could be represented by \0", b by \10", c by \110", and d by \111". This representation is more compact on average than the first one; indeed, it is the most compact representation possible (though not uniquely so). In this simple example, the modeling part of the problem is determining the probabilities for each symbol; the entropy-coding part of the problem is determining the representations in bits from those probabilities; the probabilities associated with the symbols play a fundamental role in entropy coding. One well-known method of entropy coding is Huffman coding, which yields an optimal coding provided all symbol probabilities are integer powers of .5. Another method, yielding optimal compression performance for any set of probabilities, is arithmetic coding. In spite of the superior compression given by arithmetic coding, so far it has not been a dominant presence in real data-compression applications. This is most likely due to concerns over speed and complexity, as well as patent issues; a rapid, simple algorithm for arithmetic coding is therefore potentially very useful. An algorithm which allows rapid encoding and decoding in a fashion akin to arithmetic coding is known as the Q-coder. The QM-coder is a subsequent variant. However, these algorithms being protected by patents, new algorithms with competitive performance continue to be of interest. The ELS algorithm is one such algorithm

The ELS-coder works only with an alphabet of two symbols (0 and 1). One can certainly encode symbols from larger alphabets; but they must be converted to a two-symbol format first. The necessity for this conversion is a disadvantage, but the restriction to a two-symbol alphabet facilitates rapid coding and rapid probability estimation.

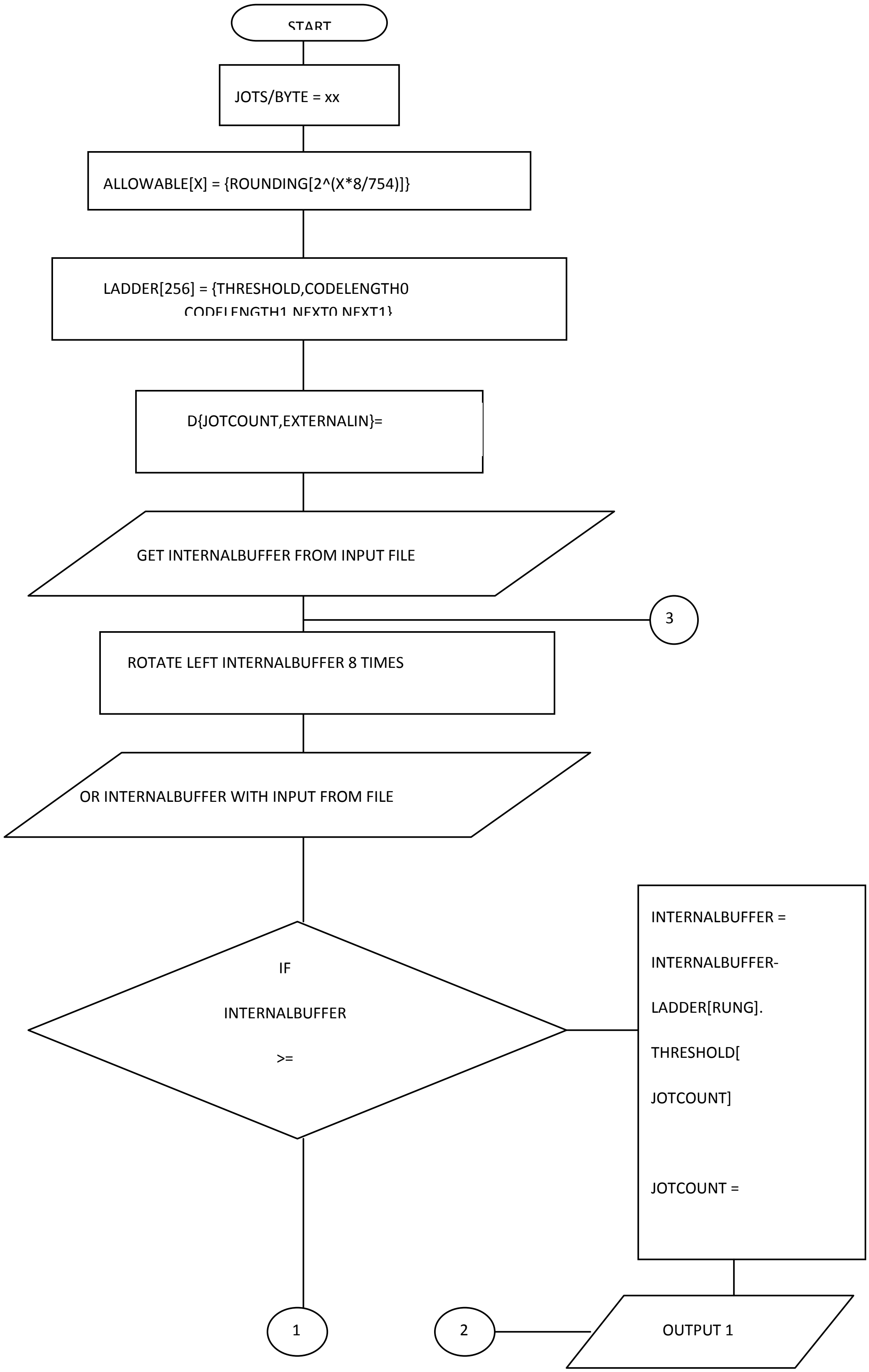
The ELS-coder decoding algorithm has already been described. The encoder must use its knowledge of the decoder's inner workings to create a data stream which will manipulate the decoder into producing the desired sequence of decoded symbols. As a practical matter, the encoder need not actually consider the entire coded data stream at one time. One can partition the coded data stream at any time into three portions; from end to beginning of the data stream they are: preactive bytes, which as yet exert no in seuence over the current state of the decoder; active bytes, which affect the current state of the decoder and have more than one consistent value; and postactive bytes, which affect the current state of the decoder and have converged to a single consistent value. Each byte of the coded data stream goes from preactive to active to postactive; the earlier a byte's position in the stream, the earlier these transitions occur. A byte is not actually moved to the external _le until it becomes postactive. Only the active portion of the data stream need be considered at any time. Since the internal buffer of the decoder contains two bytes, there are always at least two active bytes. The variable backlog counts the number of active bytes in excess of two. In theory backlog can take arbitrarily high values, but higher values become exponentially less likely.

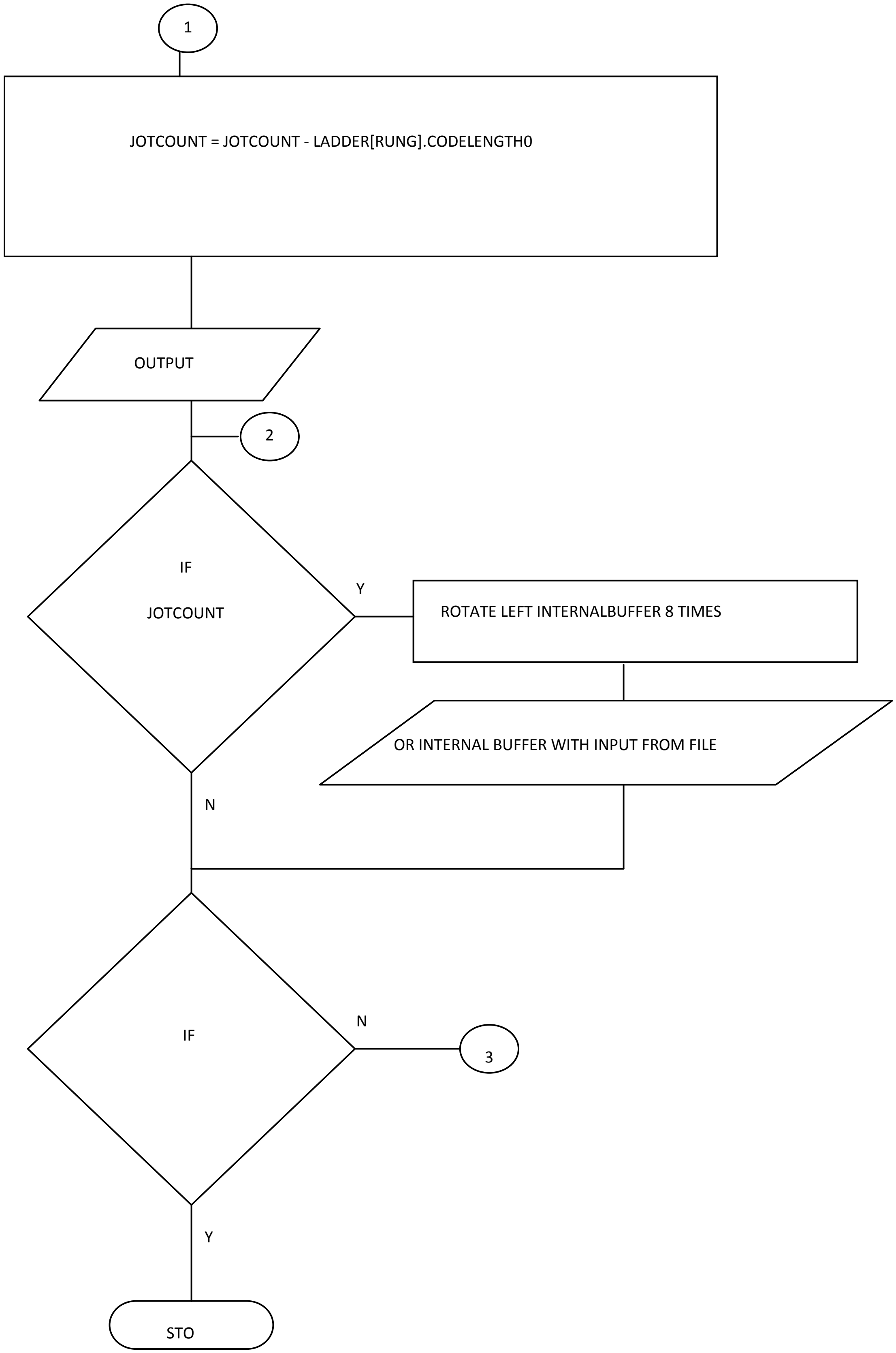
3. METHODOLOGY

Following steps will be taken for the future work

1. Creation of different sizes of databases in oracle
2. Employment of object oriented programming for compression using datawarehousing
3. Further compression of database csv files using C++
4. Comparison of time taken and compression efficiency for different sizes of databases.

4. FLOW CHART OF DATA COMPRESSION


START
JOTS/BYTE = xx
ALLOWABLE[X] = {ROUNDING[2^(X*8/754)]}
LADDER[256] = {THRESHOLD,CODELENGTH0
CODELENGTH1 NEXT0 NEXT1}
D{JOTCOUNT,EXTERNALIN}=
GET INTERNALBUFFER FROM INPUT FILE
3
ROTATE LEFT INTERNALBUFFER 8 TIMES
OR INTERNALBUFFER WITH INPUT FROM FILE
IF
INTERNALBUFFER
>=
INTERNALBUFFER =
INTERNALBUFFER-
LADDER[RUNG].
THRESHOLD[
JOTCOUNT]
JOTCOUNT =
1
2
OUTPUT 1


1
JOTCOUNT = JOTCOUNT - LADDER[RUNG].CODELENGTH0
OUTPUT
2
IF
JOTCOUNT
Y
ROTATE LEFT INTERNALBUFFER 8 TIMES
OR INTERNAL BUFFER WITH INPUT FROM FILE
N
IF
N
3
Y
STO

5. CONCLUSION

A data warehouse is an essential component to the decision support system. The traditional data warehouse provides only numeric and character data analysis. But as information technologies progress, complex data such as semi-structured and unstructured data become vastly used[2],[3]. Data Compression is of interest in business data warehousing, both because of the cost saving it offers and because of the large volume of data manipulated in many business application. The entropy is used in many areas such as image processing, document images. But in our research we used the entropy in object oriented data warehousing. Creation of different sizes of databases in oracle. Employment of object oriented programming for compression using datawarehousing. Further compression of database .csv files using C++. Comparison of time taken and compression efficiency for different sizes of databases.

