Proceedings

The Second International Symposium on Networking and Network Security (ISNNS 2010)

2-4 April 2010

Jinggangshan, China

Editors

Fei Yu
Guangxue Yue
Ming Leng
Xuange Peng

Co-Sponsored by

Jinggangshan University, China
Peoples'Friendship University of Russia
Feng Chia University, Taiwan
Nanchang HangKong University, China
East China Jiaotong University, China
Jiaxing University, China
Qingdao University of Science & Technology, China
Academy Publisher of Finland, Finland
Hunan Agricultural University, China
Technical Co-Sponsor:
IEEE
IEEE Shanghai Section

The papers in this book published by the Academy Publisher, Post: P.O. Box 40, FIN-90571, Oulu, Finland, Email: general@academypublisher.com, Internet: http://www.academypublisher.com/, Phone: +358 (0)44 525 7800, Fax: +358 (0)207 81 8199. The book is made available on-line at http://www.academypublisher.com/proc/.

AP Catalog Number AP-PROC-CS-10CN006

ISBN 978-952-5726-09-1

Additional copies may be ordered from:

Academy Publisher
P.O. Box 40, FIN-90571
Oulu, Finland,
Phone: +358 (0)44 525 7800
Fax: +358 (0)207 81 8199
Email: order@academypublisher.com

ACADEMY PUBLISHER
http://www.academypublisher.com/

Table of Contents

Message from the Symposium Chairs vii
ISNNS 2010 Organizing Committee viii
ISNNS 2010 Committee Members ix

Shot Identification Based on DBN 1
Yun Liu, Xueying Liu, and Chao Huang

Discussion on the Current Development and Reform Ideas about the Competitive Wushu Routine at Colleges in Jiangxi Province 5
Youru Yu , Zhipeng Xiao , Weimin Wang, and Hongmei Deng

Analysis of Uncertainty Information for Several Kinds of Complex Systems 9
Jiqin Peng, and Jinfang Han

Collaborative WWW-based CAI Software Achieved by Java 12
Sanjun Liu, Zhengjiang Yi, Weifeng Du, and Zhengjiang Wu

Expression of XML Implicit Semantics 15
Ying Chen, Wenping Guo , and Xiaoming Zhao

An Application Research on Video Coding H.264 Standard and Video Conference System 19
Zhan Guo, Zuming Xiao, Feng Luo, and Bin Tan

Application of Neural Networks in Medical Image Processing 23
Zhenghao Shi, and Lifeng He

Research and Application of Auto-profiling Control Algorithm for 0Cr18Ni10Ti 27
Qiang Wang, XuGuang Min , ZhongYue Xiao, and Yanhui Zhou

Research and Realization about Conversion Algorithm of PDF Format into PS Format 31
Xingfu Wang, Lei Qian, Fuyou Mao, and Zhaosheng Zhu

The Design of Ethernet Controller Interface Circuit Based on ENC28J60 35
Bin Tan, Bo Yuan, and Bing Zhu

Security Research on WEP of WLAN 39
Peisong Ye, and Guangxue Yue

Research on the Safety Assessment of Bridges Based on Fuzzy-Neural Network 43
Bo Wang, Xuzheng Liu, and Chao Luo

Influences of Powder Content on the Piezomagnetic Effect of Rubber-based Soft Magnetic Composite Film 47
Daguo Jiang

An Improved Molecular Solution for the Partition Problem 50
Xu Zhou, and ShuangShuang Huang

Research for the Algorithm of Query to Compressed XML Data 54
Guojia Yu, Huizhong Qiu, and Lin Tian

A Middleware-Based Approach for Mobile Service Delivery 58
Weidong Zhao, Haifeng Wu, Weihui Dai, Wenjuan Wang, and Guangxue Yue

DDoS Detection System Based on Data Mining 62
Rui Zhong, and Guangxue Yue

Exploration of E-Business 66
Hongmei Xiang

Research on Weights Assigning Based on Element Consistency Level in Judgment Matrice 69
Xixiang Zhang, Jianxun Liu, Liping Chen, and GuangxueYue

Improving Prediction Of Model Gm (1, 1) Based On Class Ratio Modeling Method 73
Qin Wan, Yong Wei, and Shuang Yang

Research and Implementation the Real Time Rendering and Integration Of Vector Data with 3D Terrain 77
Meiyu Zhang, Yu Li, and Qi Hu

The Evaluation System for Blending Learning Based on Blackboard Academic Suite 81
Lin Xin, Xixiang Zhang, Kede Qin, and GuangxueYue

Swarm Dynamics Behavior Analysis and Coordinated Control of Limited-Range Perceived Agents 85
Zhibin Xue, and Jianchao Zeng

An Anomaly Detection Method Based on Fuzzy C-means Clustering Algorithm 89
Linquan Xie, Ying Wang, Liping Chen, and Guangxue Yue

An Alarm Flow Decomposition Method for Security Threat Evaluation 93
Jie Ma, and Zhitang Li

Fuzzy Control of Intelligent Vehicle Based on Visual Navigation System 97
Tingjian Zhong, and Meilian Qiu

Prediction of Reservoir-induced Earthquake Based on Fuzzy Theory 101
Ming Zhong, and Qiuwen Zhang

Study on Civil Airport Safety Management System Based on Risk Management 105
Ming Cheng, and Wenjuan Liang

Analysis System for Correlative Fluctuation in Futures Market 108
Xingyun Dai, Weihui Dai, Hua Ye, and Genghui Dai

Extraction Algorithms for Using a Regular Grid DEMs 112
Kun Hou, Jigui Sun, Wei Yang, Tieli Sun, Ying Wang, and Sijun Ma

Study on the Failure Identifiability of Disaster Detection System 116
Xiaoyu Wang, and Kaiquan Wang

Provably Correct Aspect-Oriented Self-monitoring Programs 120
Li Li, and GuoSun Zeng

A New Information Hiding Method for Image Watermarking Based on Mojette Transform 124
Cui Zhang, Junyu Dong, Jining Li, and Florent Autrusseau

Modeling Seismic Wave Propagation Using Graphics Processor Units (GPU) 129
Zhangang Wang, Suping Peng, and Tao Liu

Explore Diigo for the First Time Applying to Learning Hitting the Target 133
Bo Yuan, and Bin Tan

Agent-Based Service-Oriented Dynamic Integration ERP Architecture 136
Zuoming Huang, Jincheng Zhang, and Qiushi Cong

Applications of Computer Control System in Automobile Comprehensive Performance Test 140

Zuming Xiao, Zhan Guo, Bin Tan, and Bing Zhu

Policy-based Autonomic Mobile Network Resource Management Architecture 144
Fang Mei, Yanheng Liu, Hui Kang, and Shuangshuang Zhang

A Energy-efficient Scheduling Algorithm in mWSN 149
Xinlian Zhou,and Bin Xiao

A New Method to Classify Shots in Basketball Video 153
Yun Liu, Chao Huang, and Xueying Liu

Developing Mobile Input Method System Using An Improved Design Pattern Approach 157
Fei Zhu, Hongjun Diao, and Wei Huang

Reviewing some Cloud Computing Platforms 161
Dong Xu, and Hui Liu

Weighted Suffix Tree Document Model for Web Documents Clustering 165
Ruilong Yang, Qingsheng Zhu, and Yunni Xia

Comprehensive Evaluation of Drilling Risk Based on Unascertained Measurement 170
Jun'e Liu, and Rong Wang

Spatial Information Based Medical Image Registration Using Mutual Information 174
Benzheng Wei, Zaitao Liu, and Xin Peng

Study of Awareness-based Knowledge Communication 178
Li Wang

Research on Tag-based Collaborative Filtering Strategy 182
Shuo Wan, and Huizhong Qiu

An SVM model for Water Quality Monitoring Using Remote Sensing Image 186
Wei Huang, Fengchen Huang, and Jing Song

A Study on High-Strength Communication Scheme Based on Signed Digital Envelope 190
Wenping Guo, Ying Chen, and Xiaoming Zhao

A Framework for Credibility Evaluation of Web-Based Competitive Intelligence 193
Jie Zhao, and Peiquan Jin

Power Aware Job Scheduling with Quality of Service Guarantees: A Preliminary Study 197
Congfeng Jiang, Xianghua Xu, Jian Wan, Jilin Zhang, and Yinghui Zhao

Networked Manufacturing Resources Optimization Deployment for Complicated Parts 201
Wenli Peng, Zhongbao Qin, and Wenni Zhang

The Annihilator and its Structure in Lattice Implication Algebras 206
Hua Zhu, Weifeng Du, and Jianbin Zhao

Combining Permission Based Delegation Model with Chinese Wall 210
Chunxiao Ye, Fawen Yu, Tinsen Huang, Haibo Hu, and Hong Xiang

Research on Security Policy and Framework 214
Dongliang Jiao, Lianzhong Liu, Shilong Ma, and Xiaoni Wang

An Adaptive Scheduling for QoS Enhancement Mechanism in IEEE 802.11 WLAN 218
Aiyun Zhan, Aihan Yin, Yueli Jiao, and Qingmiao Zhang

Research on Resource Locator in P2P Network 223
Zhihong Zhao, Guangxue Yue, and Libo Yang

Emotion Recognition of EMG Based on BP Neural Network 227
Xizhi Zhu

Design of Electron-agriculture System Base on Information Visualization and Windows Live Platform 230
Meiyu Zhang, and Fengfan Zou

Research of System-Switch Mechanism Based on sCPU-dBUS Architecture 234
Fengjing Shao, Xiaoning Song, and Rencheng Sun

Research and Application of Enterprise Search Based on Database Security Services 238
Yi Huang, Xinqiang Ma, and Danning Li

Study of Electromyography Based on Informax ICA and BP Neural Network 242
Guangying Yang, and Shanxiao Yang

ECG Pattern Recognition Based on Wavelet Transform and BP Neural Network 246
Shanxiao Yang, and Guangying Yang

Property Analysis of Petri Net Reduction 250
Chuanliang Xia

An Improved Volumetric Compression Algorithm Based on Histogram Information 254
Liping Zhao, and Yu Xiang

An Object Extraction Method Based on Oscillatory Network 258
Min Li, Lizhong Xu, and Min Tang

CNC System of MKS1632 External Cylindrical Grinding Machines Based on Embedded 262
Xuyan Zhou, Bin Tan, Bing Zhu, XuanGe Peng, and Yonghua Pei

Parabola Interpolation With Adaptive Error Compensation 267
Guangming Yang, and Fengqi Yu

Cloud Trust Model in E-Commerce 271
Min Wu

Database Auditing Design on Historical Data 275
Narongrit Waraporn

Author Index 282

Message from the Symposium Chairs

The 2010 International Symposium on Networking and Network Security (ISNNS 2010) was the Second in the annual series that started in Qingdao, China, during Oct. 28-30,2009(ISNNS 2009).

In the field of networking, the specialist area of network security consists of the provisions made in an underlying computer network infrastructure, policies adopted by the network administrator to protect the network and the network-accessible resources from unauthorized access, and consistent and continuous monitoring and measurement of its effectiveness (or lack) combined together.

The 2010 International Symposium on Networking and Network Security will bring together researchers from academia and industry, who are interested in the emergent field of information processing. We are soliciting papers that present recent results, as well as more speculative presentations that discuss research challenges, define new applications, and propose methodologies for evaluating and the roadmap for achieving the vision of Networking and Network Security.

Welcome to ISNNS 2010.Welcome to Jinggangshan, China, 2-4 April 2010. The 2010 International Symposium on Networking and Network Security is co-sponsored by Jinggangshan University, China;Peoples'Friendship University of Russia;Feng Chia University, Taiwan; Nanchang HangKong University, China; East China Jiaotong University, China;Jiaxing University, China; Qingdao University of Science & Technology, China; Academy Publisher of Finland, Finland; Hunan Agricultural University, China. Technical Co-Sponsors of the conference are IEEE; IEEE Shanghai Section. Much work went into preparing a program of high quality. We received 147 submissions, every paper was reviewed by 2 program committee members or 3 reviewers, and 79 papers are selected as regular papers, representing a 54% acceptance rate for regular papers. From these 79 research papers, through two rounds of reviewing, the guest editors selected 9 papers as the Excellent papers will be published by the special issues on Journal of Computers (EI Compendex, ISSN 1796-203X), Journal of software (EI Compendex, ISSN 1796-217X), Journal of Multimedia (EI Compendex, ISSN 1796-2048), Journal of Networks (EI Compendex, ISSN 1796-2056).

The purpose of ISNNS 2010 is to bring together researchers and practitioners from academia, industry, and government to exchange their research ideas and results and to discuss the state of the art in the areas of the symposium. In addition, the participants of the main conference will hear from renowned keynote speakers IEEE Fellow Prof. Jun Wang from Chinese University of Hong Kong, Hong Kong; IEEE Fellow Prof. Derong Liu, Editor-in-Chief of IEEE Trans. on Neural Networks from Chinese Academy of Sciences,China; IEEE & IET Fellow Prof. Chin-Chen Chang from National Chung Hsing University, Taiwan, and Chair of IEEE Shanghai Section, Prof. Junfa Mao from Shanghai Jiaotong University, China.

We would like to thank the program chairs, organization staff, and the members of the program committees for their hard work. We hope that ISNNS 2010 will be successful and enjoyable to all participants.

We thank Sun, George J. for his wonderful editorial service to this proceeding.

We wish each of you successful deliberations, stimulating discussions, new friendships and all enjoyment Jinggangshan, China can offer you. While this is a truly remarkable Symposium, there is more yet to come. We look forward to seeing all of you next year at the ISNNS 2011.

Fei Yu ,Guangxue Yue, Ming Leng,Xuange Peng

Second International Symposium on Networkingand Network Security

Organizing Committee

Honorary Chairs

Chin-Chen Chang, *National Chung Hsing University, Taiwan (IEEE & IET Fellow, Editor-in-Chief Journal of Computers)*

Derong Liu, *Institute of Automation, Chinese Academy of Sciences,China (IEEE Fellow, Editor-in-Chief of the IEEE Transactions on Neural Networks Associate Editor of, the IEEE Computational Intelligence Magazine, and the IEEE Circuits and Systems Magazine)*

Jun Wang, *Chinese University of Hong Kong, Hong Kong (IEEE Fellow, Associate Editor of the IEEE Transactions on Neural Networks;IEEE Transactions on Systems, Man, and Cybernetics – Part B;IEEE Transactions on Systems, Man, and Cybernetics – Part C)*

Junfa Mao, *Chair of IEEE Shanghai Section, Shanghai Jiaotong University, China*

Program Committee Chairs

Wenlang Luo,*Jinggangshan University , China*

Jian Shu,*Nanchang HangKong University, China*

Haiwen Liu, *East China Jiaotong University, China*

Yun Liu,*Qingdao University of Science & Technology, China*

Guiping Liao, *Hunan Agricultural University, China*

Organizing Chairs

Xuange Peng, *Jinggangshan University, China*

Jiexian Zeng, *Nanchang HangKong University, China*

Juefu Liu, *East China Jiaotong University, China*

Huojiao He, *Jiangxi Agricultural University, China*

Guangxue Yue, *Jiaxing University, China*

Jun Zhang, *Guangdong University of Business Studies, China*

Finance Chairs

Fei Yu, *Peoples' Friendship University of Russia, Russia*

Guangxue Yue, *Jiaxing University, China*

Publication Chair

Fei Yu, *Peoples' Friendship University of Russia, Russia*

Xuange Peng, *Jinggangshan University, China*

Roy Ng, *Ryerson University, Canada*

Local Organizing Chairs

Ming Leng, *Jinggangshan University, China*

Second International Symposium on Networkingand Network Security

Committee Members

Prof. Chen Xu, Hunan University, China
Prof. Chia-Chen Lin, Providence University, Taiwan
Prof. Chin-Chen Chang, National Chung Hsing University, Taiwan
Prof. Chu-Hsing Lin, Tunghai University, Taiwan
Prof. Dengyi Zhang, Wuhan University, China
Prof. Derong Liu, University of Illinois at Chicago, USA
Prof. Dingguo Wei, Guangdong University of Business Studies, China
Prof. Gao Yan, Henan Polytechnic University, China
Prof. GOLODOVA ZHAN NA, Peoples'Friendship University of Russia, Russia
Prof. Guangxue Yue, Jiaxing University, China
Prof. Guiping Liao, Hunan Agricultural University, China
Prof. GuoZhu Liu, Qingdao University of Science & Technology, China
Prof. Haiwen Liu, East China Jiaotong University, China
Prof. Hui Sun, Nanchang Institute of Technology, China
Prof. Huojiao He, Jiangxi Agricultural University, China
Prof. Jeng-shyang Pan, National Kaohsiung University of Applied Sciences, Taiwan
Prof. Jian Shu, Nanchang HangKong University, China
Prof. Jie Lin, Tongji University,China
Prof. Jiexian Zeng, Nanchang HangKong University, China
Prof. Jiliu Zhou, Sichuan University, China
Prof. Jinyi Fang, Guilin University of Electronic Technology, China
Prof. Juefu Liu, East China Jiaotong University, China
Prof. Jun Chu, Nanchang HangKong University, China
Prof. Jun Wang, Chinese University of Hong Kong, Hong Kong
Prof. Jun Zhang, Guangdong University of Business Studies, China
Prof. KARPUS NIKOLAY, Peoples'Friendship University of Russia, Russia
Prof. Lei Shi, Zhengzhou University, China
Prof. Limin Sun, Institute of Software, Chinese Academy of Sciences, China
Prof. Martha Russell, Stanford University, USA
Prof. Ming LI, Nanchang HangKong University, China
Prof. Naiping Hu, Qingdao University of Science & Technology, China
Prof. Qiang Liu, Qingdao University of Science & Technology, China
Prof. Qingling Li, Qingdao University of Science & Technology, China
Prof. Roy Ng, Ryerson University, Canada
Prof. Tinglei Huang, Guilin University of Electronic Technology, China
Prof. Tzong-Chen Wu, National Taiwan University of Science and Technology, Taiwan
Prof. Wanqing Li, University of Wolllongong, Australia

Prof. Weidong Zhao, Fudan University，China
Prof. Weihui Dai, Fudan University，China
Prof. Wen Chen, Shanghai Jiaotong University, China
Prof. Wenming Huang, Guilin University of Electronic Technoloqy, China
Prof. Xiaoli Wang, Tongji University, China
Prof. Yiqin Lu, South China University of Technology, China
Prof. Yongjun Chen, Guangdong University of Business Studies, China
Prof. Youfu Du, Yangtze University, China
Prof. Yu-Chen Hu, Providence University, Taiwan
Prof. Yung-Kuan Chan, National Chung Hsing University, Taiwan
Prof. Yuping Hu, Guangdong University of Business Studies, China
Prof. Zhijian Wang, Guangdong University of Business Studies, China

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 001-004

Shot Identification Based on DBN

Yun Liu, Xueying Liu, and Chao Huang
College of Information Science and Technology of
Qingdao University of Science and Technology, Shandong province, China
Email:Lyun-1027@163.com
Email:{ lxying2009, hchaopro }@qq.com

***Abstract*—A new method to identify the shots in basketball video based on Dynamic Bayesian Network (DBN) has been presented in this paper. Using the DBN, the semantic gap between low-lever features and high-lever semantic will be resolve. To identify the ball and the basketball hoop, Kalman filter and SURF (Speed up Robust Features) is applied separately. At last the DBN is employed to identify the shot events. Experimental results have shown our proposed method is effective for basketball event detection.**

***Index Terms*—shot identification; mean-shift; DBN ;SURF**

I. INTRODUCTION

In recent years, the basketball game has been more and more popular, the technology on real-time broadcast has developed rapidly and the network has been widely used, all above leads to the remarkable increase of basketball video data which people can get. In this paper we propose a novel basketball processing method that can satisfies this requirement and generate the summaries of the video.

With remarkable development in multimedia systems, many sports applications came into birth. The huge amount of data that is produced by digitizing sports videos demands a process of data filtration and reduction. These needs were addressed by the applications such as video summarization and highlight event extraction [1].

Currently, there is some work report in the domain of event identification for soccer video[2], baseball video [4] and tennis video[4]. In [2], Xu eta1 presented a framework for soccer video structure analysis and event detection based on grass-area-ratio. Rui et al. [3] developed effective techniques to detect an excited announcers' speech and baseball hits from noisy audio signals, and fuse them to extract exciting segments of baseball program.

To detect and track the ball is the primary work of most of the basketball video analysis. In soccer video, contrasting with the simple background, the ball possession and event detection has been crucially analyzed in [8].

From the above refers and *ref* [9] we can see that most of the research in sports video processing assumes a temporal decomposition of video into its structural units such as clip, scenes, shots and frames similar to other video domain including television and films.

Here we apply SURF (Speed Up Robust Feature)[11] to identify the basketball hoop and use the mean-shift and Kalman filter to identify and track the basketball.

This paper is supported by Natural Science Fund of Shandong (Y2008G09).

AP-PROC-CS-10CN006

The SURF detector is based on the Hessian matrix[6], but uses a very basic approximation. It relies on integral images to reduce the computation time and we therefore call it the 'Fast-Hessian' detector.

This paper is organized as follows. In section 2 the shot boundary will be detected based on the difference of the frames. In section 3, the SURF features will be extracted for hoop identification. Ball will be identified and tracked using Kalman filter and mean-shift and then the DBN is used to classify the event we defined at section 1. Finally, section 5 concludes the paper and introduces the directions for future research.

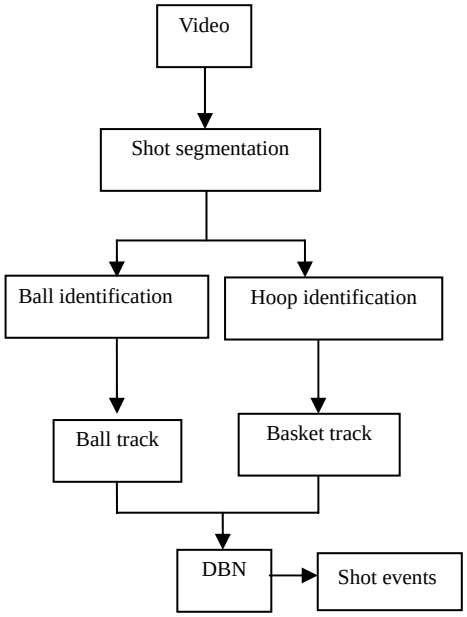


Figure1. The flow chart of the method

II. SHOT BOUNDARY DETECTION

To detect the shot boundary needs to define a variable which value is very big at the boundary while very small in shot. If the scene cuts from one shot to another, the difference between them will be very large and so is the standard deviation, for that reason we select standard deviation to measure the shot transition.

In Fig.2 we have shown different types of the frames we classify in our paper and their histograms. (a) are the frames in play and their histogram, (b)are the frames in close up and their histograms, (c) is the difference of the histogram. At the boundary the standard is very large as a single peak as shown in Fig.2.

(a) The frames In-play and their histogram

(b)The frames of close-up and their histogram

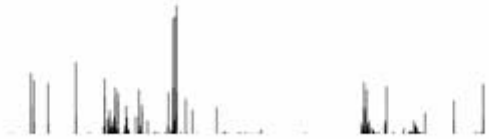


Figure 2. The frames and their histogram of the video and the last figure is the difference between the frames

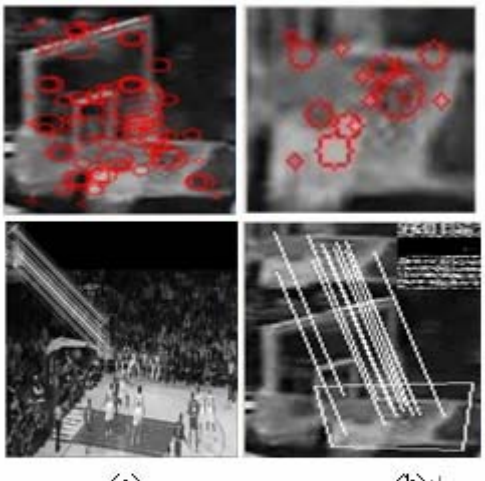
III. FEATURE EXTRACTION FOR HOOP IDENTIFICATION

The basketball hoop tracking is a coarse-to-fine approach. Firstly we choose SURF[6] method to extract features of the backboard and then identify the basketball hoop in the extracted area. For it is of better speed and accuracy compared with other features, we further extend it to our detection procedure. According to Bay *et al.*[7], SURF extraction is divided into three steps, which are feature detection, orientation assignment and descriptor components.

A. Feature extraction

Detection of SURF is based on the Hessian matrix, it is defined as follows

$$H = \begin{bmatrix} L_{xx}(\hat{x},\sigma) & L_{xy}(\hat{x},\sigma) \\ L_{xy}(\hat{x},\sigma) & L_{yy}(\hat{x},\sigma) \end{bmatrix} \quad (1)$$



(a) (b)

Figure3. The tracking results

(a) Identify the backboard

(b) Identify the basketball hoop

In which $L_{xy} = \frac{\partial}{\partial x^2} g(\sigma) * I(x,y)$, and $g(\sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}$ Others parameters are similar in the calculation.