6. REFERENCES

1. Wei-Chou Chen; Tzung-Pei Hong; Wen-Yang Lin; , "Using the compressed data model in object-oriented data warehousing ," Systems, Man, and Cybernetics, 1999. IEEE SMC '99 Conference Proceedings. 1999 IEEE International Conference on , vol.5, no., pp.768-772 vol.5, 1999
2. Wei-Chou Chen; Tzung-Pei Hong; Wei-Yang Lin; , "A composite data model in object-oriented data warehousing," Technology of Object-Oriented Languages and Systems, 1999. TOOLS 31. Proceedings , vol., no., pp.400-405, 1999
3. Shieh, J. C.; Lin, H. W.,"The Novel Model of Object-Oriented Data Warehouses", Workshop on Databases and Software Engineering, 2006
4. Chen Wei-Chou, Hong Tzung-Pei (2), Lin Wen-Yang "Three maintenance algorithms for compressed object-oriented data warehousing"
5. Boqiang Huang; Yuanyuan Wang; Jianhua Chen; , "2-D Compression of ECG Signals Using ROI Mask and Conditional Entropy Coding," Biomedical Engineering, IEEE Transactions on , vol.56, no.4, pp.1261-1263, April 2009
6. Gong, Y.; Fan, M.K.H.; Huang, C.-M.; , "On entropy-constrained residual vector quantization design," Data Compression Conference, 1999. Proceedings. DCC '99 , vol., no., pp.526, 29-31 Mar 1999
7. De, I.; Sil, J.; , "Wavelet entropy based no-reference quality prediction of distorted/decompressed images," Computer Engineering and Technology (ICCET), 2010 2nd International Conference on , vol.3, no., pp.V3-245-V3-250, 16-18 April 2010
8. De, I.; Sil, J.; , "ANFIS tuned no-reference quality prediction of distorted/decompressed images featuring wavelet entropy," Computer Information Systems and Industrial Management Applications (CISIM), 2010 International Conference on , vol., no., pp.126-131, 8-10 Oct. 2010
9. Liu, L.; Dong, Y.; Song, X.; Fan, G.; , "An entropy based segmentation algorithm for computer-generated document images," Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on , vol.1, no., pp. I- 541-4 vol.1, 14-17 Sept. 2003
10. Tu, C.; Tran, T.D.; , "Context-based entropy coding of block transform coefficients for image compression," Image Processing, IEEE Transactions on , vol.11, no.11, pp. 1271- 1283, Nov 2002
11. Chen, S.; Reif, J.H.; , "Using difficulty of prediction to decrease computation: fast sort, priority queue and convex hull on entropy bounded inputs," Foundations of Computer Science, 1993. Proceedings., 34th Annual Symposium on , vol., no., pp.104-112, 3-5 Nov 1993
12. Sang Hyun Kim; Rae-Hong Park; , "A novel approach to scene change detection using a cross

entropy ," Image Processing, 2000. Proceedings. 2000 International Conference on , vol.3, no., pp.937-940 vol.3, 2000

13. Scales, A.; Roark, W.; Kossentini, F.; Smith, M.J.T.; , "Lossless Compression Using Conditional Entropy-Constrained Subband Quantization," Data Compression Conference, 1995. DCC '95. Proceedings , vol., no., pp.498, 28-30 Mar 1995
14. Jegou, H.; Guillemot, C.; , "Entropy coding with variable length re-writing systems," Information Theory, 2005. ISIT 2005. Proceedings. International Symposium on , vol., no., pp.1529-1533, 4-9 Sept. 2005
15. Hua Xie; Ortega, A.; , "Entropy- and complexity-constrained classified quantizer design for distributed image classification," Multimedia Signal Processing, 2002 IEEE Workshop on , vol., no., pp. 77- 80, 9-11 Dec. 2002

Establishing Relationship between Complexity and Faults for Object-Oriented Software Systems

Usha Chhillar[1], Sucheta Bhasin[2]

[1] Department of Computer Science, Kurukshetra University, Kurukshetra, Haryana, India

[2] Department of Computer Science, Kurukshetra University, Kurukshetra, Haryana, India

Abstract

Controlling and minimizing software complexity is the most important objective of each software development paradigm because it affects all other software quality attributes like reusability, reliability, testability, maintainability etc. For this purpose, a number of software complexity measures have been reported to quantify different aspects of complexity. Complexity and fault proneness are two prominent parameters for improving quality of the software . The software industry is continuously facing the challenges of growing complexity of software and increased level of data on defects. To control the complexity and faults is one of the major challenges for researchers to predict different parameters which are responsible for increasing complexity and fault proneness. In this paper, faults prediction through bebugging technique has been tried through an experiment applied to C++ programs and compared the results with various object-oriented complexity measures. The results have been found encouraging. Relationship between faults and complexity has also been established.

Keywords: *Reusability, Reliability, Testability, Maintainability, Fault Proneness, Faults Prediction, Bebugging.*

1. Introduction

From time to time, various complexity metrics have been designed in an attempt to measure the complexity of software systems. Software complexity directly affects maintenance activities like software reusability, understandability, modifiability and testability. Estimates suggest that about 50 to 70 % of annual software expenditure involve maintenance of existing systems. Predicting software complexity and faults can save millions in maintenance [1,7,9,10,18]. Clearly, if complexities could somehow be identified and measured, then software developers could adjust development, testing and maintenance procedures and effort accordingly. This concern has motivated several researchers to define and validate software complexity measures and establish relationship between software complexity and faults [1, 2, 3, 5, 7, 16, 19, 20, 21]. It is accepted by both software developers and researchers that complexity of software can be controlled more effectively through object-oriented approach than traditional function-oriented approach. It is because that objected-oriented paradigm controls complexity of a software system by supporting hierarchical decomposition through both data and procedural abstraction [9]. But, the complexity of software is an essential attribute, not an accidental one [6]. Traditional software complexity metrics are not appropriate for object-oriented software systems due to their distinguish features like class, inheritance, polymorphism, coupling, and cohesion.

In this paper, faults prediction through bebugging technique has been tried through an experiment applied to C++ programs and compared the results with various object-oriented complexity measures. The results have been found encouraging. Relationship between faults and complexity has also been established.

Rest of this paper is organized as follows: Section 2 presents overview of software complexity and existing complexity measures. Faults prediction through bebugging is explained in section 3. Section 4 describes experiment design for faults prediction. Section 5 discusses the experimental results. Finally, section 6 concludes the paper with directions for future work.

2. Overview Of Software Complexity And Existing Complexity Measures

2.1 Software Complexity

In literature, software complexity has been defined differently by many researchers. Zuse [11] defines software complexity as the difficulty to maintain, change and understand software. It deals with the psychological complexity of programs. According to Henderson-Sellers [12] the cognitive complexity of software refers to those characteristics of software that affect the level of resources used by a person performing a given task on it. Basili [4] defines software complexity as a measure of the resources expended by a system while interacting with a piece of software to perform a given task. Here, interacting system may be a machine or human being. Complexity is defined in terms of execution time and storage required to perform the computation when computer acts as an interacting system. In case of human being (programmer) as an interacting system, complexity is defined by the difficulty of performing tasks such as coding, testing, debugging or modifying the software. Bill Curtis [13] has reported two types of software complexity – Psychological and Algorithmic. Psychological complexity affects the performance of programmers trying to comprehend or modify a class/module whereas algorithmic or computational complexity characterizes the run-time performance of an algorithm. Brooks [6] states that the complexity of software is an essential attribute, not an accidental one. Essential complexity arises from the nature of the problem and how deep a skill set is needed to understand a problem. Accidental complexity is the result of poor attempts to solve the problem and may be equivalent to what some are calling complication. Implementing wrong design or selecting an inappropriate data structure adds accidental complexity to a problem.

Software complexity can not be defined by a single definition because it is multidimensional attribute of software. So, different researchers/users have different view on software complexity. Therefore, no standard definition exits for the same in literature. However, knowledge about software complexity is useful in many ways. It is indicator of development, testing, and maintenance efforts, defect rate, fault prone modules and reliability. Complex software/module is difficult to develop, test, debug, maintain and has higher fault rate.

2.2 Software Complexity Measures

Software complexity can not be removed completely but can be controlled only. But, for effective controlling of complexity, we need software complexity metrics to measure it. From time to time, many researchers have proposed various metrics for evaluating, predicting and controlling software complexity. Halstead's software science metrics, McCabe's cyclomatic number and Kafura's & Henry's fan-in, fan-out are the best known early reported complexity metrics for traditional function-oriented approach [16,17,18]. But these metrics do not consider object oriented features of software for measuring the complexity of software. So traditional software complexity metrics are not suitable for measuring complexity of object oriented software.

Various researchers have proposed many object oriented metrics to compute complexity of object oriented software. Chidamber and Kemerer [1] proposed a suite of six metrics : Number Of Children (NOC) - number of immediate derived classes, Depth Of Inheritance Tree (DIT) - maximum path length from root to node in inheritance tree, Weighted Methods per Class (WMC) - sum of all methods of a class, Coupling Between Objects (CBO) - number of classes to which a class is coupled, Lack Of Cohesion in Methods (LCOM) - measures the dissimilarity of methods in a class and Response For a Class (RFC) - number of methods of a class to be executed in response to a message received by an object of that class. These metrics measure complexity of object-oriented software by using design of classes. WMC measures the complexity of a class as a sum of complexity of individual methods. Higher values of NOC and DIT are indicator of higher complexity due to involvement of many methods. CBO value for a class is the indicator of total number of other classes to which it is coupled. Mishra [14] proposed a metric for computing the complexity of a class at method level by considering internal structure of method. Fothi et al [8] designed a metric which computes complexity of a class on the basis of complexity of control structures, data and relationship between data and control structures. A metric which calculates overall complexity of design hierarchy was proposed by Mishra [14]. It computes complexity by considering inherited methods only and does not take into account internal characteristics of methods.

3. Faults Prediction Through Bebugging

The process of finding and rectifying faults in a program is called debugging. Bebugging is the reverse of debugging. In bebugging , a fixed number of artificial bugs are introduced in a source program. The complete detail of these artificial bugs is kept for identifying and removing the same from the source program after the experiment. By applying bebugging method, we may predict how many faults are still present in the source code and thus in a software system. Suppose P is the source program in which we want to predict the number of faults present through bebugging method.

Let
I = Number of artificial faults introduced in program P.

T = Total number of faults find out by the reviewer or a tester in program P.
R = Number of faults find out from I.
(T-R) = New additional faults find out
Total number of predicted faults (PF) in the program P :

$$PF = ((T - R) / R) \times I$$

The bebugging method is generally used by Zoologist for estimating the number of fish in a tank.

For example, take a sample of 100 fish from a pond. Mark them and put them back into the pond for mixing them with the total population of the tank. Again take the sample of 100 fish and find how many marked fish are in this sample. Let marked fish are 10. According to the bebugging method, there are 900 fish in the pond. In this method , we assume that the original sample was random and remixing of fish was homogenous.

Similarly, if we insert 5 bugs in a program and reviewer reports total 9 faults through bebugging process. Let out of these 9 bugs, 3 bugs are out of 5 bugs inserted by us.
Then
I = 5, T = 9, R = 3, T-R = 6.
PF = ((T-R) /R)*I = 10
It means predicted number of faults present in the program are 10.

4. Experiment Design

In this experiment, the main objective is to predict number of faults in a program by using the bebugging method described in the previous section and also to analyze the effect of faults on the complexity. It is intuitive that a programmer finds lesser number of faults in a complex program than a simple program in a given time period.

For this purpose, an experiment was conducted at the end of the academic session by involving 15 MCA fourth semester students on scheduled date and time. For this purpose, 10 programs written in C++ language were used. In each of these programs, five logical and syntax errors are inserted knowingly called artificial bugs. The purpose of the experiment was explained well to the students before conducting the experiment and they were asked to find out logical and syntax errors as many as they can. The experiment was conducted in 10 continuous sessions of 10 minutes duration each. After each session, the sheets of the program specified for the session were collected from the students . In this way 10 different programs were given to the students in 10 different sessions. One sample program has been given in table 2.

5. Experimental Results

For all the 10 programs used in the experiment, the number of predicted faults (PF) were calculated by using above mentioned bebugging technique for each of the 15 students (S1-S15) and results are tabulated in table 1. One sample program has been given in table 2. Type and description of errors inserted in sample program are described in table 3. We have also calculated three Chidamber and Kemerer's (CK) metrics – WMC, NOC, DIT, McCabe's complexity measure V(G), lines of code (LOC) metric and composite weighted complexity metric (CWT) for the programs studied here and results are shown in table 4 [1,5,15,17,18].

We also analyzed the relationship between complexity and predicted number of faults through bebugging method. For this purpose, we have drawn bar graphs among complexity metrics and faults described in table 4. These bar graphs have been given in figures 1-6. The bar graphs clearly show that fault rate is directly proportional to complexity i.e. more complexity implies more possibility of faults and hence less quality . However, the results of program no 9 and 10 vary because in these two programs coupling is the dominating factor . Due to more coupling, number of faults are more where as the value of complexity measures WMC, NOC, DIT,V(G) and LOC are less for these two programs. From this it is clear that coupling plays major role for increasing the complexity and reducing the quality of programs/software. So, it should be controlled to minimum to develop a good quality software.

Table 1: Students wise Experimental Predicted Faults (PF)

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
S1	0	0	3.3	0	3.3	0	10	3.3	25	0
S2	1.5	3.3	0	1.5	0	7.5	5	0	10	1.25
S3	1.7	0	2.5	1.7	1.7	0	0	0	5	0
S4	0.5	2.5	0	0.5	0	2.5	0	0	1.5	2.5
S5	2.5	2.5	0.5	0	0.5	2.5	0	5	5	20
S6	0	0	5	0	0	2.5	2.5	0	5	7.5
S7	0.5	0	2.5	0	0	0	0	0	2.5	1.25
S8	1.7	1.0	0.1	1.7	0	1.25	0	0	1.25	1.7
S9	20	1.5	1.5	0	0.5	2.5	20	5	5	6.5
S10	0	0.5	0.5	0	0.5	10	0	5	5	3.3
S11	5	0	1.5	0	6.5	0	10	6.5	10	1.7
S12	0	1.7	3.3	0	0	15	2.5	0	0	2.5
S13	0	2.5	1.7	5	0	1.7	3.3	0	7.5	3.3
S14	0	2.5	2.5	0	0	0	0	0	1.5	5
S15	0	0	2.5	0	1.7	2.5	0	1.7	5	0
Total	33.4	18.0	27.4	10.4	14.7	47.95	53.3	26.5	89.25	56.50

Table 2: Sample program

Program : To find out the greatest number out of three number using single inheritance

```
   #include<iostream.h>
   #include<conio.h>
01 Class Abc
02 {
03  protected:
04   int a ,b ,c;
05   public:
06   void input();
07   void output();
08  };
09 void Abc:: input()
10 {
11   cout<< enter the value  ;
12   cin >>a>>b
13   }
14 void Abc:: output()
15 {
16   cout<<" a ="<<a;
17   cout<< "b="<<b;
18  cout<<"c="<<c;
19 }
20 class xyz : private Abc
21 {
22   public:
23  void  greatest();
24 };
25 void greatest :: greatest()
26 {
27   input();
28  if (a>b)
29   if(a>c)
30 {
31   cout<<"a is the greatest number" ;
32   }
33   else{
34         cout<<" c is the greatest number" ;
35         }
36 }
37  else
38   if(b>c)
39    {
40     cout <<" b is the greatest number" ;
41     }
42  else;
43    {
44     cout <<" c is the  greatest   number";
45     }
46 void  main()
47   {
48     xyz a;
49     clrscr();
50     a.output();
51     a.greatest();
52      getch();
53     }
```

Table 3: Description of errors in sample program

LINE NO	ERROR TYPE	ERROR DESCRIPTION
11	Syntax	Undefined symbol 'enter' due to" " is missing in cout statement
12	Logical	Always c is the greatest number because the value of c is not read.
28	Syntax	Declaration terminated incorrectly due to '{' brace missing.
42	Syntax	else is terminated with ';'.
51	Syntax	abc::out not accessible due to abc class is inherited in private mode

Table 4: Values of complexity measures and errors

P.No	LOC	VG	WMC	NOC	DIT	CWT	Errors
P1	20	4	1.5	0.5	0.5	153	33.4
P2	16	2	1	0.67	0.33	112	18.0
P3	24	4	1.6	0.67	1	285	27.4
P4	11	2	1	0.5	0.5	69	10.4
P5	16	2	1.5	0.5	0.5	70	14.9
P6	35	6	1	0.67	0.33	303	47.95
P7	39	10	1.3	1	1	645	53.3
P8	24	4	1.6	0.67	1	285	26.5
P9	12	1	1	0.67	0.67	59	84.75
P10	17	1	2	0	0	71	56.50

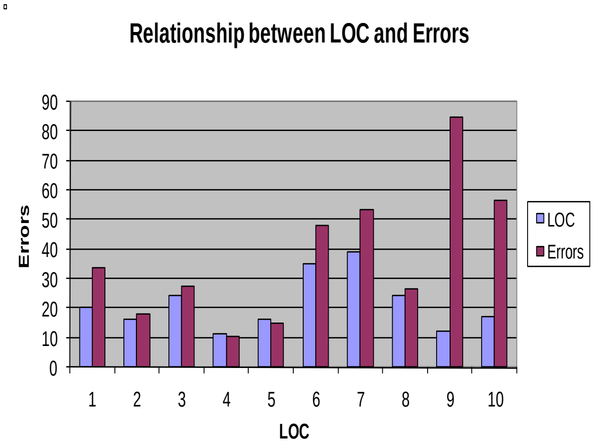


Fig. 1: Relationship between LOC and Errors

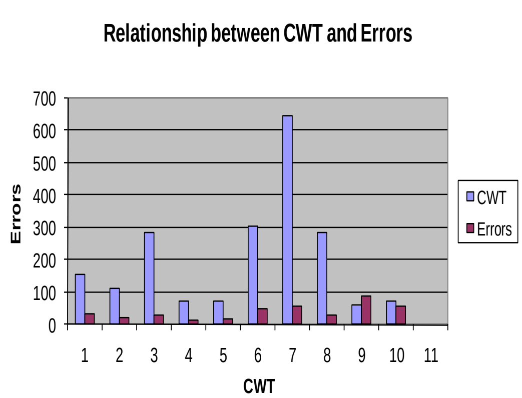


Fig. 2: Relationship between CWT and Errors

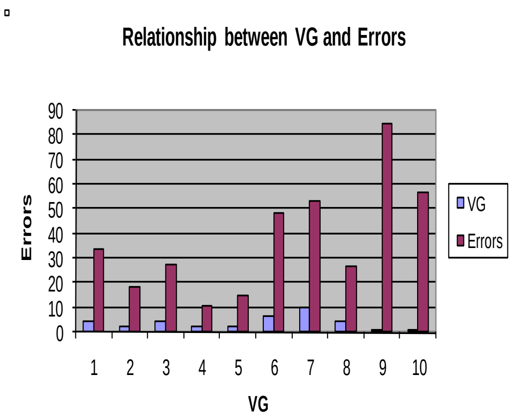


Fig. 3: Relationship between VG and Errors

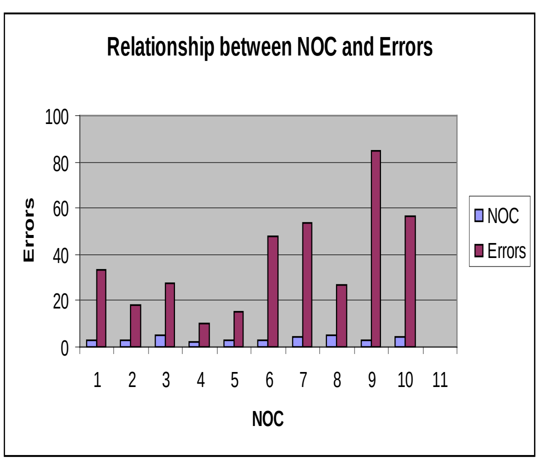


Fig. 4: Relationship between NOC and Errors

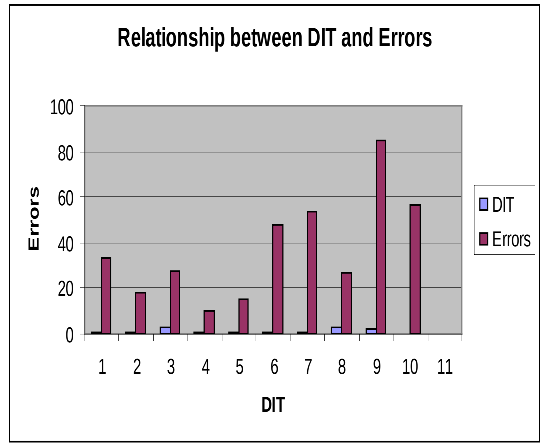


Fig. 5: Relationship between DIT and Errors

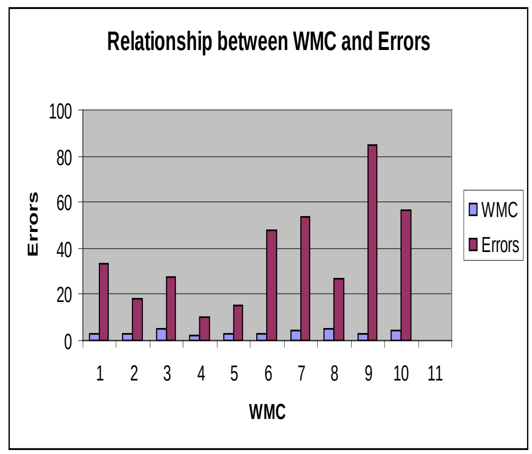


Fig. 6: Relationship between WMC and Errors

6. Conclusions And Directions For Future Work

In this paper, faults prediction through bebugging technique has been implemented through an experiment applied to C++ programs and compared the results with various object-oriented complexity measures. The results have been found encouraging. Relationship between faults and complexity has also been established. Making early decisions about complexity of a object-oriented system may help a lot to software developers in reducing design, testing and maintenance efforts and can improve its quality and reliability as well. The results appear to be logical and fit the intuitive understanding – if more complexity, then more possibility of faults. However, application of conclusions to real life situations needs further study.

So, further empirical research is required using data from industrial projects to validate these findings and to derive more useful and generalized results. Using data from industry implemented projects will provide a basis to examine the relationship between complexity and faults and we can comment on quality of software in a better way.

References

[1] Chidamber, S. R., Kemerer, C.F. A Metrics Suite for Object Oriented Design, IEEE Transactions on Software Engineering, 1994, pp. 476-492.
[2] Mark, L, Jeff, K. Object Oriented Software Metrics, Prentice Hall Publishing, .1994
[3] Basili, V.R., Biand, L., Melo, W.L. A validation of object-oriented design metrics as quality indicators, Technical report, Uni. of Maryland, Deptt. of computer science, MD, USA., 1995
[4] Basili, V. Qualitative Software Complexity Models: A Summary, In Tutorial on Models and Methods for Software Management and Engineering, IEEE Computer Society Press, Los Alamitos, CA, 1980.
[5] Singh, R., Grover, P.S. A New Program Weighted Complexity Metric, Proc. International conference on Software Engg. (CONSEG'97), Chennai, India, 1997, pp. 33-39.
[6] Brooks, I. Object Oriented Metrics Collection and Evaluation with Software Process, presented at OOPSLA'93 Workshop on Processes and Metrics for Object Oriented software development, Washington, DC, 1993.
[7] Harrison, W. Magel, K, Kluezny, R., dekock, A.: Applying Software Complexity Metrics to Program Maintenance, IEEE Computer, 15,1982, pp. 65-79.
[8] Fothi, A. Gaizler, J., Porkol, Z. The Structured Complexity of Object-Oriented Programs, Mathematical and Computer Modeling, 38, 2003, pp. 815-827.
[9] Da-wei, E. The Software complexity model and metrics for object-oriented, IEEE International Workshop on Anti-counterfeiting, Security, Identification, 2007, pp. 464-469.
[10] Brooks, F.P. The Mythical Man Month: Essays on Sofware Engineering, Addison-Wesley, 1995.
[11] Zuse, H. Software Complexity Measures and Methods, W.de Gruyter, New York, 1991.
[12] Sellers, B. H. Object-Oriented Metrics : Measures of Complexity, Prentice Hall, New Jersey, 1996.
[13] Curtis, B. Measurement and Experimentation in Software Engineering, Proc. IEEE conference, 68,9, 1980, pp. 1144-1157.
[14] Mishra, S. An Object Oriented Complexity Metric Based on Cognitive Weights, Proc. 6th IEEE International Conference on Cognitive Informatics (ICCI'07), 2007.
[15] Usha Chhillar, Sucheta Bhasin. A New Weighted Composite Complexity Measure for Object-Oriented Systems, International Journal of Information and Communication Technology Research, 1 (3), 2011.
[16] Elish, M.O., Rine, D. Indicators of Structural Stability of Object-Oriented Designs: A Case Study, Proc. 29th Annual IEEE/NASA Software Engineering Workshop(SEW'05), 2005.
[17] Halstead, M.H. Elements of Software Science, New York: Elsevier North Holland, 1977.
[18] McCabe, T.J. A Complexity Measure, IEEE Trans. On Software Engg., SE-2, 4, 1976, pp. 308-320.
[19] Aggarwal, K.K. Empirical Study of Object-Oriented Metrics, Journal of Object Technology, 5, 2006, pp. 149-173.
[20] Usha Kumari, Sucheta Bhasin. Application of Object-Oriented Metrics To C++ and Java : A Comparative Study, ACM SIGSOFT Software Engineering Notes, 36 (2), 2011, pp. 1-6.
[21] Singh, R. Improving Quality through Faults Prediction, International conference on Quality, Reliability and Information Technology at the Turn of the Millennium, December 21-23, 2000, New Delhi.

Usha Chhillar is working as Head, Department of Computer Science, A.I.J.H.M. PG college, Rohtak, Haryana, India. Currently, she is persuing her Ph.D Degree from Department of Computer Science and Applications, Kurukshetra University, Kurukshetra, Haryana, India. She obtained her Master Degree in Computer Science from Maharshi Dayanand University (MDU), Rohtak and M.Phil (Computer Science) from Ch. Devi Lal University (CDLU), Sirsa. She has total more than twelve years teaching experience. Her research interests include Software Engineering, Object-Oriented and Component-based Software Metrics.

Dr. Sucheta Bhasin is working as Associate Professor, Department of Computer Science and Applications, Kurukshetra University, Kurukshetra, Haryana, India. She has total more than 23 years teaching and research experience in the University. She has published more than 60 research papers in International/National journals and conferences. She is life member of International Forum for Interdisciplinary Mathematics and Indian Society of Information Theory & Applications. Her research areas include Networking, Software Metrics, Object-Oriented and Component-based Software Metrics.

Novel information security model using proposed e-cipher method with combining the features of cryptic-steganography

Prof. Venkateswaran Radhakrishnan, Dr. Sundaram Venkatachalam[2]

[1] Asst. Professor, GR Govindarajulu School of Applied Computer Technology and Research Scholar – Ph.D , Karpagam University, Coimbatore, Tamilnadu, India.

[2] Director , Department of CA , Karpagam College of Engineering Coimbatore, Tamilnadu, India.

Abstract

Cryptography is the art of hiding information in ways as to prevent detection of hidden messages. Secure data transmission method, which tries to alter the originality of the data files in to some encrypted form by using different methods and techniques. Encryption of data plays a vital role in the real time environment to keep the data out of reach of unauthorized people. After encryption, the files can be transferred securely by using multiple cytological methods.

In this Paper embed and de-embed processes of information hiding in various file format and carried out analysis in different approach and procedures are implemented in developing novel information security system in multimedia files like image and video, video file and other methods.

Varieties of techniques for embedding information in digital audio /video have been established. In this paper we will attend the general principles and different methodology adopted based on e -cipher model for hiding secret information using cryptographic technology, and an overview of functions and techniques, the goal of this paper is to know the different areas of information hiding and tools for providing secure data transmission with proposed e-cipher algorithms.

Keywords:* *Encryption; Decryption; data hiding; Mono Substitution; Poly Substitution; genetic keys.

1.Introduction

Information hiding technique is a new kind of secret communication technology. The majority of today's information hiding systems uses multimedia objects like image, audio and video files. Embedding secret messages in different file format is usually a more difficult process. Varieties of techniques for embedding information in digital image, audio /video have been established. In this proposed paper we will provide technique, which gives us more secure in information hiding system using cryptographic methods.

The importance of information and communications systems for society and the global economy is intensifying with the increasing value and quantity of data that is transmitted and stored on those systems. At the same time those systems and data are also increasingly vulnerable to a variety of threats, such as unauthorized access and use, misappropriation, alteration, and destruction. Proliferation of computers increased computing power, interconnectivity, decentralization, growth of networks and the number of users, as well as the convergence of information and communications technologies, while enhancing the utility of these systems, also increase system vulnerability.

Security of information and communications systems involves the protection of the availability, confidentiality and integrity of those systems and the data that is transmitted and stored on them. Availability is the property that data, information, and information and communications systems are accessible and useable on a timely basis in the required manner. Confidentiality is the property that data or information is not made available or disclosed to unauthorized persons, entities and processes. Integrity is the property that data or information has not been modified or altered in an unauthorized manner. The relative priority and significance of availability, confidentiality and integrity vary according to the information or communication systems and the ways in which those systems are used. The quality of security for information and communication systems and the data that is stored and transmitted on them depends not only on the technical measures, including the use of both hardware and software tools, but also on good managerial, organizational and operational procedures.

Cryptography is an important component of secure information and communications systems and a variety of applications have been developed that incorporate cryptographic methods to provide data security. Cryptography is an effective tool for ensuring both the confidentiality and integrity of data.