B. Hoop identification

From the video we find that the backboard is entirely different from the background region. In every frame we track the backboard to find the candidate area of the basketball hoop with their matching SURF feature. To find their precise position we apply the SURF feature once again. The tracking result is shown in Fig.3.

IV. DETECTION AND TRACKING OF THE BASKETBALL

The aim of an object tracker is to generate the trajectory of an object over time by locating its position in every frame of the video. Object tracker may also provide the complete region in the image that is occupied by the object at every time instant.

A. The ball detection

In this paper Kalman filter is used to build the kinematical model and predict the object's motion so as to narrow the scope of matching, thereby accelerate the speed of target matching. Then the target in the next frame within the specified area is matched in order to establish the association between targets. At last, update the kinematical model to form the moving target chain and get moving target trajectory.

With the Kalman filter we can predict the ball and we can select the filter depending on the actual condition, and then through the above equations estimate the system state.

B. Tracking the ball

Mean Shift is a classic object tracking algorithm with non parameter estimation which is widely used in object tracking. Kalman filter is an algorithm to estimate the linear minimum variance inaccuracy of a dynamic system's states sequence. Because of the complexity of the basketball match video, ball detection and tracking is a complex approach. First, Kalman filter is used to predict a position which is to be the initial search center in Mean-Shift algorithm to track. Then, Mean-Shift gets a new target position which is used to be the input parameter of the next Kalman filter.

Tracking the objects by converging at the real location, mean shift continuously iterative calculate the mean-shift vectors by its astringency, for which the tracking algorithm requires the overlap region exist in the target standard and the possible target. In this paper Kalman is used to predict the location of the target to confirm the stability and robustness of the Mean-shift tracking.

V. EVENT ANALYSIS BASED ON THE DBN

Bayesian net is a directed acyclic graph, which reflects the dependencies on the probability inter the variables. Dynamic Bayesian Network combines Bayesian Net with the time information to form a new random

model that can deal with time series data based on the Probability Network. It considers factors outside the system as well as the inter-link

A DBN consist of two parts (a priori Network and Transition Network), having which a DBN will be formed in any length. To facilitate the processing, we suppose the DBN can meet the conditions: (a).the network topology dose not change over time; (b) the network meet the first-order Markov condition. Satisfying the above two conditions the DBN can be seen as unfolding in time sequence.

Compared with the existing works, we use a stronger temporal signal processor-----DBN (Dynamic Bayesian Network)[16]to identify the shot event. On the one hand, considering the transition probability between the various moments, dynamic Bayesian networks extend Bayesian network modeling capabilities of the timing signals. On the other hand, dynamic Bayesian networks allow the use of multiple state variables at the same point in time, while Hidden Markov model uses only one state variable. Based on these considerations, we believe that Dynamic Bayesian Networks is more suitable for sports video content analysis, especially for the semantic analysis of events and their mutual relations.

A. The introduction of the DBN theory

Suppose $P(X_t \mid X_{t-1})$ present the probability of the current status in the condition that the last status has been known. X_t^i is the i^{th} variable's value in t, and $P_a(X_t^i)$ is its parent node. If there are N variables in two status,

$$P(X_t \mid X_{t-1}) = \prod_{i=1}^{N} P(X_t^i \mid Pa(X_t^i)), \quad (3)$$

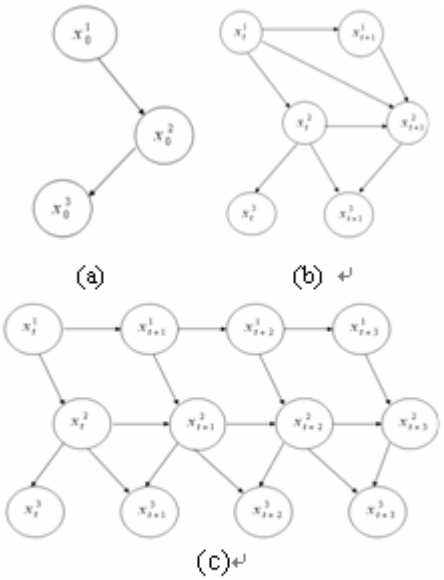


Figure 4 Schematic diagram of the DBN

(a) The initial network of DBN

(b) The transition network of DBN

Similarly the other nodes' joint distributions are:

$$P(X_{1:T}^{(1:N)}) = \prod_{i=1}^{N} P_{B_0}(X_1^i \mid Pa(X_1^i)) \times \prod_{t=2}^{T} \prod_{i=1}^{N} P_B(X_t^i \mid Pa(X_t^i)) \quad (4)$$

In the context of the basket has been identified, we track the ball's location and measure the Mahalanobis distance[15] between them. When the distance satisfies a condition, we judge the relative position of the previous state and the next state, and then deduce the score event.

B. The training process

The state-space needs some parameters to define its transmission probability model and observation model, and training can estimate the parameters.In the condition that the network structure is unknown and the samples are totally difference, Friedman proposed the structured EM algorithm[16] Three parameters will be taken into account in the node that is the distance between the ball and net, the pre status of the node, the next status of the node.

C. The event identification

Having the conditional probability and the prior probability for each node, we can interpret the video shots. The video shots are processed by several video analyzers such as the basketball and the basketball hoop identify by the SURF[6] descriptor and the mean-shift algorithm. These low -lever features are extracted as the input of the DBN. We used the evidence propagation procedure from the low-lever features to the relationship features to identify the video shot events and scoring events.

VI. EXPERIMENT RESULTS

In this section, we describe the experiments aimed at evaluating the proposed method, which integrated into a system that was tested using two basketball video sequences.

The test videos are two basketball videos NBA games with a total length of fifty minutes, which is segmented into video shots from which we select about 332shots to analyze.

In our experiments results comparison all the ground truths are assigned manually. And we demonstrates the results by using the precision and the recall defined as follows,

$$\text{precision} = \frac{\text{number of correct detection}}{\text{number of correct detection} + \text{number of false drop}} \quad (5)$$

$$\textit{recall} = \frac{\text{number of correct detection}}{\text{number of correct detection} + \text{number of miss}} \quad (6)$$

A. Shot segmentation

From table 1 the accuracy of our shot boundary detection algorithm is about 93.7% (the number of correctly detected shots divided by the number of total shots, which is counted manually). Table 1 lists the performance in terms of the precision and recall.

TABLE 1. THE RSULT OF SHOTS BOUNDARY DETECTION

	manually	auto	precision	recall
Video 1	147	136	94.4%	93%
Video 2	253	221	96%	92%

The dissimilarity matching on two frames are all based on the histogram of the frames shown in Fig.2. A promising performance, Recall 85%-93%, and precision 90%-95%, has been achieved.

B. Events detection

After the shot segmentation stage we cluster the shots manually and auto by the method and the results are shown as follows:

TABLE 2. THE RSULT OF SHOTS CLASSIFICATION

	manually	auto	precision	recall
In play	180	166	96.4%	91%
Close-up	420	412	98%	96%
Free-throw	18	14	66%	80%

In Fig.5 we present a user interface of score events.

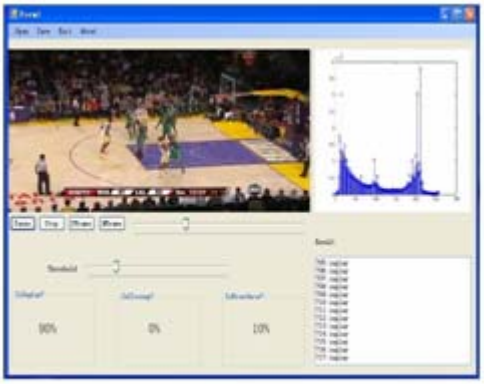


Figure 5 The shot identification

CONCLUSIONS

We have proposed a novel basketball video shot identification system based on DBN. The main contribution of this paper is to develop an inference system consisting of some linkages between unobservable concepts and observable concepts. We integrate the feature extraction with inferable DBN to fill the gap between the low-level visual domain and the high-level semantic classes. In the experiments, we have proved that our system can identify the shot events and scoring events effectively.

REFERENCES

[1] M. H. Kolekar, K. Palaniappan Semantic Event Detection and Classification in Cricket Video Sequence in Sixth Indian Conference on Computer Vision, Graphics & Image Processing 2008

[2] P. Xu, L. Xie, S-F Chang, A. Divakaran, A. Vetro, and H. Sun, "Algorithm and Systems for Segmentation and Strncture Analysis in Soccer Video", In Proc. of IEEE International Conference on Multimedia and Expo, Tokyo, Japan, Aug. 22-25,2001.

[3] Y. Rui, A. Gupta and A. Acero, "Automatically Extracting Highlights for TV Baseball Programs'', In *Proc. OfACMMultimedia*, Los Angeles, CA, pp. 105-115,2000.

[4] G. Sudhir, J. C. M. Lee, and A. K. Jain, "Automatic Classification of Tenis Video for High-level Content-based Retrieval", In *Proc.* of *IEEE International workshop on Content-Based Access of Image and Video Database, 1998.* pp. 81-90

[5] M. Xu, L.Y. Duan, C.S. Xu, M.S. Kankanhalli, and Q. Tian."Event Detection in Basketball Video using Multiple Modalities," Proc. IEEE-PCM'03, Singapore, Dec. 2003.

[6] H. Lu and Y.-P. Tan, "Content-based sports video analysis and modeling," in Proceedings of 7th International Conference on Control, Automation,Robotics and Vision (ICARCV '02), pp. 1198–1203, Singapore, December 2002.

[7] Y. Fu, A. Ekin, A. M. Tekalp, and R. Mehrotra, "Temporal segmentation of video objects for hierarchical object-based motion description," IEEE Transactions on Image Processing, vol. 11, no. 2, pp. 135–145, 2002.

[8] X.Yu, Q. Tian, and K. W. Wan. "A novel ball detection framework for real soccer video," Proc. IEEE-ICME'03, pp. 265-268, 2003.

[9] V. Chasanis, A. Likas, and N. Galatsanos. Scene detection in videos using shot clustering and symbolic sequence segmentation. *in IEEE Workshop on Multimedia Signal Processing*, 2007.

[10] H. Bay, T. Tuytelaars, and L. Van Gool. SURF: Speeded up robust features. Proceedings of Ninth European Conference on Computer Vision(ECCV), Graz, May 7, vol. 13:pp. 404-417, 2006.

[11] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," Conference on Computer Vision and Pattern Recognition, pp. 511-518, 2001.

[12] Cheng Y. Mean shift，mode seeking and clustering[J] IEEE Transactions on Pattern Analysis and Machine Intelligence. 1995. 17(8)：79 — 799

[13] M.Luo,Y.F.Ma,and H.J.Zhang,"Pyramidwise Structuring for Soccer Highlight Extraction,"Proceedings of IEEE Pacific-Rim Conference on Multimedia,2003.

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 005-008

Discussion on the Current Development and Reform Ideas about the Competitive Wushu Routine at Colleges in Jiangxi Province

Youru Yu[1], Zhipeng Xiao[1], Weimin Wang[1], and Hongmei Deng[1]
[1] School of P.E., Jinggangshan University, Ji'an, 343009, China
yuyouru@jgsu.edu.cn , zpxiao@jgsu.edu.cn , wmwang@jgsu.edu.cn , hmdeng@jgsu.edu.cn

***Abstract*—Based on the new technology development direction of "high, difficult, beautiful, new" of the Taolu competition in Chinese martial arts, this paper analyses the current situation to the development of the competitive Wushu Routine at colleges in Jiangxi province by way of questionnaires, inquires and comparison. It discovers that the development level of the competitive Wushu Routine at colleges in Jiangxi province is worrying. The paper also discusses the causes of this phenomenon, and points out a new way to accelerate the development of the competitive Wushu Routine at colleges in Jiangxi province.**

***Index Terms*—Jiangxi Province, Colleges, the Competitive Wushu Routine, Reform and Development.**

I. INTRODUCTION

The Competitive Wushu Routine is undergoing an important historical period, in which it is going along the new technology development direction of "high, difficult, beautiful, new" and showing its unique fascination[1][2]. So it is no doubt to put forward higher request for selecting the Taolu players scientifically and offering better training methods. Universities should not only provide advanced facilities and equipments for Taolu players and their training, but also arouse the teachers and the students' participation awareness of boxing and raise the colleges' reputation by way of the competitive Wushu Routine. During the whole process of Wushu competition in Jiangxi province 12th Sport Games, we watched and found that the ability of the high level group of Wushu from undergraduate colleges in Jiangxi provinces was far worse than that of national Wushu team and the level of the common lever group, compared to the three previous university groups in Jiangxi, showed a decline tendency in a whole. This is not optimistic for the Taolu competition in Jiangxi's universities. Therefore, we analyze the restricted factors for the common level group of Wushu from undergraduate colleges in Jiangxi province from many perspectives and point out a new way to accelerate the development of the competitive Wushu Routine at colleges in Jiangxi province

II. RESEARCH OBJECTS AND METHODS

A. Research objects

Research objects are composed of the whole athletes, who participated in the Wushu competition in Jiangxi province 12th Sport Games. All the players are composed of 12 male competitors, 10 female competitors in high level group, 55 male competitors and 33 female competitors in undergraduate ordinary group.