2.Background study of this research work

1. Basic concepts of cryptography, cryptology, cryptosystem and fundementatal concepts of caeser cipher, features and break analysis and various related cipher like mono, homo and PolyGram and other substitution ciphers.

2. Detailed study about transposition, substitution, transformation and other related Encryption types symmetric and asymmetric algorithms and related key concepts.

3. Gathered information about different ciphers like block cipher and stream cipher methodology related issues, challenges and other features and draw backs of this system.

4. Various attacking methods especially for cipher text, concentrated on cryptanalysis and brute force attack.

5. Mathematical concepts of substitutions, permutation, modulus functions, factorization concept and related issues.

6. Study about Block size, Key size, Number of rounds, Sub key generation algorithm, Round functions, Fast software encryption or decryption Braking analysis

7. Differentiate various attacking methods for cipher text in the form Cipher text, Known plaintext, Chosen plaintext, Chosen cipher text, Chosen text and analysis of various essential ingredients of symmetric system, secret key, cipher text, encryption and decryption and algorithm development.

8. Various key concepts private keys, public keys, session keys, master keys and proposed genetic keys.
9. Study about symmetric and asymmetric algorithms like, DES, AES and other related concepts, it was analyzed in various ways performance, time taken analysis, processing power and other issues based on cryptography aspects.

10. Done base work based on the different analysis of various substitution ciphers, exiting algorithms, related issues of attacking cipher text, features and international journals published recently on the web and other related articles and books.

3. Data hiding and retrieval process in multimedia file.

The vast improvement of the Internet and the digital information revolution caused major changes in the overall environment in the world. Flexible and simple-to-use software and decreasing prices of digital devices have made it feasible for consumers from all over the world to create, edit and exchange multimedia data. Broadband Internet connections almost an errorless transmission of data helps people to distribute large multimedia files and makes identical digital copies of them. In modern communication system Data Hiding is most essential for Network Security issue. Sending sensitive messages and files over the Internet are transmitted in an unsecured form but everyone has got something to keep in secret. Multimedia based data hiding method is one of the most effective ways to protect your privacy.

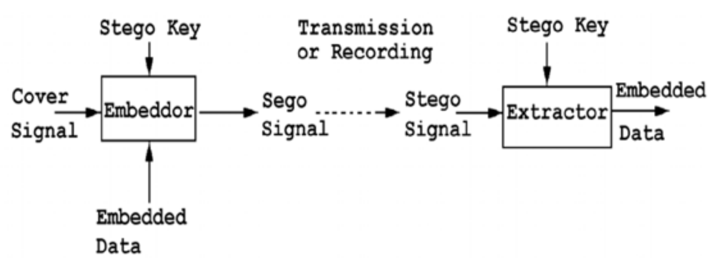


Fig1: Block Diagram of Data hiding and Retrieval

4. Objectives of this research work

· ***Authentication:*** This means that before sending and receiving data using the system, the receiver and sender identity should be verified.
· ***Secrecy or Confidentiality:*** Usually this function (feature) is how most people identify a secure system. It means that only the authenticated people are able to interpret the message (date) content and no one else.
· ***Integrity:*** Integrity means that the content of the communicated data is assured to be free from any type of modification between the end points (sender and receiver).

The basic form of integrity is packet check sum in IPv4 packets.

· ***Non-Repudiation:*** This function implies that neither the sender nor the receiver can falsely deny that they have sent a certain message.

· ***Service Reliability and Availability:*** Such systems should provide a way to grant their users the quality of service they expect.

5. Analysis of various model with proposed EC method.

The various analyses have taken to find strengthens and weakness of the many systems available using substitutions cipher method, based on the background study in different approached, finally we have got the poly alphabetic genetic keys substitutions cipher method. This method is used as a base method for security and then the output of this algorithm is passed to multimedia file (audio and video files) for high security in network.

Table 1. Shows analysis on various substitutions cipher Model

Mono Substitution Cipher
Methodology used: Shift Characters by fixed amount
Demerits: Easy to break algorithm based on frequency analysis poor security.
Example: Caeser Cipher & Vernam Cipher
Poly Substitutions Cipher
Methodology used: More than one replacement applied
Demerits: Message and keys are long, easy to break based on frequency analysis
Example: Vignere Cipher and Beaufort Cipher
Transposition Cipher
Methodology used: Permutes the symbols of the message according to a preset pattern.
Demerits: Insecure algorithms
Example: Row and Column Transposition Cipher Model
PolyGram Substitution Ciphers
Methodology used: Arbitrary Substitution for group of characters
Demerits: Reparative analysis to decipher the cipher text.
Example: Hill cipher, Flay air Cipher model
Proposed Substitution Model
Methodology used: ASCII with multiple key substitutions method
Merits: Genetic key used, fast, high security, embedded with multimedia files

The various analyses have taken to find strengthens and weakness of the many systems available using substitutions cipher method, based on the background study in different approached, finally we have got the poly alphabetic genetic keys substitutions cipher method. This method is used as a base method for security and then the output of this algorithm is passed to multimedia file (audio and video files) for high security in network.

Poly-alphabetic substitution ciphers the plaintext letters are enciphered differently depending upon their placement in the text. As the name poly alphabetic suggests this is achieved by using several two, three keys and random keys combinations instead of just one, as is the case in most of the simpler crypto systems.

Using two keys, we take 2 keys e1, e2 and let the ASCII values of e1 be 1 and e2 be 2 and take the text, add ASCII values of e1 to first character and ASCII values of e2 to second character. Alternatively add the value of e1 and e2 to consecutive characters.

5.1 Goal of my research

- Efficient to use.
- Must be available for all users.
- Fast.
- Provides high security.
- Improved version than Existing one.

6. Proposed information security system model

In polyalphabetic substitution ciphers the plaintext letters are enciphered differently depending upon their placement in the text. As the name polyalphabetic suggests this is achieved by using several two, three keys and random keys combinations instead of just one, as is the case in most of the simpler crypto systems.

6.1 About proposed Model

The majority of today's steganographic systems uses multimedia objects like image, audio and video etc as cover media because people often transmit digital pictures over email and other internet communication. Depending upon the nature of cover object, steganography can be divided into 5 types: Text steganography, Image steganography, Audio steganography, video steganography, and Protocol steganography. We hereby propose new novel information security system, which gives us high security system with features of cryptography tools and methods. There are various methods described in this paper about e-cipher model which is base Model for multiple data hiding system in the following methodology

- Text Cryptography, Image cryptography, Audio cryptography, Video cryptography, Unicode cryptography

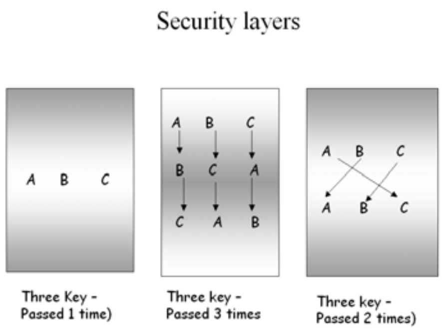


Fig. 2 Enhanced Security model using e-cipher model

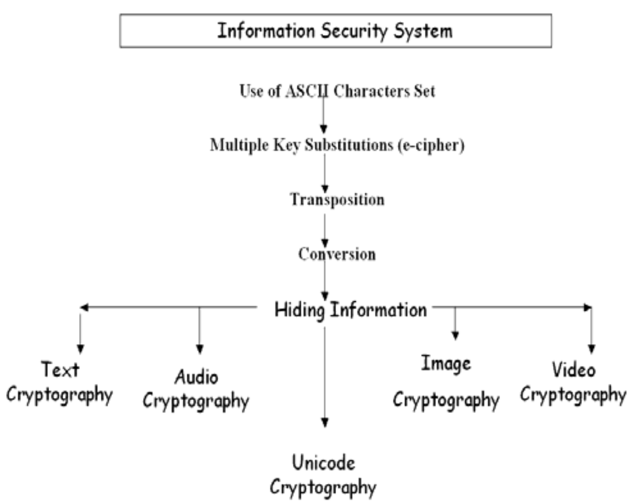


Fig. 3. Diagram shows difference security layers

6.2 Genetic keys

We can use Genetic keys (Three private keys) for text encryption by 2 keys and 3 keys and even more then 3 keys to make the decryption process more complicated.
•Using two keys, we take 2 keys e1, e2 and let the ASCII values of e1 be 1 and e2 be 2 and take the text, add ASCII values of e1 to first character and ASCII values of e2 to second character. Alternatively add the value of e1 and e2 to consecutive characters.

6.3 E-Cipher model

- Take the example text
- •Take three key e1, e2, e3 and assign a character e1 be ‘a’ and e2 be ‘D’ and e3 be ‘s’
- •Let ASCII value of e1 be 1 and e2 be 2 and e3 be 3 and take the text, add ASCII value of e1 to value of first character, and e2 to second character and e3 to third character, alternatively add the value of e1, e2, e3 to consecutive characters.
- •Three layers to be applied to each three consecutive letters and same to be continued thru the remaining text.
- •After adding ASCII value of all values of given text, the resultant text is an encrypted message. And it generate a combination of 3* (256 * 256 * 256) letters encrypted coded text with 128 bit manner.
- •Transposition takes place in each character after all the process are over that is moves or change one bit either LSB or MSB, the end result is increasing security
- •Finally takes the decimal values of each updated character in the given text and print in the encrypted format.

6. 4 Audio Steganography

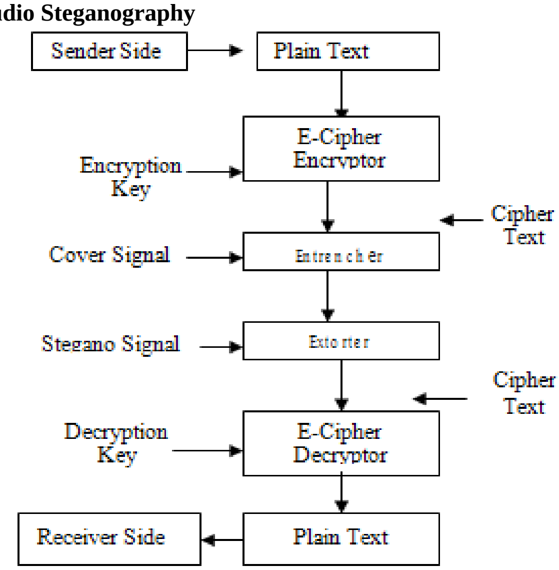


Fig 4. Audio steganography data flow model

In this work we propose a new model Information Security System – Information Hiding in Audio Signal - Embedding Text in Audio Signal that embeds the text with encryption that gains the full advantages of cryptography. In audio steganography, secret message is embedded into digitized audio signal which result slight altering of binary sequence of the corresponding audio file. For more security not only altering the bits of audio files, we embed high security e-Cipher cryptography algorithm in audio signal for data hiding.

6.4.1 Audio steganography base algorithm

Step 1: Load the audio file (AF) of size 12 K. Step 2: Input key for encryption

Step 3: Convert the audio files in the form of bytes and this byte values are represented in to bit patterns.

Step 4: Using the key, the original message is encrypted using E-Cipher algorithm.

Step 5: Bisect the audio file bit patterns horizontally.
•Step 6: Split the Encrypted message bit patterns vertically

Step 7: Insert the LSB bit of the vertically spitted encrypted text file (TF) into the LSB bit of the horizontally spitted audio file.

Step 8: Repeat Step 7 for the remaining bits of encrypted text file.
Step 9: If size (AF) ≥ size (TF) then
Embedding can be done as explained above
Else
The next higher order bit prior to previous bit position can be used
Until it is exhausted.

6.5 Image Steganography

In Image Steganography, There are a variety of methods using which information can be hidden in images. Least Significant Bit Replacement Technique: In image steganography almost all data hiding techniques try to alter insignificant information in the cover image. Least significant bit (LSB) insertion is a common, simple approach to embedding information in a cover image. For instance, a simple scheme proposed, is to place the embedding data at the least significant bit (LSB) of each pixel in the cover image. The altered image is called steno-image. Altering LSB doesn't change the quality of image to human perception but this scheme is sensitive a variety of image processing attacks like compression, cropping etc. We will be emphasizing more on this technique using e-cipher with LSB processing model gives us high and enhanced security for data transmission over networks.

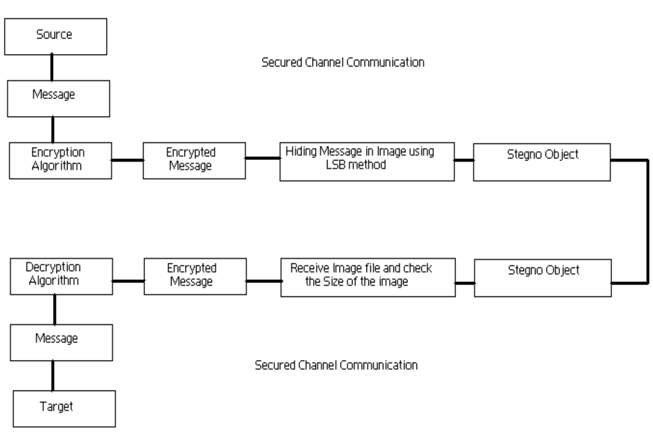


Fig 5. System Flow Diagram –Cryptic-steganography

6.5.1 Image Steganography base algorithm

1. Input encrypted file using EC algorithm.
2. Reading the text file and converting each char. To 8 bits in array (Conversion)
3. Reading the bmp file and adding each byte with the mask bits. (Watermarking)
4. Then adding to it 1 or not depending on the array. (Transposition)
5. Saving the result back to the bmp file.

7. Applications of the proposed work.

Audio/Video data hiding can be used anytime you want to hide data. There are many reasons to hide data but most important is to prevent unauthorized persons from becoming aware of the existence of a message. In the business world Audio data hiding can be used to hide a secret chemical formula or plans for a new invention. Audio data hiding can also be used in corporate world.