B. Research methods

1) Observation

Observation method means to watch each competitor's show in live, who are participated in the Wushu competition in Jiangxi province 12th Sport Games.

2) Questionnaire method

Questionnaire method is carried out among 11 coaches, 23 competitors from colleges in the sports event. The recovery rates are 100% and the effective rate is about 94%.

3) Interview

Interview method is carried out on-site visit with one chief referee, six referees, ten local coaches.

4) Comparison

Comparison method is based on two levels of competitions, according to the competition rules and grading standards examined and approved by Chinese martial arts association of Wushu in 2006. One is to compare the two groups between the high level group of Wushu from undergraduate colleges in Jiangxi province and the athletes participated in the Taolu match of the 10th National Games. Another is to compare the two groups between the common lever group this time and the three previous times.

III. RESULTS AND ANALYSIS

A. Poor state of the athletes' ability and achievement in the sports event

1) Low level of Martial arts choreography

Martial arts choreograph reflects the technical characteristics of the art of attack and defense. Its action elements are generally divided into three basic action types. The basic actions belonging to "form" include series Wushu specifications, such as hand form, step form and body form etc., which should be complied with the high quality of the specification of Wushu competition rules defined. Those basic actions belonging to "essentials" include series Wushu movements, such as hand technique, footwork, leg technique, body work and eye technique. In the process of dynamic movement, the athletes' movements should be soft and firm, while conforming to the outfit and inner side advantage. The

ones belonging to jumping is full of light footsteps and skilled movements. When the athletes take off jumping, they should jump as high and light as possible[3][4]. In the Wushu competition, the movements acted by the high level group of men and women's athletes are in agreement with the requirement of the match rules basically, but the movement acted by common lever group of men and women's athletes is in next levels on average and lack of basic Wushu skills. Their forms are lack of specification. Their essentials are lack of attack and defense connotation as well as the stress of energy and their actions of jumping are lack of higher state and elegant appearance state, and so on.

2) Low degree of difficulty of jumping movement

In the current Wushu competition rules, the three levels of difficulty of movement pattern are clearly defined, as can assess how a competitive Wushu athlete masters the level of difficulty and measure an athlete physical condition. In every game in martial arts, the jumping difficulty acted by the high level group is not too high, and their ground movements are lack of stability. While the athletes of common lever group of can only accomplish some simple jumping movements, such as flying kick with body turn, 3600 whirlwind kick, 3600 flying lotus leg swing etc and their movements are lack of standard. Only seven of male athletes do the action of side somersault in the lateral pattern matches.

3) Lack of beauty for the skill movements

Attack and defense is the essential feature of the martial arts. Therefore, when the athletes perform, their actions should be perfectly reflected in action specification, power and romantic charm for different school Quans, the holding methods of different weapons. At the same time, their skill movements should represent the beauty of martial arts movement skills in the light of the nature law of Wushu, so as to show the style and features of a wushi. The Wushu Taolu competition rules define the specific criteria of evaluation for different movement lever. However, the athletic ability and performance of common lever group in the sports event are not good, which is mainly embodied in several aspects: lack of the coordination between the body motion and the inner mentality, vital energy and spirit, lack of the beauty of smooth channeling of energy, moving like flowing water, guiding by consciousness, unity of form and spirit, lack of charming of martial arts movements skills. In addition, their actions are lack of accuracy, exertion of force, hardness and softness and beauty of rhythm.

4) Lack of novelty of the establishing action and artistic ideas

In the development trend of Competitive Wushu Routine, the new is put forward to the coaches and competitors about their creative thinking. After watching this year’s match, we can find that the movements from establishing action to artistic ideas performed by the common level group of Wushu from undergraduate colleges are lack of novelty. The level of two-fifths of competitors remains the 2nd international competitive Taolu, even some competitors remove only the important parts about side somersault and flying gun with body turn. The Survey Research shows that the establishing action and artistic ideas of this match are worse than those of the Taolu match of the 10th National Games.

B. The factors affecting the Taolu competition in Jiangxi’s universities

1) The shortage of College Wushu talent for the Special Test Project of Wushu sport excluded from college entrance examination

By means of questionnaire and interviews with experts, we find that the development of Wushu education in Jiangxi at colleges is facing atrophy trend since the Special Test Project of Wushu sport excluded from college entrance examination in 2000.The primary cause has two: First, some middle school students in the nature of good level of martial art have no chance to study in college through sports college entrance because of Wushu sport excluded from college entrance examination; second, the middle school students are not motivated to practice Wushu Taolu in some way because of the Wushu sport canceled, which causes that most students of P.E. colleges almost have no basic skills.

2) Competitors of Wushu Routine Training in a short time, and the basic skills in low level

By means of the survey of 11 college coaches, we find that the undergraduates of P.E. at colleges in Jiangxi are weak in basic skills and lack interest in practicing Wushu. To participate in Wushu Taolu competition, the best are selected inside the dwarfs, then they practiced well several routines after several months hard training. They wanted to have good achievement in the competitions by harmonious society, also they wanted to learn each other, now matter win or lose. In the other hand, an interview is carried out on-site visit with twenty-three competitors, we find that 16 of 23 players really begin to practice Wushu Taolu when they took the compulsory course of Wushu. Even worse, one player never touched on Wushu before she trained for Wushu Taolu competition, she only spent her spare time on intensive training for the large-scale provincial tournament. More notably, the player gained the third in 42-pattern Taoji Quan of woman group at a common level. In the light of the above facts analysis, it shows the basic skills of the Wushu Taolu players at colleges in Jiangxi.

3) Wushu Routine in Children's Amateur Sports School of county and city at the edge of dying

The current situation of amateur sports schools, provided Wushu Taolu training, at various countries in Jiangxi province is same as the saying goes, “The gate of the town is on fire, affecting the fish in the pool.” When the interview is carried out on-site visit with 11 local coaches, we find out that: because of the Special Test Project of Wushu sports excluded from college entrance examination, the students heavily dropped out of local Wushu schools, which caused the survival of amateur sports schools of physical training at various countries in Jiangxi province to the edge of extinction. For actively organizing team to take part in Wushu Taolu competition of the youth group, five of nine local amateur sports schools enlarged an army from other Wushu schools to

training for the Wushu competition in Jiangxi province 12th Sport Games. The survey suggests that the coaches from 78.3% of the local amateur sports schools were forced to look for another job for the dropout of the students.

IV. Some ideas about speeding up the upgrading of Competitive Wushu Routine at colleges in Jiangxi province

In order to change the present state of the competitive Wushu Routine at colleges in Jiangxi province in the new century and make it boom, several reform ideas are proposed.

A. Adding Wushu Taolu in the college entrance examination of P.E.

For improving the development level of the competitive Wushu Routine at colleges in Jiangxi province, it is necessary for Wushu Taolu to be included in College P.E. enter examination. There are two main reasons. First, as a starting point, it should be broken the current situation of Wushu talent and construct a platform of introducing talented people. Second, with the College P.E. enter examination as the leading role, it should provide impetus to the development of Wushu Taolu as middle school in Jiangxi so as to provide the excellent reserve force for colleges. Furthermore, it can activate the education markets in amateur sports schools of physical training in various countries in Jiangxi province. At the same time, when parents'concerns are eliminated, they will let their children practice martial art in amateur sports schools and it will promote the students' level of Wushu Routine and lay a foundation for providing more good seeded players of Wushu Taolu.

B. Improving the curriculum construction of Wushu at colleges in Jiangxi and offering the training classes of Wushu Taolu

With the reform in the curriculum of P.E. Major at Colleges in our country, the class hours of Wushu Taolu are reduced relatively, which hinders the development of the competitive Wushu Routine. For this reason, in view of the poor present situation of competitive Wushu Routine at colleges, the colleges can enroll the seeded players from their own students and offer a long-term training classes of Wushu Taolu to promote the scale and level of Competitive Wushu Routine at colleges to a new stage while the colleges go along with the he reform in the curriculum of P.E. Major.

C. With the target of the new technology development direction of "high, difficult, beautiful, new", the seeded players selected.

The Wushu Taolu competition rules issued in 2006 further focuses on the new technology development direction of "high, difficult, beautiful, new", which requires the present competitive Wushu Routine to reveal the following characteristics: high exercise quality, big jump difficulty, beautiful skill movements, novel establishing action and artistic ideas. In the light of the slow growth of Competitive Wushu Routine at colleges in Jiangxi province, we should think about the players' qualities from several perspectives: acceleration, force, suppleness, harmony, agility, viqour, figure and looking when we select the first-rank competitors who participate in the Wushu Routine match, so as to lay a foundation for making good mark.

D. Strengthening the coaches' awareness of innovation and promoting scientific training

With the technology development of the Competitive Wushu Routine, innovative ideas are not only the need for the Competitive Wushu Routine itself, but also the need for the competition adapted[1][5]. The "new" in the development direction of the Competitive Wushu Routine puts forward new requirements for creative thinking of the coaches and competitors. The coaches should constantly adapt to changes in the rules of competition, break through the stereotypes and encourage their competitors to innovate on the basis of their own characteristics, without violating the aim of attack and defense. Also, the coaches can adopt innovative Wushu movements acted by the competitors and assure the novelty of the establishing action and artistic ideas. In their teaching, the coaches should actively guide the competitors to understand the intention of Taolu innovation, reasonably maser the instruction laws and avoid sport injury and the dull training methods.

E. Creating a healthy cultural environment for the Competitive Wushu Routine

To ensure creating a healthy cultural environment for the Competitive Wushu Routine, we should adopt three measures. First, the leaders of universities should give more attention to the Competitive Wushu Routine and give a strong support from two sides. One is to provide the training place equipped with advanced facilities, to increase the technological content of the Competitive Wushu Routine and to increase the research dollars. The other is to set the outstanding award gainer so as to activate the competitors to join in the training. Second, the leaders of universities should strengthen the communication each other, holding on the large-scale tournament to stimulate the creativity, initiative and enthusiasm of the participation in boxing, which can increase the chances of participation in big game for competitors. Third, the colleges and universities should hold on the Wushu Routine match or show of Wushu Team and present the charm of Chinese martial arts to arouse more teachers and students to participate in Wushu exercises.

V. Conclusion

The Competitive Wushu Routine at colleges in Jiangxi province should be increasingly recognized and supported from all levels of leaders and brought into the Special Test Project of college entrance examination again, which can be a shortcut to the talent selection of good level of martial art and promote the steady development of Wushu Routine at elementary and secondary schools and ctivate the education markets in

amateur sports schools of physical training at various countries in Jiangxi province so as to drive the vigorous development of Wushu Routine in Jiangxi province. In addition, with the Competitive Wushu Routine as the leading role, the leaders of universities should provide the human and material resources to dig coaches out of their innovation awareness and carry on the scientific selection of Taolu players and their training. Furthermore, Colleges and universities should strengthen the communication each other to promote the level of Wushu Routine.

ACKNOWLEDGMENT

The authors wish to thank the athletes and local coaches for their answering the Questionnaires and accepting the interviews, who participated in the Wushu competition in Jiangxi province 12th Sport Games.

REFERENCES

[1] JIANG Xia, HUANGTian-lin,WANGYi-ming,WANG Rui-lin,"Investigation About the Present Condition of Competitive Martial Arts Routine in Shaanxi Province",JOURNAL OF XI ′ AN PHYSICAL EDUCATIONUNIVERSITY, Vol. 24, No. 5, China, Sep· 2007, pp.35-38 .

[2] DING Baoyu,"A Study on Development of Technical Features of Routine Movements of Competi-tive Wushu in“Post-Olympic Games”Era"， Journal of Capital Institute of Physical Education, Vol.21 No.5 , China, Sep. 2009, pp: 528-531.

[3] WangMing-jian,"An Evolution Analysis of the TechnicalDegree ofDifficulty of Series of Skills and Tricks in CompetitiveW ushu",Journal of Chengdu SportUniversity, Vo.l 33,No. 6, China, 2007, pp:88-90.

[4] XIAO Hong-zheng,MA Min-qing,ZENG Yu-jiu,"Technical development of competitive Wushu routines",Journal of Wuhan Institute of Physical Education, Vol.40, No.8,China, Aug. 2006, pp: 77-79

[5] DU Guo-you,LIN Zhi-hua,"Development of competitive Wushu routines",Journal of Wuhan Institute of Physical Education,Vol.41, No.6, China, June 2007, pp: 79-82.

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 009-011

Analysis of Uncertainty Information for Several Kinds of Complex Systems

Jiqin Peng, and Jinfang Han
Science College, Hebei University of Science and Technology, Shijiazhuang 050018 ,Hebei China
pengjiqin8787@126.com , Jfhanemail@126.com

***Abstract*—In this paper, three kinds of uncertain complex system are discussed by analyzing uncertainty of information, and their respective connotation characteristic along with the main difference and connection are elaborated from theory method, which make it convenient for people to best distinguish and handle uncertain problem. Finally, a kind of uncertain information system just born and a vital problem remaining to be solved are provided.**

***Index Terms*—uncertainty System, Information, Random, Fuzzy ,Grey, Unascertained**

I. INTRODUCTION

With respect to the uncertainty of complex system, Professor Wang Qingyin [1] have partly discussed and inquired. Some study methods and mathematics relation between four kinds of uncertainty have been provided and discussed, which open up train of thought, establish basis for further recognizing and deeply studying complex system. The uncertainty of system is uncertain information System contain. Uncertain information is the information with uncertainty. Some uncertain information will accompany with the description of system information characteristic. In this paper, the uncertainty of system is introduced and discussed in other point of view; the respective characteristic of various kinds of uncertain system are expounded. The paper sum up, compare, contrast in many aspects to explain that which method is used to solve the uncertain problem of complex system under what situation (condition).

II. UNCERTAINTY OF SYSTEM INFORMATION

To identify a system is to identify its information characteristic, information characteristic is mainly the unity of system's essential factor, structure, function. A system can't be described and expressed without these information characteristic. But, in respective aspects from the generation of information to the description of information characteristic, the information grasped is uncertain owing to various reasons. For example, in the process of information generation, transmission and receiving, owing to the disturbance of outside disturbance source, some distortion phenomenon usually happen when source information arrive at clinic information through channel. When recognizing and describing system's information characteristic, owing to the objective thing's complexity, people subjective knowledge's limitation, natural language's intrinsic uncertainty, the quality's appearance, the quantity's provision, relationship's expression and law's find, people can't truthfully reflect objectively existing things. Distortion and no easily exact reflection make information uncertain. For uncertainty is universal in objective reality, people have to study, build up and apply uncertain information system. Via analyzing several kinds of uncertain information, different uncertain systems are expounded

Random Information

A lot of actual problems involve the analysis and design of system disturbed by environment uncertainty. The uncertain disturbance may be caused by different source. Among them, a kind of disturbance source is a certain causality that does not happen between condition and event because the condition is insufficient. Hence, it is uncertain whether the event happen or not (namely the event maybe happen or not). The kind of uncertainty is called randomness.

For random system, or in handling random problem in reality, the main mathematics tool adopted is statistical mathematics, that is, probability statistics (containing random process). It may be called the theory method of handling random system.

The characteristic of probability statistics is many data. Probability statistics faces uncertainty of infinite data but no law; it solves the problem of "big sample uncertainty", the data demanded is classical distribution. For example, the distribution used frequently is binomial distribution, Passion distribution, uniform distribution, normal distribution…etc. Among them, the normal distribution is used most. The basis of probability statistics is Cantor set, the set of "1" and "0". The elements of set have the property of Yes or not, the means of probability statistics is "statistical", namely, acquiring law via statistics. The goal of probability statistics is historic statistical law. A big sample studied contains all the data of reality and unreality time zone, reality time zone is transient relative to unreality time zone, therefore, via statistical way, all statistical data embody historic statistical law. Hence, the thinking way of probability statistics is "appear again". The information criterion of probability is infinite information. Because of footing "big sample uncertainty", probability statistics handle problem in infinite information space.

Fuzzy Information

Owing to the complexity of things, for many things, its boundary isn't obvious, characteristic is not explicit, so the information about them is uncertain, namely fuzzy. The things that have existed and happened in objective

world, are all certain things. But for some certain things, the information grasped in people's brain may be fuzzy .For example, people's face itself have not any uncertainty, but, even if the information in the husband (or wife)'s brain is fuzzy too, there is no quantitative certain data, but people can accurately distinguish the man known via the kind of fuzzy information. Therefore, we say, in present, the fuzziness of things considered is mainly uncertainty caused by the distinct definition and judgment standard that can't be given for something. Or say the difference between things display either this or that in middle transition. As inconsistent law expound "when the complexity of a system add, people's capacity to make it exact decrease, when arrive at a certain threshold, complexity and exactness repel each other, fuzziness accompany with complexity" [2].

In fuzzy system, or say when people handle fuzziness problem in reality, the mathematics tool adopted is fuzzy mathematics (namely fuzzy theory method). For fuzzy mathematics, its characteristic is depending on experience, Fuzzy mathematics faces uncertainty caused by inadequate knowledge, namely, solves the uncertain problem of "knowledge uncertainty". Depending on experience talked here is membership function is given via experience. In practice, constructing a continuous membership functions usually needs infinite experience data (there are exceptions in specific). Hence, we may say, the information criterion of fuzzy theory is experience information. The involved is experience's infinite information space (there are exceptions). The basis of fuzzy mathematics is fuzzy set. Fuzzy set are between 0 and 1, the characteristic value of element may be arbitrary number between "0" and "1". The means is bound value . The operation of fuzzy set "max" $\vee$, "min" $\wedge$, namely take a value of bound . For fuzzy set theory, the "demand" is distribution according to membership function, the goal is knowledge expression. The original intension of fuzzy set is the knowledge of conception is described with number relation by people, so the goal of fuzzy mathematics is knowledge expression. The thinking way of vague mathematics is denotation quantification. For the characteristic of fuzzy set is denotation is uncertain, fuzzy mathematics' thinking way is making uncertain denotation expressed by membership degree, namely, denotation quantification.

Grey Information

Problems in complex big system often involve a lot of information (data and knowledge). Because the noise in channel (objective) disturb, or the capacity of receiving system (subjective) is limited, We only infer the general range that information reflect, but didn't know all the confirmed content, in addition, when some problem is handled, all information concerned is unnecessarily or impossibly obtained (only a part is known), which is information incompleteness. Grey system is the system which partial information is distinct and partial information is indistinct. It can be said information incomplete system.

Grey system theory's intension is "small sample uncertainty". Grey system faces "the uncertainty" caused by a few data, namely, solves the problem of "a few data uncertainty". It demands data is arbitrary distribution. Its goal is "reality law". Because a few data studied is the data distributing in reality time zone. The data embody the reality law. The basis of grey system theory is "grey hazy set". Grey hazy set is the set, under a certain proposition, whose element is from distinct to indistinct, information is from a little to a lot, and can be replenished unceasingly; the set from grey to white, from abstract to concrete; the set can evolve, have "life", have "effectiveness for a given period of time"; the set is compatible to 0,1 property, cantor set and [0,1] property; the set have four kinds of forms (embryo, growth, maturity, actual evidence).[3], The Grey System theory is on the basis of information covering, that is, via information covering to describe, analyze, handle grey target whose information is incomplete, uncertain. The intension of information covering is to contain, cover information of given proposition with a group of information. The basis means of grey system theory is grey generation. Grey generation or generation is the data processing (concluding accumulation, ,transform, reject, interpolation…). Its goal is to provide comparable data for analysis; to provide reasonable data base for building model; to provide extremity uniform sample for making decision. Its characteristic is a few data; its thinking way is many angle of view; its information criterion is a little information. On the basis of "uncertainty of a few data", only a little information can be obtained, hence, grey system theory involves finite information space.

Above, three kinds of uncertain system and their theory (mathematics) method are discussed, their respective characteristic is analyzed from several aspects. The goal is make people to exactly apply uncertain system theory to solve real problem. The brief summary is given as follows

TABLE

aspect	**kind**		
	probability statistics	***fuzzy theory***	***grey theory***
intension	big sample uncertainty	knowledge uncertainty	small sample uncertainty
basis	Cantor set	fuzzy set	grey hazy set
characteristic	many data	experience（data）	a few data
means	statistical	Bound value	generation
according to	probability distribution	membership function	information covering
demand	classical distribution	function（distribution）	permit arbitrary distribution
goal	historical statistical law	knowledge expression	reality law
thinking way	appear again	denotation quantification	many angle of view
information criterion	infinite information	experience information	the least information

III. REMARK

There exists another kind of uncertain information, which our country's famous scholar Guangyuan Wang academician finds and first provided, that is, unascertained

information, called weak uncertain information too. “Unascertained information” is the information that is must be used but unknown when making a strategy owing to restriction of objective condition. Or say, owing to inadequate evidence (information) grasped by decision maker is not enough to ascertain thing’s real state and quantity relation, which lead to the unascertained information of pure objective knowledge. The system with unascertained information is unascertained system. For unascertained system, its mathematics tool is “unascertained mathematics”, its basis is unascertained set. Reader can consult reference [4-6]

It is noted that various kinds of uncertain system as well theory method is respectively discussed above, but in objective world, some big system may have not merely a kind of uncertainty, may have two or three kinds of uncertainty, which demands people plan as a whole in dealing with uncertain information. In aspect of handling system with two kinds of uncertainty, professor Wang Guanyuan and Zhang yue have established “fuzzy” random system theory[7-8], then, how to handle the problem of the other two or more kinds of information coexisting? It wait people to tap and open up.

ACKNOWLEDGMENT

This paper is supported by the National Natural Science Foundation of China (70271006, 60674107), and by the Natural Science Foundation of Hebei Province (F2006000571, F2006000343), and by the Natural Science Foundation of Education Committee of Hebei Province (20070607)

REFERENCES

[1] Wang Qingyin, et al. “Exploration of Research Method in Systems with Uncertainties”. U.S.A. *ASSA Lnauguration Issue*,1995,pp.136-138

[2] Zadeh L A. “Outline of a new approach to the analysis of complex systems and decision processes”, *IEEE Trans. Syst Man Cybern*, 1973,(3),pp.28-44

[3] Deng Julong, “ Introduction to Grey Mathematics”. *Wuhan: HUST Press*, 1992

[4] Wang Guanyuan, “Unascertained Information and Its Mathematical Treatment”. *Journal of Harbin Architecture engineering Institute*, 1990，23（4）

[5] Wang Guanyuan, et al. “A Preliminary Study on Unascertained Mathematics”, *Journal of Hebei Mining and Civil Engineering Institute (Higher Education Research Edition)*

[6] Wang Guanyuan, et al. “Unascertained Mathematics”. *Journal of Hebei Mining and Civil Engineering Institute*, 1992，（3）pp.86-130

[7] Wang Guanyuan and Zhang yue, “Fuzzy Random Process Theory”,.*Guiyang: Guizhou Science Press*, 1994

[8] Wang Guanyuan and Zhang yue, “Fuzzy Random Dynamic System”. *Beijing: Science Press*, 1993

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 012-014

Collaborative WWW-based CAI Software Achieved by Java

Sanjun Liu[1], Zhengjiang Yi[2], Weifeng Du[3*], and Zhengjiang Wu[4]
[1] Institute of Information Engineering, Jiaozuo University, Jiaozuo, Henan, China
Email: liusanjun1975@163.com
[2] Institute of Economic Management, Jiaozuo University, Jiaozuo, Henan, China
Email: yizhengjiang@163.com
[3] School of Mathematics & Information Engineering, Jiaxing University, Jiaxing, Zhejiang, China
Email: woodmud@tom.com
[4] School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, Henan, China
Email: jiang2021987@163.com

***Abstract*—The WWW application model based on the internet as a convenient and rapid means of sharing information is popular with users. The users view multimedia teaching contents by accessing hypertext pages, and execute free explore learning by clicking hyperlinks. However, the teaching based only on the hypertext pages has many shortcomings. The computer-assisted teaching software described by this article overcomes the many shortcomings. The software is achieved by J + + and consists of the server section, the students section, the teaching section by the teachers, the courseware maintenance section. This software actually is a teaching framework system, and all teaching materials are prepared by the teachers, and the teachers and students go into their by visiting their roles by accessing their own URL addresses. The students execute the learning, discussion, exercises and testing in the on-site control of the teachers.**

***Index Terms*—real-time; data-source data; frame**

I. FUNCTION OF THE SYSTEM

A. Learning of hypertext pages to learn

The teachers unified guide the learning pages of the students. The teachers choose the learning pages, send the URL address of the learning pages to the students, and guide the students to the Web pages to learn. This process increases the pertinence of learning, and meets the organized and controlled teaching needs.