Audio/Video data hiding can also be used in the non-commercial sector to hide information that someone wants to keep private. Terrorists can also use Audio data hiding to keep their communications secret and to coordinate attacks. In the project which aims to embed animation parameters into audio and video contents. Data hiding in video and audio is of interest for the protection of copyrighted digital media, and to the government for information systems security and for covert communications.

It can also be used in forensic applications for inserting hidden data into audio files for the authentication of spoken words and other sounds, and in the music business for the monitoring of the songs over broadcast radio.

This work is more suitable for automatic control of robotic systems used in military and defence applications that can listen to a radio signal and then act accordingly as per the instructions received. By embedding the secret password in the audio signal the robot can be activated only if the predefined password matches with the incoming password that reaches the robot through audio signal. It can then start functioning as per the instructions received in the form of audio signal. More such sort of applications can be explored but confined to audio medium usage.

8. References

[1] Dr D Mukhopadhyay, A Mukherjee, S Ghosh, S Biswas, P Chakarborty: An Approach for Message Hiding using Substitution Techniques and Audio Hiding in Steganography, IEEE 2005

[2] Pradeep Kumar Singh,R.S. Agarwal, "Enhancement of LSP based steganography for hiding data., IJCSE , 02, No. 05, 2010, Page No.1652-1658

[3] Maram Balajee, Unicode and color integration tool for encryption and decryption , IJCSE, Vol. 3 No. 3 Mar 2011

[4] Nalani N, G. Raghavendra Rao,' Cryptanalysis of Simplified Data Encryption Standard via Optimization, Heuristics;IJCSNS, Vol.6 No.1B, January 2006

[5] William Stallings," Cryptography and Network Security: Principles and Practice", 2/3e Prentice hall, 2010.

[6] V. Lokeswara reddy, Dr.A. Subramaniyam, Dr. P. Cheena reddy. Implementation of LSP Steganograpy and its evaluation of various file formats , Int. J. Advanced Networking and Applications, Volume: 02, Issue: 05, Pages: 868-872 (2011)

[7] Sujith Ravi, Kevin Knight,'Attacking Letter Substitution Ciphers with Integer Programming',Oct 2009,33,4; Proquest Science Journals Pg.321.

[8] K.Geetha, P.V. Vanthia muthu," International journal of Computer Science and Engineering" vol 2 No.4 PG No: 1308-1312, Year 2010

[9] Darrell Whitley,'A Genetic Algorithm Tutorial', Computer Science Department, Colorado State University, Fort Collins, CO 80523.

7. Conclusions and future enhancement

This work proposes a new model for data transmission at higher degree of secrecy by using Image steganography, audio steganography, video steganography and proposed e- cipher methods and other technologies are presented here. This proposed study provides an efficient method for hiding the data from the eavesdropper. LSB data hiding technique is the simplest method for inserting data into audio signals.

And audio steganography model is able to ensure secrecy with less complexity at the cost of same memory space as that of encrypted text and the user is able to enjoy the benefits of cryptography and steganography combined together without any additional overhead. This work is more suitable for automatic control of robotic systems used in military and defense applications that can listen to a radio signal and then act accordingly as per the instructions received. By embedding the secret password in the audio signal the robot can be activated only if the predefined password matches with the incoming password that reaches the robot through audio signal.

It can then start functioning as per the instructions received in the form of audio signal. Some methods usage image steganography, which provide new way to hide the information in a secured way. Based on the perspective analysis, still we have to analyze more methods and add few updating in future and make the whole system which definitely provides a platform for the research engineers to help them for more innovation in this area and helps us to transfer our data more secure on the net.

10. BIOGRAPHY

Prof. R. Venkateswaran received his professional degree MCA and MBA (IS) from Bharathiar University, Tamilnadu, India, He received his M.Phil in computer science from Bharathidasan University, Tamilnadu, India, and He is currently a Ph.D Scholar in the Karpagam Academy of Higher Education, Karapagam University, Tamilnadu, India, in the field of Cryptography and Network Security. Presently he is working as an Asst. Professor of Computer Applications, GR Govindarajulu School of Applied Computer Technology, Coimbatore, Tamilnadu. He has 12 years of teaching experience and 3 years of research experience. He has participated in many national level conferences and workshops, published papers in five international conferences proceedings and published four papers in international refereed journals.

He is a member of CSI, IAENG, IACSIT, CSTA and many online forums. He has completed his course in Oracle 9i at Oracle University. His research interests are in cryptography and network security, information security, software engineering and database management systems.

Prof. Dr. V. Sundaram received his professional degree M.Sc. in Applied Mathematics from the University of Madras in the year 1967 and he received his Professional Doctoral Degree Ph. D in Mathematics from the University of Madras in 1989. He had 45 years of teaching as well as Research experience at PSG College of Technology and Polytechnic, Kumuraguru College of Technology and also worked in Ibra College of Technology, Sultanate of Oman. He is currently working as Director, Department of Computer Applications in Karpagam College of Engineering, Tamilnadu, India; He is a research Guide for Anna University as well as Karpagam University in the field of Computer Science and applications.

He has delivered guest lectures in the areas of computer applications and he had organized one international conference and six national level conference / symposium in the academics. He has attended and organized many faculty development programmes. He published several papers in International Journals and Conferences and also published 13 books in the area of engineering mathematics and he is the life member of ISTE and ISIAM. His research interests are in Cryptography and network security, Applied Mathematics, Discrete Mathematics, Network etc.

Cluster Based Hybrid Niche Mimetic and Genetic Algorithm for Text Document Categorization

A. K. Santra[1], C. Josephine Christy[2], B. Nagarajan[3]

[1] **Dean, CARE School of Computer Applications , Trichy – 620 009, India.**
[2] **Research Scholar, Bharathiar University, Coimbatore – 638401, India.**
[3] **Associate Professor, Bannari Amman Institute of Technology, Sathyamangalam – 638401, India.**

Summary

An efficient cluster based hybrid niche mimetic and genetic algorithm for text document categorization to improve the retrieval rate of relevant document fetching is addressed. The proposal minimizes the processing of structuring the document with better feature selection using hybrid algorithm. In addition restructuring of feature words to associated documents gets reduced, in turn increases document clustering rate. The performance of the proposed work is measured in terms of cluster objects accuracy, term weight, term frequency and inverse document frequency. Experimental results demonstrate that it achieves very good performance on both feature selection and text document categorization, compared to other classifier methods.

Keywords:
Document categorization, Feature Selection, Niche Mimetic Algorithm, Genetic Algorithm

1. Introduction

In the last ten years, content-based document management tasks have gained a prominent status in the information system field, due to the increased availability of documents in digital form and the ensuring need to access them in flexible ways [2]. Clustering is an important task in unsupervised learning. The essence of the clustering problem is to partition a set of objects into an a priori unknown number of clusters while minimizing the within cluster variability. Then it is maximizing the between cluster variability. Data clustering is a common technique for statistical data analysis and has been used in a variety of engineering and scientific disciplines such as biology.

While a wide range of classifiers have been used, virtually all of them were based on the same text representation, bag of words, where a document is represented as a set of words appearing in this document. Features used to describe a word are usually the ones which express whether the word appears in a document or how frequently this word appears. Above all, while the frequency of a word expresses the intuition that the more frequent, the more important, the compactness of the appearances of a word shows that the less compact, the more important and the position of the first appearance of a word shows that the earlier, the more important. Experiments suggest that the distributional features are useful for text categorization.

Document clustering groups similar documents into clusters on the basis of their contents. The documents in the resultant clusters exhibit maximal similarity to those in the same cluster and, at the same time, share minimal similarity with documents from other clusters. In addition, existing monolingual document clustering techniques can be classified broadly into non-LSI-based and LSI-based approaches. Memetic algorithms (MA) represent one of the recent growing areas of research in evolutionary computation. The term MA is now widely used as a synergy of evolutionary or any population-based approach with separate individual learning or local improvement procedures for problem search.

Here we suggest a unified criterion for simultaneous clustering and feature selection based on a well-known scatter separability index. A GA-based evolutionary procedure is then proposed to optimize the criterion. In order to allow simultaneous clustering and feature selection without the number of clusters being known a priori, a composite representation is devised to encode both feature election and cluster centers with a variable number of clusters. As a consequence, the crossover and mutation operators are suitably modified to tackle the concept of composite chromosomes with variable lengths. Additionally, we hybridize the proposed procedure with local search operations, which are introduced to refine the feature selection and cluster centers, respectively. These local searches move solutions toward local optima and allow a significant improvement in the computational efficiency. Finally, a niche method is integrated with the resulting hybrid GA to preserve the population diversity and prevent premature convergence.

2. Literature Review

S. Areibi and Z. Yang [1] have proposed several local search operations to effectively design an MA for simultaneous clustering and feature selection. which incorporate local searches with traditional GAs, have been proposed and applied successfully to solve a wide variety of optimization problems. These studies show that pure GAs are not well suited to fine tuning structures in complex search spaces and that hybridization with other techniques can greatly improve their efficiency. J. Shi and J. Malik [2] and S. Wu *et al.*[3] have proposed about data clustering is a common technique for statistical data analysis and has been used in a variety of engineering and scientific disciplines such as biology (genome data). Y. Zhao and G. Karypis [4] have proposed the purity of a cluster represents the fraction of the cluster corresponding to the largest class of documents assigned to that cluster; thus, the purity of the cluster.

One way of approaching this challenge is to use stochastic optimization schemes, prominent among which is an approach based on genetic algorithms (GAs). The GA is biologically inspired and embodies many mechanisms mimicking natural evolution. It has a great deal of potential in scientific and engineering optimization or search problems. Recently, hybrid methods [2], [5], [6], which incorporate local searches with traditional GAs, have been proposed and applied successfully to solve a wide variety of optimization problems. These studies show that pure GAs are not well suited to finetuning structures in complex search spaces and that hybridization with other techniques can greatly improve their efficiency. GAs that have been hybridized with local searches are also known as memetic algorithms (MAs) [7].

Traditional GAs and MAs are generally suitable for locating the optimal solution of an optimization problem with a small number of local optima. Complex problems such as clustering, however, often involve a significant number of locally optimal solutions. In such cases, traditional GAs and MAs cannot maintain controlled competitions among the individual solutions and can cause the population to converge prematurely [8].

To improve the situation, various methods [9], [10] (usually called niche methods) have been proposed. The research reported shows that one of the key elements in finding the optimal solution to a difficult problem with a GA approach is to preserve the population diversity during the search, since this permits the GA to investigate many peaks in parallel and helps in preventing it from being trapped in local optima.

GAs are naturally applicable to problems with exponential search spaces and have consequently been significant source of interest for clustering [6]. For example, in [11] proposed the use of traditional GAs for partitioned clustering. These methods can be very expensive and susceptible to becoming trapped in locally optimal solutions for clustering large data sets. Tsai et al. [6] introduced hybrid GAs by incorporating clustering-specified local searches into traditional GAs.

In contrast to the methods proposed in [3] and [5], clustering based on hybrid GAs can be more efficient, but these techniques can still, however, suffer from premature convergence. Furthermore, all of the above methods may exhibit limited performance, since they perform clustering on all features without selection. GAs have also been proposed for feature selection [3], [6]. However, they are usually developed in the supervised learning context, where class labels of the data are available, and the main purpose is to reduce the number of features used in classification while maintaining acceptable classification accuracies.

The second (and related) theme is feature selection for clustering, and feature selection research has a long history, as reported in the literature. Feature selection in the context of supervised learning [7], [6], adopts methods that are usually divided into two classes [2], [3] filters and wrappers based on whether or not feature selection is implemented independently of the learning algorithm. To maintain the filter/wrapper distinction used in supervised feature selection, we also classify feature selection methods for clustering into these two categories based on whether or not the process is carried out independently of the clustering algorithm.

The filters in clustering basically preselect the features and then apply a clustering algorithm to the selected feature subset. The principle is that any feature carrying little or no additional information beyond that subsumed by the remaining features is redundant and should be eliminated.

3. Document Clustering

Document clustering is very much for categorizing documents into meaningful groups. The usefulness of categorization is fully appreciated with labeling the clusters with the relevant feature words or phrases which describe various text document associated with them. A highly accurate key phrase extraction algorithm, called Core Phrase is proposed for this particular purpose.

Core Phrase works by building a complete list of phrases shared by at least two documents in a cluster. Phrases are assigned scores according to a set of features calculated from the matching process. The candidate phrases are then ranked in descending order and the top L phrases are output as a label for the cluster. While this algorithm on its own is useful for labeling document clusters, it is used to produce cluster summaries for the collaborative clustering algorithm.

Cluster key phrase summaries are exactly what is used to succinctly inform remote nodes of the content of local document clusters, which in turn is used to judge the similarity between remote data and local clusters. A distributed version of this algorithm is also used in the hierarchically-distributed clustering algorithm (described below) to produce summaries for the globally distributed clusters.

• Key phrase extraction can be applied to a single document for labeling the document; this is mainly used in generating metadata (e.g. title, description, keywords) that can be associated with the document.

• A centralized document cluster can be summarized and labeled using key phrase extraction.

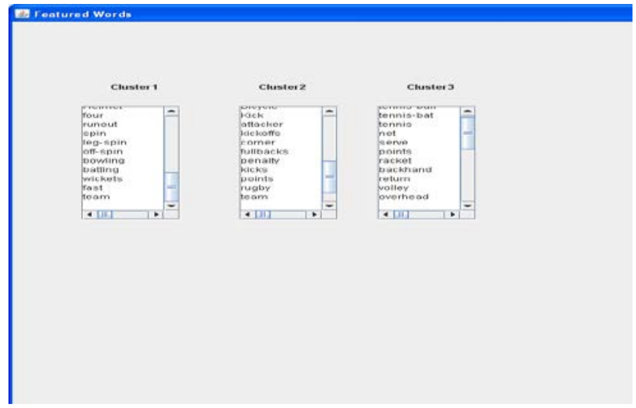
• Distributed document clusters in a flat peer-to-peer network can be summarized. Cluster summaries can be exchanged between peers to facilitate collaborative clustering.