B. Real-time discussion

The real-time discussion is an essential feature of collaborative teaching software. The system provides two kinds of tools to support real-time online discussion. One tool is the text conversation. If the speaker does not select the users, the conversation is public speak, otherwise the conversation is privately speak and is only sent to the selected users. There is a text field in each user's discussion panel to display the speaker and the corresponding speech of each speak. The other tool is the shared whiteboard. The teachers and students can write contents on the whiteboard, and the contents on the whiteboard of all users are consistent. The shared whiteboard supports lines, circles, rectangles, text, graphics and other basic operations, and can choose the color of lines.

C. Interactive exercises

The teachers send the guided message of practice, the students automatically enter the practice interface, and the system automatically transferred questions from the questions database for the students to do. In this process, the system compares the answers of the students with the standard answers to judge if the answers of the students is correct, and pops dialog boxes to give students the encourages information. The interactive exercises help students to farther understand the teaching contents.

D. Online Testing

The testing is a necessary teaching step in the process of teaching. The teachers can check the study of the students by the testing, and understand the learning results of the students to provide a basis for improving the teaching. The students can know their mastery of the contents to increase the interest in learning. Currently, each test paper is prepared by choosing questions from the test database and set the corresponding scores by the teachers. Tests are time-limited.

II. OVERALL DESIGN OF THE SYSTEM

To accomplish the functions from the system perspective, there are two ways of working: one ways is the client / server approach (such as students do questions), the client-side send an access application, the service-side gives the respondence.

The other ways is web-based collaboration applications (such as real-time discussion, etc.). In this way, any party could be a message sender, while the other party becomes the recipient of the message. The sender's action is caused by all kinds of event-driven of the user interface, and its timing is accidental. The receiver must always be in a listening state to receive.

Since Java's security restrictions, the socket connection between the two users' browsers can not be directly

* Corresponding author: Du Weifeng , School of Mathematics & Information Engineering, Jiaxing University, Jiaxing, Zhejiang, China, Email:woodmud@tom.com

AP-PROC-CS-10CN006

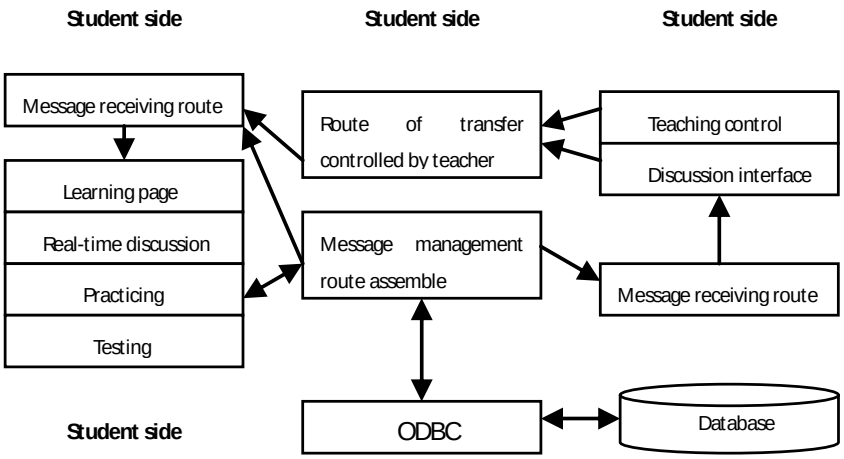


Figure 1. Overall design of the system.

established. The communication between the two users' browsers must be transferred by the Java application program in a Web server. To sum up, in order to achieve the function of the system, the overall design of the system is shown in Figure 1.

Instructions:

a) The students and teachers, respectively, establish communication links with the Java application program in the Web server by means of client/server. The connection monitoring route of the server side is responsible for monitoring the Socket connection application from the clients. The server side should establish a message management route for each linked client, and the message management route is responsible for receiving messages from the corresponding client, and analyzing and managing the messages.

b) The students and teachers, respectively have a message receiving route which is responsible for receiving message from the serve side. The message receiving route reflects the requirement of the received messages on the user interface by calling the method in the object of the user interface. Therefore, during the process of designing the message receiving route, it is a basic method to achieve the exchange visits between the objects in Java that the objects of the user interface are set as a parameter passed to the objects in the message receiving route.

c) The process of accessing the test database is executed as follows: the students send the corresponding message to the serve, the serve side accesses ODBC-based database via the COM component RDO to access, and feedbacks the accessing results to the students.

III. SEVERAL KEY TECHNOLOGY

A. Database Access

Visual J + + do a special extension for the Java class database according to the Microsoft operation system, and the users can access any software supporting COM in the Java programs. For accessing to the database, two methods of DAO(Date Access Object) and RDO (Remote Data Object) are provided in J + +. The software in this article adopts the RDO objects to realize the data access to the ODBC data source.

The security of IE should be checked strictly, and Java classes are divided into credible classes and discredible class. The discredible class can not use COM services. All classes not loaded from the classpath are not credible, the discredible classes includes the classes downloaded from the internet, unless the classes are packaged in a CAB file with a digital signature. In order to avoid the trouble raised by the security and reduce the ODBC configuration of the client side, the author adopts the method of visiting the database by the Java application program of the serve side to access the test database, and this method has no any security restrictions. The Applet program of the browser side passes accessw requests and results by means of communicating with the application program of the server side.

B. Message Design

The whole system is built based on the messages, the server side and client side both bear message management work. For a large number of messages based on collaborative applications, the serve side is only responsible for transferring such messages and such messages are managed by the client side. Each kind of messages is consisted of the message header and the content, and the message header is used for the system discrimination. The whole system mainly comprises six kinds of messages: chat messages; whiteboard messages; guide messages; users log in and log off messages; accessing to the test database messages, and other messages (such as boarding grade and checking grade).

C. Achieve of URL Guide

Inputting and transferring of URL The teachers click on "Page guide" button, the system accesses learning content index and pop up a dialog box so that the teachers can select the chapters and problems to study. Then the system sends the corresponding URL address to the student client side via the Socket path, and a new browser displaying the page is pop up for the confirmation.

Displaying of URL The steps are as follows:

a) using the URL class to form URL objects according the receiving character strings;

b) obtaining Applet Context of the small application program by means of get Applet Context of the Applet class;

c) displaying the content of URL by means of show Document of the Applet Context interface. The specific format and parameters of the method are described as follows:

Wherein, the target parameter is used to appoint the location of displaying of the file contents, and the specific value and the role of the target parameter are shown the Table I:

TABLE I.

Value	Location of file display
"_self"	Current frame
"_parent"	Parent frame
"_top"	Top frame
"_blank"	Displaying in a new open browser window
"Frame name"	Displaying in an appointed frame

In the teacher side, the author adopts the "-blank" method, which open another browser window to display the URL content for the teachers to confirm the content URL. In the student side, the author adopts the multi-frame method, which arranges the multi-frame in the HTML file, and one frame is used to display the Java Applet, the other frame displays the content of URL pages sent by the teachers. The size of the specified frame in the HTML file can be adjusted, so that the students can adjust the frame border to read the content more easily.

IV. CONCLUSION

This article describes a specific implementation of WWW-based collaborative teaching software system. The software actually is a teaching framework system, and the specific teaching contents are arranged and designed by the teachers by using the maintenance parts of the system. Thus, the subjective initiative of the teachers can be well exerted. The author intends to make the following improvements to the system: a) providing the function of automatic test paper; b) introducing a more comprehensive model of student to estimate the learning level of the students; c) providing a variety of teaching methods, so that the system can be operated in both circumstances of the attendance and absence of the teachers to increase the intelligence of the CAI.

ACKNOWLEDGMENT

This work was partially supported by Zhejiang province fatal project (priority subjects) key industrial project (2008C11011).

REFERENCES

[1] Translated by WANG Zhi-hai, etc., Publicated by InmonW H, Establishment of data warehouse [M]. Beijing: Mechanical Industry Press, 2000

[2] Liu Fei, Design and optimization of database connection pool based on Java [J]. Micro-computer applications .2008

[3] Peng Chen-Yang, Development guide of JAVA application system, Beijing: Mechanical Industry Press, 2004

[4] Publicated by Jamie J aworski (U.S.), translated by CAO Kang, Java Development guide [M]. Beijing: China Water Conservancy and Hydropower Press, 1996

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 015-018

Expression of XML Implicit Semantics

Ying Chen, Wenping Guo , and Xiaoming Zhao
School of Mathematics and Information Engineering, Taizhou University, Linhai, China
Email: ychen222@163.com , guo_wp@126.com , tzxyzxm@yahoo.com.cn

***Abstract*—The problem that XML formally governs syntax only but not semantics has been recognized as a serious barrier for XML-Based data integration and the extension of current Web to the semantic Web. This paper presents expression mode of XML implicit semantics, so that XML document authors may express the semantic information more clearly.**

***Index Terms*—XML, Semantics,XML Semantics ,Implicit Semantics, Semantic Web**

I. INTRODUCTION

Since its being the W3C(The World Wide Web Consortium) recommended standard of data denotation and exchange in the web, Extensible Markup Language(XML)[1]has achieved great success, not only it enjoys wide application in the industry, but also related techniques follow fast. However, one of its serious weaknesses is more and more obvious, say conveying syntax only without formally semantics [2]. Instead of being fully given expression, the syntax information contained in XML document is just kept in the human head. When developing the XML application program, researchers need to know implications of label and document structure in advance, and then encode them into the program directly [3]. Without any normative explanation to the label implication, XML users may directly encode their different understandings into the program which results in the impossibility of applications' mutual operation. As a result, it becomes a main difficulty in XML data integration and application integration.

It comes to be a research focus for web of next generation, the semantic Web [4], to study out how to deliver XML implicit semantics so as to give XML document users full play to express semantic information clearly.

II. SEMANTICS

Known as an ambiguous word, semantics contains different meanings if being introduced in different fields such as computer science, philosophy, psychics and linguistics. We need to analyse what it means in different situations to understand its semantic description and XML essence.

In Semantic Web, semantics is a key factor to develop our familiar internet [5]. Semantics can be simply regarded as an implication from its concept represented by the data (symbol), relations among the implications and abstract or higher-level logic expression to the data. In the field of computer science, semantics generally is a kind of explanation to computer symbol which are used by users to describe the real world. To be simple, it is a way for users to connect computer symbol with real world. To relational database as an example, simply speaking, data in the database is stored into formats. For example, students' basic information is stored into "Students' Format", then as to the ensemble from data of each column in the format, its implication is corresponding concept of the column, which reflects designers' given name to corresponding property of this column, such as NAME, GENDER etc. The relation among properties is equivalent to that among data's concept. Totally, all are the property of the entity of the students. Properties and relations in database format can be recognized as date semantic information.

Surely, semantics is absolutely not so simple; relation it represents may be more sophisticated, even more than expression range of modelling language of database such as E-R model. In fact, semantic is a new concept introduced to IT. Instead, it is long employed in the database to distinguish model and data, and creates a niche in database modelling, search and tasks management technique. In a word, semantics is a key element to ensure the expansibility, high effectiveness and robustness of data management system.

A. Expression Form of Semantic in Computer Science

The understanding of computer to the semantics of a symbol actually is a semantic expression that is understandable for a computer. That is to say, semantics embodies the description to the symbol.

As for most computer application nowadays, the information is not described in a uniform syntax format, which causing information disparity in the semantic description and semantic heterogeneity in the information expression. As to explanation to the semantics, meaning of the information, it is written into the application in a hard coding way, this so-called procedural semantics [6] isolates information semantics and its syntactic description and solidifies information understanding to certain code logic in the way of information semantic model, all which leading to its great limitation of without expansibility and generality and difficulty in maintenance and integration. The opposite as information modelling way of procedural semantics, declaration semantic modelling way has nothing to do with material data and application logic with nice adaptable and reusability by isolating data semantic description and certain application logic, fitting data semantic description to data layer, and leaving data semantic explanation to general formalization system (such inference system as first-order logic, description logic, datalog). Data semantic description is to describe data semantics in declaration

AP-PROC-CS-10CN006

semantic model way so as to alleviate the pressure of application processing date.

B. XML Semantics

It is known that XML semantics is semantic information in XML document.

As to information description, it is ideal to describe all information with same syntax. As an effective way to share the data among applications or machines, XML makes it possible to put data with different types into same format. While, the understanding to XML semantics is different in different fields, among which, three such understandings have solved different problems of XML.

1) Mapping to UML: Mapping XML document to XML diagram as its semantics. MDL (Meaning Definition Language) [8] is such a kind of language, which chooses XML nodal points with Xpath expression words and maps the points to UML object, attribute and relation. While, XML data binding is a technique to map XML document OO of object model.

2)Basing on format semantic way of procedure language: XML semantics can be defined format semantic way of procedure language, for example, W3C[9] standard defines Xpath and Xquery operation semantics with judge and inference rule; Wadler [10] describes model semantic of Xpath and XSLT with traditional referring semantics; Yang[l1]describes Xpath semantics with formalizing standard language of Object-Z.

3)Basing on attribute grammar: Finding the similarity between XML marked document and character string generated from BNF grammar, Psaila & Crespi-Reghizzi[l2] put forward a way to transfer DTD element description to formal EBNF rule and Semantics rule definition (SRD) and gives an algorithm to calculate XML relation semantics of the rule definition. After that, Havasi[13] further advanced a new semantic rule congregation and expression language of semantic rule meta-language (SRML), which extend XML. Attribute grammar checks semantic consistency by examining and evaluating attribute value.

III.XML IMPLICIT SEMANTICS

Though having been the base of web technique framework, with its application in data exchange and application congregation, XML is more and more known for its important shortage: expression only in data grammar, not in formalizing semantics.

Basic XML idea: data in XML document is described by labels in a meaningful and self-description way, especially label names selected by experts [14]. The label reflects common recognition. For example, label of <age> means the age which can be inferred that data marked on the label refers to age. However, XML including DTD or XML Schema does not give formalizing system to explain what the label mean which results in XML processor's failure to understand what label mean in XML document. In fact, there is no difference in the meaning between label <age> and HTML label <HI>. That means XML shows grammar rather than formalizing semantics.

While, XML label and XML document framework contain some meaning, the label means something to human beings, just like XML fragment:

```
<Lecturer name="David Kevin">
<course>
      <name>Computer Network</name>
      <type>Compulsor </type>
   </course>
</lecturer>
```

expresses rich semantic information as follows: a lecturer named "David Kevin" teaches a compulsory course of "computer network". The nest between nodal points of lecturer and course shows a relation of professor. Semantic information in XML document is called as XML semantics [2]. XML semantics is expressed by document framework and label such as element nest and nest adjacency, and XML attribute. It is worthy to mention, there is no corresponding rules to bind XML 's semantic expression way, so XML implicit semantic expression ways are diverse with irregular formats, such as nest and adjacency of node, and detailed meaning of node order. So it is difficulty to automatically analyse and extract relevant semantic information from XML document.

IV.XML IMPLICIT SEMANTIC EXPRESSION MODEL

In fact, XML is designed to be a unified data format and not suitable to show concept modal, which including concept, attribute and their relation. However, when users express data to XML document, data semantic information has been in label and framework of XML document such as element nest and adjacency, XML attribute nodal order and ID / IDREF. But it is impossible for the machine to understand semantic information; furthermore, most semantic information has been lost.

However, XML has visual semantics; there is some meaning in the label which means something to human beings. In addition, XML document framework also mean something and XML semantics is impliedly expressed in XML document.

XML has rich semantic information and diversified semantic expression ways, generally including word, individual, and data attribute and object attribute of OWL DL.

A. Expressions of words

In XML, word is generally showed by the value of text node and XML attribute. If data pattern defined in XML Schema, it corresponds to OWL pattern word. But element sometimes can be expressed into character, say as data attribute value, while these elements owns no attributive node. For example, an XML document to show students' scores.

```
<History>
<student name="Wang Yinghua" score="88" No.="0708101032"/>
<student name="Zou Yuting" score="83" No.="0708101017"/>
```

</History>

If we define element of student as an object, it not only has such attributes as name, score and student number, also contains attributes of course name, and name of element history is the value of course name attribute. In this situation, this element shows a king of character, its equivalent XML can be expressed as:

<student name="Wang Yinghua" score="88" No.="0708101032" course=" History "/>

<student name=" Zou Yuting " score="83" No.="0708101017" course=" History "/>

B. Expressions of Conditional Semantic

XML semantic expression is probable to have conditions which means if only satisfying certain condition, this element can express an example in some class or some attribute. The condition may be diverse, may be about attribute value or element character value, or about document framework, for example, some sub element is a must. For example:

<college coltype="sub-college" name="Foreign Language"/>

Among which, attribute of coltype is used to choose subclass of the college, object expressed by element college is sub-class practical example decided by the value of attribute of coltype.

XML data attribute expression is probable to be conditional. For example,

<cigarette>

<cigaretteProperty propertyName="Nanking" propertyValue="180"/>

<cigaretteProperty propertyName="Mild Seven" propertyValue="7"/>

</cigarette >

Among which, data attribute in element of Cigarette Property depends on the value of property Name node.

Similarly, the expression of object attribute may be conditional.

C.Expressions of Others

In XML, document order contains important meaning; for example, in one language with XML to show structural data, nodes in the same name has different meaning in different positions.

For example, a XML fragment expressing individual normal form [18]

<Address>

<string>China</string>

<string>Zhou Xiaqing</string>

<string>605 East Road </string>

</Address>

Among which, the first string fragment expresses attribute of state, and the second attribute of name. Besides, XML element may be divided into groups according to the attribute. Like:

<MarketOn date="20090122”>

<truck code="Kingkong0001023”price="47600" units="7"/>

<car code="Panda002046" price="37500" units="17"/>

</MarketOn>

<MarketOn date="20090122”>

<truck code="Kingkong0011072”price="37600" units="23"/>

<car code="Panda0001943" price="57500" units="9"/>

</MarketOn>

V. CONCLUSIONS

This paper introduces the concept of semantics, then analyzes semantic expression way in the computer science and shows XML semantics is implicit, failure to express formalizing semantics. Finally, a model to express XML implicit semantics is put forward.

Because the aim of XML Schema and DTD is to constraint XML document framework, and label and structure in XML document are not always just corresponding to the object and attribute. In addition, same XML structure may have many semantic explanations, such as nest structure, while XML Schema and DTD cannot distinct the possibilities clearly. As a result, to get the semantics automatically from XML document, XML Schema designers must be supposed to follow the thought of knowledge expression, designing XML document format in a normative way.

Nevertheless, due to the diversity of XML implicit semantic expression way, this suppose seldom meets, which makes it very difficulty to get semantics automatically from XML document, so the further study in the future is necessary.

REFERENCES

[1] Tim Bray, Jean Paoli, C. M. Sperberg-Mc(aueen, and Eve Maler. Extensible markup language (xml) 1.0 (second edition) w3c recommendation, 6 October 2000

[2] Liu SP, Lin ZQ, Mei J, Yue AB. A model-theoretic semantics for XML. Journal of Software, 2006,17(5):1089-1097..

[3] Uschold M. Where are the semantics in the semantic Web. AI Magazine, 2003,24(3):25-36.

[4] Berners-Lee T, Handler J, Lassila O. The semantic Web. Scientific American, 2001,184(5):34-43.

[5] Amit Sheth, Robert Meersman. Amicalola Report: Database and Information Systems research Challenges and Opportunities in Semantic Web and Enterprises. SIGMOD Record Special Issue on Semantic Web, Database Management and Information Systems, December 2002.

[6] Stefan Decker, Sergey Melnik, Frank Van Harmelen, Dieter Fensel, Michel Klein, Jeen Broekstra, Michael Erdmann, and Ian Horrocks. The semantic web: The roles of xml and rdf. IEEE Expert, 15(3), October 2000.

[7] Michael Uschold. Where are the semantics in the semantic web? AI Magazine,24(3):25-36, 2003.

[8] Robert Worden. Mdl: a meaning definition language, version 2.06, 2002.

[9] Denise Draper, Peter Fankhauser, Mary Fernrdez, Ashok Malhotra, Kristoffer Rose, Michael Rys, Jerome Simeon, and Philip Wadler. Xquery 1.0 and xpath 2.0 formal semantics, w3c working draft 20 February 2004.

[10] Philip Wadler. A formal semantics of patterns in xslt and xpath. Markup Languages,2(2):183-202, 2000

[11] Hong Li Yang, Jin Song Dong, Ke Gang Hao and Jun Gang Han., semantics of xslt using object-z. In APWeb 2003 Formalizing pages 120-131, 2003
[12] G. Psaila and S.Crespi-Reghizzi. Adding semantics to xml. In Second Workshop on Attribute Grammars and their Application, WAGA'99, pages 113-132, Amsterdam, The Netherlands, 1999
[13] Ferenc Havasi. Xml semantics extension. Acta cybernetica, 15(4):509-528, 2002
[14] Cover R.XML and semantic transparency.1998
[15] Grigoris Antonious, Frank van Harmelen. A Semantic Web Primer. MIT Press,2004
[16] Liu SP. A Model-Theoretic Semantics for XML and Its Applications.2005
[17] Horrocks I, Patel-Schneider PF. Reducing OWL entailment to description logic satisfiability. Journal of Web Semantics, 2004,1(4):345-357.
[18] Henry S. Thompson. Normal form conventions for xml representations of structured data. 2001.

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 019-022

An Application Research on Video Coding H.264 Standard and Video Conference System

Zhan Guo[1], Zuming Xiao[1], Feng Luo[1], and Bin Tan[2]
[1] Dept. of Physics, Jingdezhen Comprehensive College,Jingdezhen Jiangxi ,China
e-mail: teachergz@163.com
[2]School of Information Science and Communication, Jinggangshan University,Ji-An Jiangxi ,China
e-mail: jatanbin@163.com

Abstract—**This paper introduces an application research on video coding H.264 standard and video conference system , given the basic structure of the system design, and video processing module of the system has been optimized, tested and compared to verify the system has the good real-time and stability.**

***Index Terms*—video compression standard, H.264, video conferencing systems, video encoding and decoding technology, network architecture, video, images, multimedia communications**

I. INTRODUCTION

With the development of information technology and the popularity of network, the need for supplying high quality video/audio multimedia services grows rapidly. Videoconference is a multimedia communication tech that can achieve the real-time 、 visual and interactive communication with people who are in deferent places. Apply of Videoconference tech can save customer's time and improve the work efficiency etc[1]. Videoconference tech has a wide applicable scope, and it has a good prospect. H.264 applies many mostly used technologies with the advantage of former video compresssion standard. It provides fairly good video quality while achieving better coding efficiency and more simple syntax expression, and it's the most efficient video compression standard at present.This article introduces an application research on video coding H.264 standard and video conference system.Relativing hardware video conferencing system,it invests few on the hardware equipment, maintains the quantity to be small, thus becomes low cost high product plan.the System to network adaptiveness very well,the system may penetrate the firewall, attends the conference the flexibility to be very well, the system's mobility is strong as well. To enhance the real-time nature of video conferencing system to use the H.264 standard for video processing module optimization.video conferencing systems and providing basic structural design, as well as a brief description of system modules.

II. SUMMARY OF SYSTEM

The video conferencing system has five main functional modules[2]: Video processing module、Audio Processing Module 、 Network TCP/IP Transmission Module、Network Quality Control Module、Meeting Manager module, Shown in Figure 1.

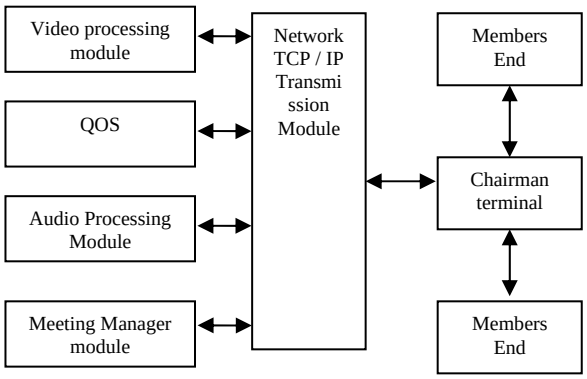


Figure 1. Software Video Conference System diagram

A. Video Processing Module

Video processing modules include video data collection, encoding, decoding and playback. Playback of video data collection and use of standards in line with Microsoft's software development kit VFW (Video For Windows), mainly by calling AVICap32.dll created AVICAP window class, window class by AVICAP in the news, macro functions, structure and the callback function to achieved. For the H.264 video codec, which has more than one bright spot in technology, such as the unified symbol VLC coding, high-precision, multi-mode displacement estimates, based on 4x4 Integer Transform block, intra prediction, deblocking filter, support for SI and SP frames, multi-frame reference, layered coding syntax and so on, these measures have made H. 264 algorithm has a high coding efficiency, in the same reconstructed image quality of the next. Than H.263 bit-rate savings of about 50%; and the H.264 bit stream structure of the network adaptability, increased error resilience can be well adapted to IP and wireless networking applications; application target range is wider, meet the different rates, different resolutions and different transmission (storage) occasion demands. The basic system is open, use without copyright, which will reduce the system cost. However, the computational complexity is too high, the complexity of the encoder for the H.263 ten times, its H.263 decoder complexity is 3 times. H.264 codec to reduce the complexity of this system has become the key to whether the real-time operation[4].