• Distributed document clusters in a hierarchical peer-to-peer network can be summarized. Cluster summaries can be accessed at different levels of the hierarchy, thus providing variable scope of summaries ranging from specific to broad.

Evaluation of the accuracy of Core Phrase shows that it can accurately extract those phrases that match the manually labeled topic of clusters, and is able to rank those matching phrases in the top two or three key phrases.

Document clustering is used to organize a large document collection into distinct groups of similar documents. It discerns general themes hidden within the corpus. Applications of document clustering go beyond organizing document collections into knowledge maps. This can facilitate subsequent knowledge retrievals and accesses. Document clustering, shown in Fig 1. for example, has been applied to improve the efficiency of text categorization and discover event episodes in temporally ordered documents. In addition, instead of presenting search results as one long list, some prior studies and emerging search engines employ a document clustering approach to automatically organize search results into meaningful categories and thereby support cluster-based browsing.

Fig 1. Document Clustering



Various document clustering techniques have been proposed, but most deal with monolingual documents (i.e., all target documents are written in the same language). However, the globalization of business environments and advances in Internet technology often cause an organization to maintain documents in different languages in its knowledge repositories. Evidently, organizations face the challenge of multilingual document clustering (MLDC). Such MLDC requirement is also prominent in other scenarios. For example, with advances in cross-lingual information retrieval (CLIR) technology, many search engines now offer a functionality that retrieves, for a user query expressed in one language, relevant documents in different languages.

In this case, to facilitate cluster based searching, it would be preferable if the search engine were capable of clustering search results in different languages into distinct categories, each of which contains documents similar in their contents.

In our work feature selection is carried out to categories document with clustering method. The feature selection process is accomplished with integrated niche memetic and genetic algorithm which are explained in the following sub sections.

3.1 Feature Selection

Feature selection is important for clustering efficiency and effectiveness because it not only condenses the size of the extracted feature set but also reduces any potential biases embedded in the original (i.e., non-trimmed) feature set . Previous research commonly has employed feature selection metrics such as TF (term frequency), TF×IDF (term frequency × inverse document frequency), and their hybrids. A sample example is shown in Table 1.

Unlike the non-LSI-based document clustering approach, which typically involves a feature selection phase, the LSI-based approach to clustering monolingual documents employs LSI to reduce the dimensions and thereby improve both clustering effectiveness and efficiency. Its process generally commences with feature extraction, followed by document representation.

Name	Frequency	Inverse Frequency	Weight
tennis ball	0.33	3.0	1.0
tennis-bat	0.33	3.0	1.0
Tennis	0.33	3.0	2.0
Net	0.33	3.0	1.0
Serve	0.33	3.0	1.0
Points	0.33	3.0	1.0
Rocket	0.33	3.0	1.0
backhand	0.33	3.0	1.0
Return	0.33	3.0	1.0
Volley	0.33	3.0	1.0
Overhead	0.33	3.0	1.0

Table 1: Term Frequency - Tennis

4.1. Niching Memetic Algorithm

In our hybrid scheme of text document categorization, we used a niche MA for simultaneous clustering and feature selection by optimizing the unified criterion. This algorithm works with variable composite chromosomes, which are used to represent solutions. The operation of the algorithm consists of using a niche selection method for selecting pairing parents for reproduction, performing different genetic operators on different parts of the paired parents, applying local search operations (i.e., feature add and remove procedures and one step of K Means) to each offspring, and carrying out a niche competition replacement.

The evolution is terminated when the fitness value of the best solution in the population has not changed for g generations. The output of the algorithm is the best solution encountered during the evolution.

The flow of the algorithm is given as follows:

Step 1: initialize p sets of solutions randomly which encode both feature selection and cluster centers

Step 2: Calculate unified criterion J2 and set its fitness value as f ¼ J2.

Step 3: Repeat the following steps until the stopping criterion is met:

i. Select pairing parents until p=2 parent pairs are selected.
ii. Generate intermediate offspring by applying different genetic operators on the different parts of the paired parents.
iii. Apply feature add and remove procedures to the offspring.
iv. Run one step of K Means on the offspring.
v. Pair the offspring with the most similar solution found during a restricted competition replacement.
vi. Calculate J2 according to (4) for each of the offspring. If the fitness of the offspring is better than its paired solution, then the latter is replaced.

Step 4: Provide the feature subset and cluster centers of the solution

The accuracy rate can be calculated by utilizing the Nichie Memetic algorithm is shown in the Fig 2.

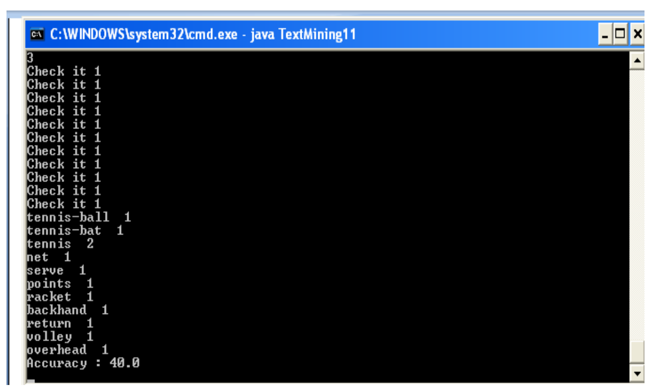


Fig: 2. Accuracy Rate of Nichie Memetic Algorithm

4.2. GA Algorithm

In a genetic algorithm, a population of strings (called chromosomes or the genotype of the genome), which encode candidate solutions (called individuals, creatures, or phenotypes) to an optimization feature selection of text from documents, evolves toward better solutions. Traditionally, solutions are represented in binary as strings of 0s and 1s, but other encodings are also possible. The evolution usually starts from a population of randomly generated individuals and happens in generations.

For the experiments, the system set genetic parameters as follows:

Generation limit = 100

Population size = 100

Cross-over fraction = 0.8

Mutation fraction = 0.1

Reproduction fraction = 1

Top N selection = 100

In each generation, the fitness of every individual in the population is evaluated, multiple individuals are stochastically selected from the current population (based on their fitness), and modified (recombined and possibly randomly mutated) to form a new population. The new population is then used in the next iteration of the algorithm. Commonly, the algorithm terminates when either a maximum number of generations has been produced, or a satisfactory fitness level has been reached for the population. If the algorithm has terminated due to a maximum number of generations, a satisfactory solution may or may not have been reached. The flow of the algorithm is given as follows:

step1: Initialize the Population

step2: Evaluate fitness of each member

step3: Reproduce with fittest members

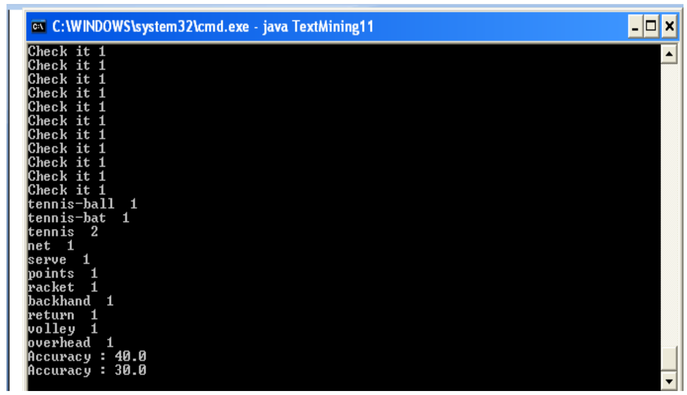
step4: Introduce random mutations in new generation

step5: Continue (2)-(3)-(4) until pre-specified number of generations are complete

Finally the hybrid niche memtic and genetic algorithm joins the process of feedback on the traditional foundation frame "training → Categorizing " algorithm. It expands the algorithm process as "Training → Categorizing feedback judgment → feedback". This kind of method is more close the real meaning machine learning. It show that the proposed hybrid algorithm has certain degree cognition self- determination in text document categorization using clustering methods.

The accuracy rate can be calculated by utilizing the Genetic Algorithm is shown in the Fig 3.

Fig 3: Accuracy Rate of GA Algorithm



5. Experimental Result and Discussion

In the experiment, we used Reuters-578, which has 178 documents collected from the Reuters newswire, as training sample set. Of the 35 categories in Reuters 578, only the most populous 10 are used. In data pre-processing, we applied stop word removal and tfc feature selection, and removed the commoner morphological and inflexion endings from words using The Porter Stemming Algorithm. Each category is employed as the positive class, and the rest as the negative class. For each dataset, 30% of the documents are randomly selected as test documents, and the rest are used to create training sets as follows: γ percent of the documents from the positive class is first selected as the positive set P. The rest of the positive documents and negative documents are used as unlabeled set U. We range γ percent from 10% - 50% to create a wide range of scenarios.

Preliminarily, documents were subjected to the following pre-processing steps: (1) First, we removed all words occurring in a list of common stopwords, as well as punctuation marks and numbers; (2) then, we extracted all n-grams, defined as sequences of maximum three words consecutively occurring within a document (after stopword removal)5; (3) at this point we have randomly split the set of seen data into a training set (70%), on which to run the GA, and a validation set (30%), on which tuning the model parameters. We performed the split in such a way that each category was proportionally represented in both sets (stratified holdout).

Based on the term frequency and inverse document frequency, the term weight will be calculated.

Term Weight = Term Frequency * Inverse Document Frequency

$$\text{Term Frequency} = \frac{\text{Term Count}}{\text{Total number of documents in count.}}$$

$$\text{Inverse Document Frequency} = \frac{\text{Total no. of documents}}{\text{No. of documents in the term}}$$

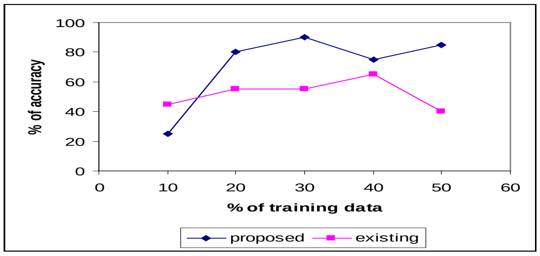


Fig4. Accuracy Vs Training Data

The above figure shows that the proposed technique gives better performance. Researchers showed 68% accuracy using the existing method with 31% test data while the technique is better both in accuracy and percentage of test data. Moreover it required processing for each class during training. But the proposed Algorithm does not require such process during training phase and hence reduces time.

6. Conclusion

The cluster based niche memetic and genetic algorithm have been designed and implemented by optimizing feature selection of the text in the documents repository. The efficacy of niche memetic is in evaluating optimal feature selection of text from the given set of documents. The contribution of genetic algorithm works on the evaluation of fitness function to cluster the relevant feature selected text which categorize documents to its most relevant cluster. In genetic algorithm, parameter tuning plays an important role for optimal feature text selection. In our approach, a text is segmented into groups of syllables with various lengths. We should build an auto parameter tuning scheme based on text length not a rigid one. This will speed up the processing time a lot.

The hybrid algorithm efficiency is shown in the experimental results, which confirms simultaneous global clustering and feature subset optimization mechanism is effective in text document categorization. The resulting algorithm is generally able to select relevant features and locate appropriate partitioning with the correct number of clusters and outperforms other methods implemented for comparison. We presented a new hybrid technique for text document clustering. The existing algorithms require more data for training as well as the computational time of these algorithms also increases. In contrast to the existing algorithms, the proposed hybrid algorithm requires less training data and less computational time. In spite of the randomly chosen training set we achieved 78% accuracy for 50% training data. Though 85% accuracy was observed in 30% training data, a class could not be classified, so we dropped this position and increased training data set for more acceptable result.

7. References

[1] S. Areibi and Z. Yang, “Effective Memetic Algorithms for VLSI Design Automation = Genetic Algorithms + Local Search + MultiLevel Clustering,” Evolutionary Computation, vol. 12, no. 3, pp. 327- 353, 2004.

[2] J. Shi and J. Malik, “Normalized Cuts and Image Segmentation,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, no. 8, pp. 888-905, Aug. 2000.

[3] S. Wu, A.W.C. Liew, H. Yan, and M. Yang, “Cluster Analysis of Gene Expression Database on Self-Splitting and Merging Competitive Learning,” IEEE Trans. Information Technology in Biomedicine, vol. 8, no. 1, 2004.

[4] Y. Zhao and G. Karypis, “Empirical and Theoretical Comparisons of Selected Criterion Functions for Document Clustering,” Machine Learning, vol. 55, no. 3, pp. 311-331, 2004

[5] F. Sebastiani. Machine learning in automated text categorization. ACM Computing Surverys, 2002, 34 (1): 1-47.

[6] H.K. Tsai, J.M. Yang, Y.F. Tsai, and C.Y. Kao, “An Evolutionary Approach for Gene Expression Patterns,” IEEE Trans. Information Technology in Biomedicine, vol. 8, no. 2, pp. 69-78, 2004.

[7] P. Baldi and G.W. Hatfield, DNA Microarrays and Gene Expression. Cambridge Univ. Press, 2002

[8] Xiaoli Li, Bing Liu, Learning to classify text using positive and unlabeled data. The International Joint Conference on Artifical Intelligence (IJCAI) (2003)

[9] W. Sheng, A. Tucker, and X. Liu, “Clustering with Niching Genetic K-Means Algorithm,” Proc. Genetic and Evolutionary Computation Conf. (GECCO ’04), pp. 162-173, 2004

[10] C. Wei, C.S. Yang, H.W. Hsiao, T.H. Cheng, Combining preference- and content-based approaches for improving document clustering effectiveness, Information Processing & Management 42 (2) (2006) 350–372.

[11] J. Kogan, C. Nicholas, and V. Volkovich, “Text Mining with Information-Theoretic Clustering,” IEEE Computational Science and Eng., pp. 52-59, 2003

A. K. Santra received the P. G. degree and Doctorate degree from I.I.T., Kharagpur in the year 1975 and 1981 respectively. He has got 20 years of Teaching Experience and 19 years of Industrial (Research) Experience. His area of interest includes Artificial Intelligence, Neural Networks, Process Modeling, Optimization and Control. He has got to his credit (i) 35 Technical Research Papers which are published in National / International Journals and Seminars of repute, (ii) 20 Research Projects have been completed in varied application areas, (iii) 2 Copy Rights for Software Development have been obtained in the area of Artificial Neural Networks (ANN) and (iv) he is the contributor of the book entitled **“Mathematics and its Applications in Industry and Business”**, Narosa Publishing House, **New Delhi.** He is the recognized Supervisor for guiding Ph. D. / M. S. (By Research) Scholars of Anna University-Chennai, Anna University-Coimbatore, Bharathiyar University, Coimbatore and Mother Teresa University, Kodaikanal. Currently he is guiding 12 Ph. D. Research Scholars in the Department. He is a Life member of CSI and a Life member of ISTE.

C. Josephine Christy received her M.Sc., M.Phil., M.B.A.,from Bharathiar University, Coimbatore. Currently she is working as Asst.Professor in Bannari amman Institute of Technology, Sathyamangalam. Her area of interest includes Text Mining, Web Mining. She has presented 2 papers international conferences and 5 papers in national Conferences. She is a Life member of Computer Society of India and a Life member of Indian Society for Technical Education.

B.Nagarajan received his Ph. D. degree in Pattern Recognition in the year 2010 from Anna University, Chennai. He has been in the teaching profession for more than a decade since 1997. His areas of academic interest are Image Processing, Pattern Recognition and Neural Networks. He has worked as
Also, he has published 14 papers in International Journals and presented 18 research papers in the National/International Conferences. He is the reviewer / editorial board member of 10 international journals from various countries like Singapore, Hong Kong, Korea, United States, Thailand, Romania and India.

Modeling Component-based Bragg gratings Application: tunable lasers

Rachida HEDARA, Mohammed .CHIKH-BLED
Telecommunication Laboratory, Department of Telecommunications
Technology Faculty, B.P.119 University Aboubekr Belkaïd
Tlemcen, 13000 Algérie

Abstract

The principal function of a grating Bragg is filtering, which can be used in optical fibers based component and active or passive semi conductors based component, as well as telecommunication systems. Their ideal use is with lasers with fiber, amplifiers with fiber or Laser diodes. In this work, we are going to show the principal results obtained during the analysis of various types of grating Bragg by the method of the coupled modes. We then present the operation of DBR are tunable. The use of Bragg gratings in a laser provides single-mode sources, agile wavelength. The use of sampled grating increases the tuning range.

Keywords: *Integrated optics, WDM, Guide of wave, Bragg grating , Telecommunications optical.*

1. Introduction

Optoelectronic components appear as key players in telecommunication architecture systems based on optical fiber, as they are present at all levels of information transmission, in order to transmit an information from a point to an other it is necessary to use a laser source for generating signals in form of a coherent optical wave. Then we must use an optical modulator that encodes the information in two levels ("0 " and "1 " optics). Afterward the optical fiber is used as an underlying support of transmission.
Regardless the significant progress made for reducing propagation losses in optical fibers. It is necessary to amplify the signal during its propagation, in order to avoid losses induced by the fiber [1].

The decreasing in spectral congestion has allowed evaluating the transmission mode giving the possibility to use multiple wavelengths in the same optical fiber. This is called wavelength multiplexing, which forms several independent transmission channels: each channel corresponds to a wavelength that is sent into the fiber. The demultiplexing is the inverse operation that performs the separation and the collection of signals coming from different channels.
The wavelength of these channels is located around 1.55µm which is the band of low attenuation in optical fibers.
In optical grating currently installed, DBR lasers play perfectly the role of a transmitter. Nevertheless, the constraints imposed on telecommunications and the emergence of new applications will promote the implementation of new laser generation sources.

2. Generality of the Bragg grating

Interests in the study of periodic structures, as well as of wave propagation in periodic medium are a phenomenon well known in solid physics.
The propagation of electromagnetic waves like light in a waveguide with a periodicity is a phenomenon quite similar, except that this wave is so confined, and the interactions do not appear in the direction of propagation. The introduction of such periodic modulation in waveguide structures has given rise to many devices such as optical filters.

A Bragg grating in an optical fiber is a periodic variation of refractive index in the core which is medium of optical signal propagation [2].
The periodic modulation introduced into a passive waveguide is the index of refraction of one or more of its constituent materials. This is called grating case volume figure Fig.1. (b).

A technique applicable to semi-conductors, for which the effects presented earlier are not sufficient, this technique introduce a periodic modulation through a physical corrugation in a waveguide structure of the by an engraving techniques.

This is called surface grating, figure Fig. 1. (a).

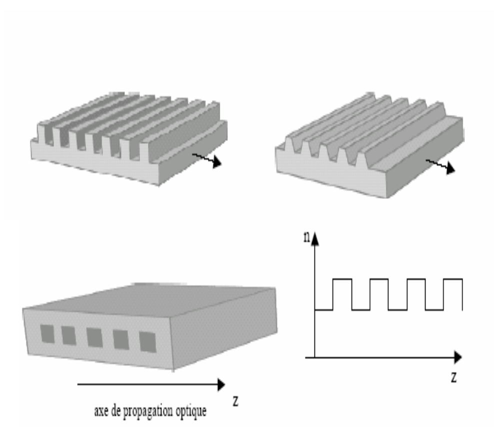


Fig. 1 (a) Surface grating, (b) Density grating

The Bragg grating can also be shorter in order to be easily integrated compared to fiber-based components in photonic circuits.
The physical phenomenon involved in a Bragg grating and their theoretical analysis, are similar for all profiles.
A fine periodic corrugation engraved on the surface of a waveguide creates a uniform coupling between light propagating in the forward direction and the one propagating in the backward direction. The grating is then analogous to a mirror formed by stacking periodic dielectric layers having different refractive indices.

3. Reflection coefficient of a Bragg grating

The Bragg grating presented a maximum reflectivity for the wavelength satisfying the Bragg condition [4].

$$\lambda_B = 2n_{eff}\Lambda \tag{1}$$

λ_B : Bragg wavelength

n_{eff} : the effective index of optical mode

Λ : Bragg period

We then obtain the expression of the maximum reflectivity of a uniform Bragg grating.

$$R_{MAX} = R(\lambda_B) = \tanh^2(kL) \tag{2}$$

The maximum reflectivity directly depends on the product's value kL, therefore we can define two types of grating: high and low-coupled grating.

3.1 Case of low coupling

When the multiplication kL is smaller than 1, the spectral response in reflection of the Bragg grating has a profile similar to the square of the sinc function centered on the Bragg frequency.

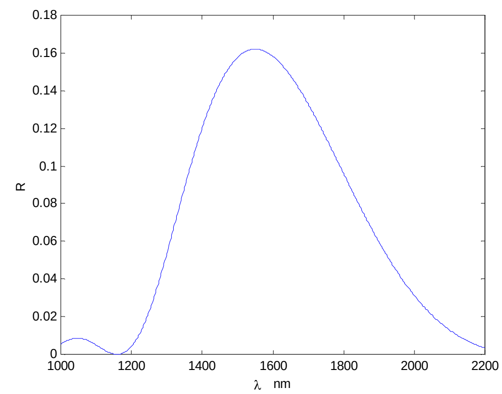


Fig. 2 Responses spectral reflection. Case of low coupling $(kL \leq 1)$

3.2. Case of strong coupling

When the multiplication kL is widely above 1, the profile of the reflection spectrum has a plate at R = 1 becoming more pronounced when kL increases.
In this case, the network has very high reflectivity. Also we can observe band of wavelengths, called stopband, where the light is considered totally reflected. It is centered on the Bragg wavelength (λ_B =1550).

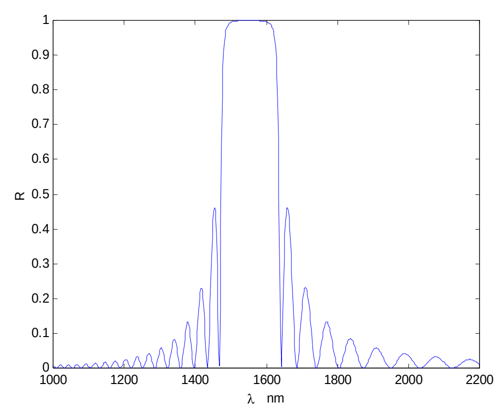


Fig. 3 spectral response in reflection. Case of strong coupling (kL >1).

4. Analytical formulation of power reflector [3]

If most indices are purely real and we consider the stacking to its wavelength design for which the layers are exactly quarter of the wave, then completely analytical formulations of the reflectance. Taking the conventions of mirrors we obtain:

$$R_{2N} = \left[\frac{1 - \frac{n_S}{n_0}\left(\frac{n_L}{n_H}\right)^{2N}}{1 + \frac{n_S}{n_0}\left(\frac{n_L}{n_H}\right)^{2N}} \right]^2 \tag{3}$$

n_S :the index of the substrate, n_0 index of incident medium, n_H :the index of the material with high index
n_L: the index of the material with low index.

The use of this simplified formula shows that it is advantageous to have a contrast between the indices *nL* and *nH* is very important in the mirror to minimize the ratio *nL/nH*.

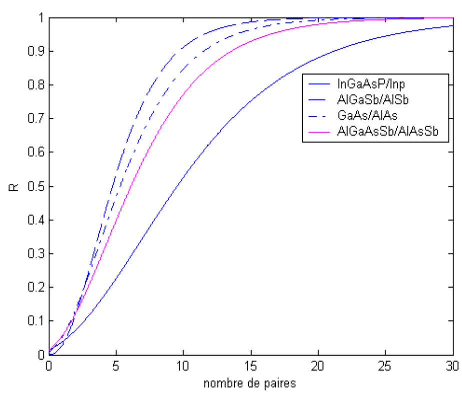


Fig. 4 Reflectivity as a function of the maximum number of pairs of layers

The advantage provided by materials with high index ratio *nL/nH* is evident on this graph. Although it seems impossible to obtain a reflectance higher then 99.9% with the system InGaAsP/InP on InP, while this is quickly achieved with systems AlGaSb/AlSb on GaSb and GaAs/AlAs on GaAs and lesser with the system AlGaAsSb/AlAsSb on InP.

5. Bragg grating individuals

Until now we have assumed that gratings are perfectly periodic along structure with starting point and end well defined. Often, either by design or because of used manufacturing techniques, the Bragg grating deviates slightly from the perfect structure.

The use of Bragg gratings in various optical components, in its simplest form, called the uniform grating is widespread. Nevertheless, the emergence of new functions of passive and active components needs more efficient grating. This is why this gratings are more particular than the sampled grating and vertical grating become necessary.

6. Sampled Bragg grating

A sampled grating consists of a right grating at a wavelength probably defined, multiplied by a sampling function. It is possible to model the reflectivity complex of sampled grating using the theory of coupled mode and the theory of transfer matrices, but it must first define the key parameters which modify its properties in phase and amplitude.

A sampled grating is a conventional grating from which we remove portions periodically. In other words, there is an alternation between sections with gratin and sections without a grating as described in Fig.

A Bragg grating wavelength λ_B ((defined by the step Λ and the effective index n_{eff}) and located on a distance Z_1 with a coupling coefficient the grating is then repeated m times (number of sampling periods) with a period Z_0. The total length of sampled grating is $L_{Tot}= mZ_0$. The spectral response of this grating is the Fourier transform of the profile index. The result is a comb Bragg reflectors, regularly spaced with the sampling frequency ISL_{SBG} most often expressed in GHz.

$$ISL_{SBG} = \frac{c}{2n_{eff}Z_0} \tag{5}$$

6.1 Model Description

The multiplication of a Bragg grating with a sampling function provides a sampled grating. The Fourier components of the sampled grating can be obtained by the convolution of the Fourier component of a Bragg grating by a comb of Fourier components of a sampling function.

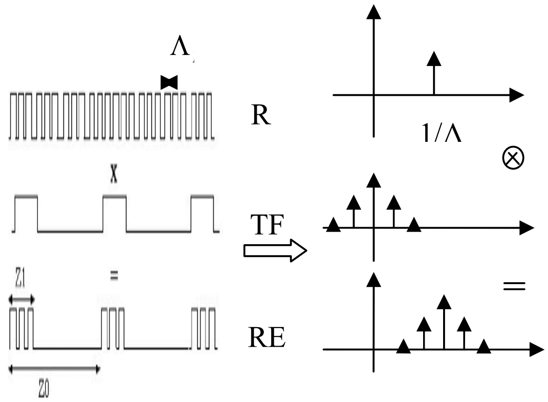


Fig. 5 grating sample space to real space Fourier

The reflectivity of this structure can be obtained from the coupled mode theory, which predicts that each component of the spatial Fourier series decomposition of the permittivity create a peak in the overall spectrum of reflections.

The q th component Δ ε (q) on the sampled grating is connected to the single Fourier component Δ 0 0 ε of an unsampled grating by the relation [5]:

$$\Delta\varepsilon(q) = \Delta\varepsilon_0 \frac{Z_1}{Z_0} \frac{\sin \pi \ qZ_1 / Z_0}{\pi \ qZ_1 / Z_0} \exp[-j\pi qZ_1 / Z_0] \tag{6}$$

If the coefficient of coupling portions of the grating is k_0 then the coupling coefficient of the qth Fourier component of the total sampled grating is:

$$k(q) = k_0 \frac{Z_1}{Z_0} \frac{\sin \pi \ qZ_1 / Z_0}{\pi \ qZ_1 / Z_0} \exp[-j\pi qZ_1 / Z_0] \tag{7}$$

It is possible to know the reflection coefficient of the field in a Bragg grating,by introducing the coupling coefficient of the sampled grating, it is then possible to determine the reflectivity of complex grating sampled.

$$r(\lambda) = \sum_{-q}^{q} \frac{jk^*(q) \ \ \sin(Q(q))L_{Tot}}{Q(q)\cos(\theta(q)L_{Tot}) - j\Delta\beta(q)\sin(Q(q)L_{Tot})} \tag{8}$$

Or

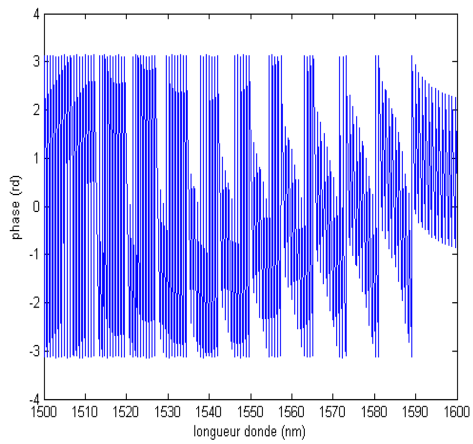
$$\Delta\beta(q) = \frac{2\pi n_{eff}}{\lambda} - \frac{\pi}{\Lambda} - \frac{\pi q}{Z_0} \quad ; \quad (Q(q))^2 = (\Delta\beta(q))^2 - |k(q)|^2$$

The evolution of the field in a periodic structure is described by its complex reflectivity. The parameters for modifying the spectral properties are the total length of L_{SBG}, the coupling coefficient of the grating $k(q)$, the sampling period Z_0 and the opening ratio h. The influence of these parameters can be modeled in a sampled grating but also in the particular cases of a non-sampled Bragg grating with opening ratio unit (Z0 = Z1). We can also define the bandwidth of the peak q between two zeros by:

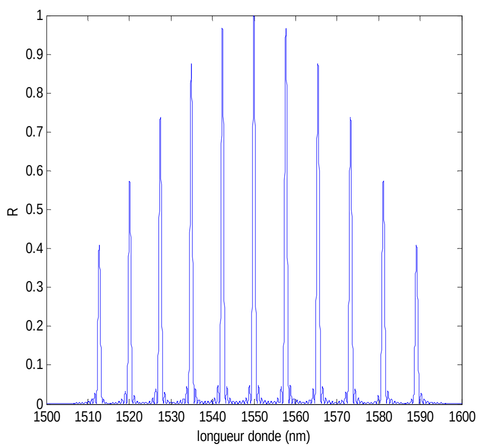
$$\Delta\lambda_{bw}(q) = \frac{\lambda^2}{\pi \ n_g} \sqrt{|k(q)|^2 + \frac{\pi^2}{{L_{Tot}}^2}} \tag{9}$$

6.2 Simulation of a Sampled Bragg grating

Two additional parameters define the sampling function (opening ratio h sampling period Z0). Modeling allows the calculation of the complex reflectivity of a sampled grating phase φ[λ] = Arg[r[λ]] and amplitude.
The models below are the typical characteristics of sampled grating, engraved in InP [6] [7]. These dimensions widely studied and optimized are: Z_0= 450µm, k_0 =150cm^{-1} et h=10%.



(a)



(b)

Fig. 6 (a) Phase (rad) in a sampled grating
(b) Reflectivity (amplitude) in a sampled grating

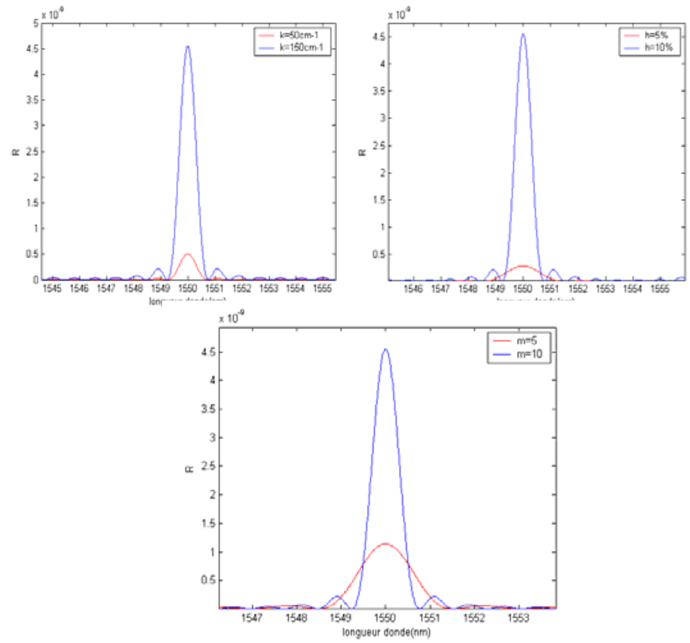


Fig. 7 Influence of (, m, h) on the maximum reflectivity and bandwidth

The modification of one parameter changes the overall properties of the sampled grating

The variation of coupling coefficient can increase the reflectivity of the grating and the bandwidth of the peaks.

In addition, the modified opening ratio changes the Envelope of The Sampled grating (number of peaks in the envelope) but also changes the reflectivity.
The number of periods and consequently the length of the sampled grating reflectivity will increase grating and reduces the bandwidth of the peaks.
The value of a grating is sampled to obtain a structure which has a low coefficient of effective coupling. From a technological point of view, it is difficult to reach values of coupling coefficients low.

The bandwidth of the reflection peak depends directly on the value of the effective coupling coefficient of the structure. So if you have a very small value of coupling coefficient, while retaining a correct length of the structure we obtain a peak very close and very selective what is interesting in multiplexing systems in dense wavelength (DWDM).

We saw here the main advantages offered by the individual networks that are the networks sampled. However we have also seen that their use requires compromises that can sometimes be difficult.

7. DBR tunable lasers

7.1 Principle of Operation

The passive DBR of a length of 1 mm is located in the RW section near the back face of the laser. while the Bragg grating and RW were defined by a lithography projection using a wafer stepper I-line and transferred to the surface of the semiconductor by dry etching process [9].
An effective reflectivity of the DBR around 60% was obtained by well designed length of the no engraved region within the grating [9].

The spectrally selective element is a policy grating, forcing the laser emission at the Bragg wavelength, while the output power and beam properties are imposed by the conical structure amplifier. Indeed, Bragg gratings have become essential optical components in recent years, with high spectral selectivity, low loss, limited drift wavelength (~ 0.01 nm / K) and the reflectivity of some 0.1% at 100 % [10].
Once the laser is manufactured, the spectrum of the gain condition is likely to continue since it is intimately linked to the material and structure. That leaves two possibilities:

A. **Changing the Status Phase**

The phase condition is related to the geometry of the cavity. In reality, the distances seen by the optical wave are n × L. We can not of course change the physical length of a laser made from a solid material.
However, it is possible to modify slightly the index of the material that composing it by several techniques.

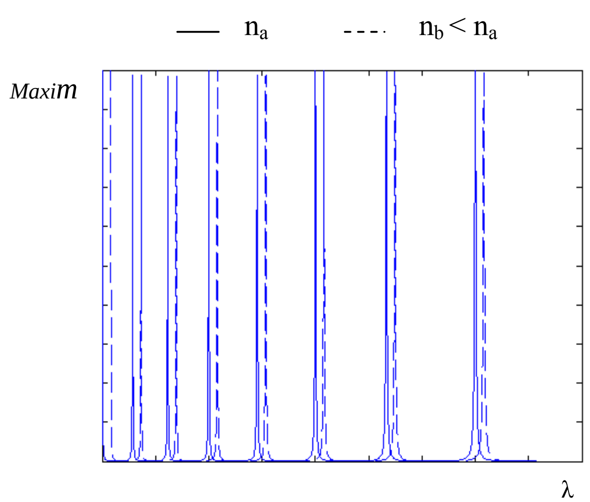


Fig. 8 Changing the phase condition

By changing the index of the optical wave guide, we change the distribution of the maxima in the core-Perot (Fig. 8), since it creates an additional phase shift.
We can also note that the evolution of the wavelength depending on the index is linear. Thus, the evolution of the wavelength depending on the phase parameter is done continually.

B. Changing the filtering Bragg Area

The filter response is also dependent Bragg indices of the materials considered

"The change in each of these two parameters causes significant changes in the spectrum emitted by the laser".

We assume here that the changes made to the index of the material are relatively low, and quantitatively the same for each of the two indices considered. We will see later that this corresponds to the behavior of compounds III-V.

$$n_1 \to n_1 + \Delta n$$
$$n_2 \to n_2 + \Delta n$$

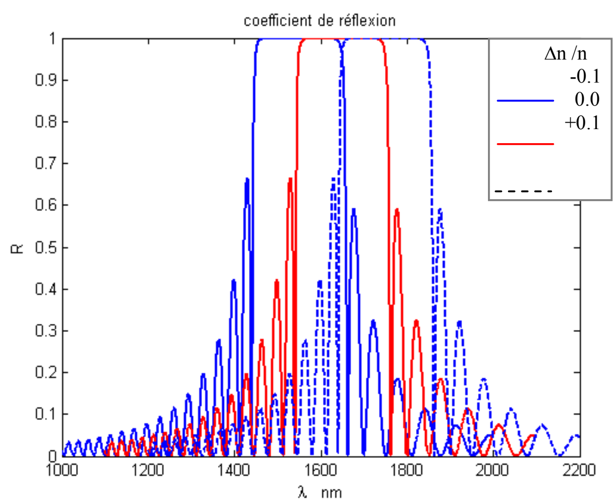


Fig. 9 Evolution of the Bragg filter according to the indices of materials

By injecting this variation in the numerical model, we obtain the results shown in Figure 9.

Thus we observe an effect of translation of the spectral bandwidth available, with a negligible effect on the bandwidth.

We simulated the evolution of the central wavelength of the Bragg mirror based on the relative variation of the index layers.

We obtain an almost linear deviation of the center wavelength depending on the index variation.

Thus, the filter Bragg will be able to move to spectrally to select a mode of Fabry-Perot and change the condition of laser oscillation.

7.2 Accordabilité of a Laser

We have identified two ways to tune a laser DBR:

A. Continuous tunability: the phase condition of the Fabry-Perot
B. Discontinuous tuning: by translation of the Bragg filter.

The tunability with the phase condition is often insufficient: the range of wavelengths obtained is too small.
However, the tunability of the Bragg filter allows a good tuning range, but achieves certain wavelengths, those defined by the Fabry-Perot.

The tunability of the phase region allows for continuous excursions of about 100 GHz, that allowed by the Bragg zone corresponds to excursions of 2 THz. This limitation is due to the low index variation depending on the injection current, and not to the spectral spread of the condition of material gain. The dependence of output power, the quality and optical spectrum of the input currents will be shown.

Various measures used to improve the modulation efficiency of a diode laser DBR similar to those presented in [11] and [12], emitting at 1060 nm.

8. Conclusion

The realized work in this paper concerns the theoretical depth of the Bragg gratings and sees their applications in optical telecommunications. The objective of this work is to study types of integrated structures periodic Bragg gratings or sampled.

Bragg gratings are components in telecommunications systems they are ideal when used in combination with fiber lasers, fiber amplifiers or laser diodes. In addition to their flexibility and their design makes them very attractive for applications to the user's needs [8].

The use of Bragg gratings in a laser provides single-mode sources, agile wavelength. The Bragg grating acts as a mirror that is reflecting only one wavelength, thereby effectively select a single mode source.

References

[1] Brackett, C.A; « Dense Wavelength Division Multiplexing networks: principles and applications», IEEE Journal of selected areas in communications, 1990, Vol 8, N°8, p.948-964.

[2] S.Garidel ; « Fabrication des réseaux de Bragg particuliers par lithographie électronique : application à la réalisation de dispositifs photoniques et optoélectroniques sur matériaux de la filière InP »,Thèse de doctorat, Lille, 2004.

[3] T. Zeno.; «Optoelectronique-composants photonique et fibres optiques», Ellipses-2001.

[4] M. Born , E .Wolf; «Principles of Optics», Pergamon Press, New York, 1970.

[5] A .Yariv; « Coupled-mode theory for guided-wave optiucs», IEEE journal of Quant um electronics Volume: 9 Issue: 9, septembre 1973, page: 199-933.

[6] G .Bjork, O .Nilsson; «Anew exact and coefficient numerical matrix theory of complicated laser structures: proprieties of asymmetric phase-shifted DFB lasers», journal of light wave technology, volume: 5 Issue: 1, Jan 1987, page: 140-164.

[7] B .Mason, J .Barton, G.A .Fish, L.A .Coldren, S.P .Denbaars; «Design of sampled grating DBR lasers with integrated semiconductor optical amplifier», IEEE photonics Technology letters, Volume: 12 Issue: 7, July 2000, page: 762-764.

[8] Pierre Aschieri; «Introduction aux télécommunications optiques», note cours optique université Nice version 12, 2004.

[9] K-H. Hasler, B .Sumpf, P .Adamiec, J .Fricke, H .Wenzel, G .Erbert, G .Tränkle; «Improving the modulation efficiency of high-power distributed Bragg reflector tapered diode lasers» Applied Physics B lasers and optics Springer- Verlag 2010.

[10]g.lucas-leclin1,d.paboeuf1, p .georges1, holm2 p, j. andersen2 ,b .sumpf3, g .erbert; «Wavelength abilization of extended-cavity tapered lasers with volume Bragg gratings» Applied Physics B Lasers and Optics ringer- Verlag 2008.

[11] B.Sumpf, K.H.Hasler, P. Adamiec, F. Bugge., J .Fricke, Ressel. P, H .Wenzel, G .Erbert, G .Tränkle, Proc. SPIE 7230, 72301E (2009)

[12] B.Sumpf, K. H .Hasler, P .Adamiec, F .Bugge, F .Dittmar, J .Fricke, H .Wenzel, M .Zorn, Erbert , G .Tränkle, IEEE J. Sel.Top. Quantum Electron. 15, 1009 ,2009.

IJCSI CALL FOR PAPERS JANUARY 2012 ISSUE

Volume 9, Issue 1

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: 30th November 2011
Notification: 04th January 2012
Revision: 12th January 2012
Online Publication: 31st January 2012

- 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).

For more information, please visit the journal website (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