To this end, pairs of H.264 can be considered in the following aspects of the corresponding optimization [6]: The most common approach is to optimize the

compression ratio (algorithm complexity) and the compression efficiency (real time) between the trade-offs. First, we can shield the most time-consuming some of the features of H.264, such as B frames, the encoder with only the I frames and P-frame coding, did not use B frames, SI-frame, SP frame. Entropy coding using only general-purpose variable-length coding (UVLC), and the codec in its process, by adjusting the code table look-up table used in the binary tree algorithm, etc., further reducing the computational complexity. Second, can the motion estimation, multiple reference frames and the adaptive time-consuming sub-block size of the larger parts of the algorithm optimization. In intra prediction, using the neighboring blocks have been intra-coded coding mode, some pixels of the Cost values and valve limits, a reasonable choice of the forecasting models the current block in advance out of the forecasting process[5].

In the inter-frame encoding mode selection in the use of a rapid inter-frame encoding mode selection algorithm. The search process in motion by adding rate-distortion optimization system, a balanced consideration of the quality of coding and computational complexity, the use of processed blocks of the adjacent projections information, a reasonable estimate of the current block is most likely to use forecasts frame, block shape and predicted position, time out of forecasting process. In the integer search using the improved asymmetric cross-shaped multi-layered hexagonal search algorithm (UMHexagonS), this algorithm is compared with the original H.264 fast full-search algorithm used in the calculation of the amount of the savings of more than 90%, especially in the Large movements of high bit-rate image sequence coding, while maintaining a good rate-distortion performance under conditions of very low computational complexity; in 1 / 2 pixel, 1 / 4 pixel search using fast algorithm for fractional pixel motion estimation, greatly reducing the the computational complexity and improve the search accuracy.

In the deblocking filter, taking into account the final results of video encoding and decoding receiver object is the human eye, the human eye to color brightness sensitivity is far greater than the sensitivity; and under the principle of deblocking filter can be aware of , when the intensity of the more hours, blocking the more obvious, and therefore intensity of less than 3, not do chroma filtering Although this would be detrimental to image quality, but the experiments prove that damage is not obvious, Deblocking Filter amount of computation is reduced by 30% or so.

Audio processing module, including audio data collection, encoding, decoding and playback. In this system, using the Windows Multimedia Audio of the wave function is set and ACM (Audio Compression Manager) to achieve the audio data collection, coding and decoding, playback using DirectX components of the process is one of the DirectSound audio codec technology adoption the G.723, G.711 and G.729 technology for the user to choose.

B. Network TCP / IP Transmission Module

The system uses a distributed multipoint conferencing mode, rather than using a centralized multipoint conferencing mode. The main reason is that centralized multipoint conferencing method must be used MCU. While the MCU is often very expensive, not suitable for ordinary users. Distributed Multipoint mode using multicast technology, which can effectively reduce the burden on the network to avoid unnecessary waste of resources.

Network Transmission Systems transport layer is divided into modules and systems network layer transport layer protocols include TCP and UDP protocols both too, UDP protocol Fortunately, real-time performance in the LAN. However, in WAN. As the ever-changing network environment, will inevitably produce delay, jitter and other issues, it will seriously affect the amount of audio and video effect.

To this end, the system in the above agreement on the basis of the two plus the RTP / RTCP protocol. In order to meet the real-time audio and video, audio and video information is transmitted using UDP protocol; and apply to enter the meeting, the President permission to get in and out of meetings and other basic control information using TCP protocol for reliable transmission; data reconciliation packages this package modules to complete. System, the transport layer from the Windows platform, complete the WinSock specification, the system object by all of the communication processes in the Microsoft c + + class derived CAsyneSoeket made. These objects include a video communication process, the audio communication process, to receive quality information feedback communication process, the meetings management process and other communications in the network layer is to comply with TCP / IP in the IGMP protocol, the system uses the standard Windows Socket2.0-specific multicast preparation of multicast communication module API.

The aforementioned system, several communication process. Where are all based on UDP multicast communication process.

C. Network Quality Control Module

Quality control module in the network to fully integrate the H.264 codec features and audio codec features are basically solved the video, audio, due to the phenomenon of discontinuous bit stream structure of H.264 network adaptability, using a "network-friendly" the structure and syntax, and have error recovery capability, which would facilitate the handling of error and packet loss, but also makes the network quality control than H.263, MPEG-4 and other video conferencing systems better. First, by a source within the RTP header description, serial number, time stamp and other information, and the number of RTCP provides contracting, contract time, the number of received packets, received the same package network-related information, to achieve the right to determine the status of network transmission. In determining the network conditions used in two ways[6]:

(1) compared the amount of data sent each other and their own side to receive the amount of data to calculate the loss rate, determine the network data transmission situation;

(2) the packet round-trip time to judge.

Secondly, H.264 encoder based on network of independently adjustable quantization step size to adapt to channel coding. In the H 264 and also made use of data partitioning method to deal with the channel bit rate changes. Generated in the encoder has a different priority of video data to support the network quality of service

QoS, each frame data is divided into several parts according to their importance in the network congestion, buffer overflow will be discarded less important information; encoder will be the status of the network independently of the frame for some important non-coding, non-transmission.

Taking into account the sensitivity of people to the sound signal is higher than the video signals and audio signals in the video conference is more important than the video signal. When the two data streams to compete on the same network resources, give priority to the amount of audio data in order to reduce the pressure on the network, in the audio encoder detected an increase of silence, only speaking participants to send the audio data; in order to ensure the continuity of audio data, In the audio decoder side opened a larger buffer, when the data received enough for a long time. Through the data packet to determine whether the timestamp on the data for the same time, evenly played.

D. Conference Management Module

Conference management support throughout the meeting process, not only to manage and coordinate all the resources used in meetings, including the transmission path, communication protocols, codecs, media access, display and related information, but also manage the meeting process, including initiation, maintenance, from the will and voice control. Conference Manager has six components: membership information database, the registered membership, speaking handling, process management, information query, from the Club Management. The management of the entire conference by the President ended the meeting manager and members of the coordination of client meetings manager to resolve. This system allows the Conference Manager can be dynamically expanded scale of the meeting, members are free to join, from the council[7].

III. SYSTEM TESTING

As a result of this system is the full software program, simply P4 level of PC machines can be achieved, which will greatly reduce the user's costs for the general ordinary users. H.264 codec software through the machine to complete. As the complexity of H.264 is the H.263 coding ten times. Coding is the key to the whole system, of which the video encoding rate, encoding rate and decoding speed determines the overall system performance[8].

Desktop video conferencing systems throughout the test platform is P4/2.4GHz 512M RAM of the PC machines and Win2000 professional system, test software is used in the official test H.264 source code, by the German Institute for hhi responsible for the development of the JM8.6[3] generated by VC7.1 compile release version, JM8.6 using basic subset (baseline profile). In the parameter setting off some of the H.264 encoding performance, the contribution of the small but very high degree of computational complexity of the new features, such as multiple reference frames, unnecessary inter prediction block mode, CABAC and so on.

When the video resolution of 176x144, frame rate of 15 frames per second, audio to G.723 (5.3 k) mode conditions, all the way video takes about close to 150kHz bandwidth. Way audio to account for less than 6kHz bandwidth, the bandwidth occupied primarily video. System supports a maximum resolution of images is 352X288, video frame rate of 30fps, the maximum rate of 2Mbps.

TABLE I.
IPERFORMANCE OF VIDEO ENCODING SYSTEM

Test Sequence	Coding time	PSNR（y）	PSNR（u）	PSNR（v）	Actual Rate
foreman	4.446S	29.86dB	37.37dB	37.94dB	64.3 Kbps
Mother_D	3.238S	35.32dB	39.86dB	40.12dB	64.2 Kbps

Table l lists the PC to complete the video encoding with pure efficiency. All the sequence 150, target rate of 64Kbps. The source frame rate to 15fps, target frame rate of l5fps.

In JM8.6 default parameters, using the basic subset (baseline profile) and use the same test sequence. Measured performance data listed in Table 2.

TABLE II.
H.264 VIDEO CODING PERFORMANCE（JM8.6 BASELINE PROFILE）

Test Sequence	Coding time	PSNR（y）	PSNR（u）	PSNR（v）	Actual Rate
Foreman	347.81S	35.79dB	39.14dB	40.51dB	134.53 Kbps
Mother_D	311.63S	37.37dB	41.58dB	41.93dB	126.28 Kbps

IV. ANALYSIS

Data Analysis: From the comparison of before and after optimization, we can see that the PSNR drop 1-5dB circumstances, the subjective image decline in the quality is not significant. But comparing the before and after the encoding rate, is indeed a huge increase, and this in the video conferencing system to improve the system for real-time is very critical[9].

(1)a software-based H.264 video conferencing system uses the H.264 standard, can be achieved at low bit rates higher quality image transmission, image quality in the same circumstances, the coding standard than the traditional savings about 50% of the rate.

(2)the reasonable choice for some of the complexity of the H.264 high performance, low-impact on the encoding characteristics of the optimization, so that the system has good real-time performance and stability.

At the same time, P4 class processor support MMx, SSE, SSE2 instructions, etc., and these instructions for achieving parallel processing to speed up computation speed played an important role. Therefore, the full use of processor resources, and the use of water mechanism, rational design and scheduling the various modules of the encoder, the H.264 codec to achieve the overall speed, but also can improve the systems real-time effects[10].

REFERENCES

[1] H. 264/ MPEG-4 Part 10 White Paper-Transform & Quantization. Available at www .vcodex .com

[2] Halbach Till, Wien Mathias. Concepts and Performance of Next-generation Video Compression Stan dardization. In Proc. Nordic Signal Processing Symposium on board Hurtigruten, Norway, Oct. 2002.

[3] Joint Video Team(JVT), reference software JM 86, http://iphome.hhi.de/suehring/tml/download/old_jm/jm86.zip

[4] Wiegand Thomas, Sullivan Gary J, et a1. Overview of the H. 264／AVC Video Coding Standard, IEEE Trans. on Circuits and Systems for video technology, July 2003.

[5] Schafer RaIf, Wiegan Thomas Schwarz Heiko. The Emerging H. 264C standard. EBU Technical Review. Janary 2003.

[6] Meng Bojun, et a1. Efficient intra-prediction algorithm in H. 264. Image Processing, Sept. 2003.

[7] Doulamis, A Doulamis, N Ntalianis, K Kollias, S An efficient fully unsupervised video object segmentation scheme using an adaptive neural-network classifier architecture. Neural Networks, IEEE Transactions on , Volume.14 ,NO.3, May 2003:616~630

[8] ISO/IEC International Standard 11172-2.Coding of Moving Picture and Associated Audio Information: Video. Nov.1994

[9] ITU-T. Draft ITU-T Recommendation H.263 (Video Coding for Low Bit rate Communication).Jan.1995

[10] Dogan S., Sadka A.H., Kondoz A.M. Efficient MPEG-4/H/263 video transcoder for interoperability of heterogeneous multimedia networks. Electronics Letters, 1999.35(11):863～864

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 023-026

Application of Neural Networks in Medical Image Processing

Zhenghao Shi[1], and Lifeng He[2]
[1] School of Computer Science and Engineering, Xi'an University of Technology, Xi'an, China
ylshi@xaut.edu.cn
[2] School of Information Science and Technology, Aichi Prefectural University, Nagoya, Japan
helifeng@ist.aichi-pu.ac.jp

***Abstract*—This paper reviews the application of artificial neural networks in medical image preprocessing, in medical image object detection and recognition. Main advantages and drawbacks of artificial neural networks were discussed. By this survey, the paper try to answer what the major strengths and weakness of applying neural networks for medical image processing would be.**

***Index Terms*—Neural Network; Medical Image preprocessing; Object Detection;Computer Aided Diagnosis**

I. Introduction

In the past years, artificial neural networks (ANNs) have seen an an increasingly interests in medical image processing[1]-[2]. According to our searching results with Google Scholar, more than 33000 items were found on the topic of medical image processing with ANNs during the past 16 years. The intention of this article is to cover those approaches introduced and to make a map for ANN techniques used for medical image processing. Instead of trying to cover all the issues and research aspects of ANNs in medical image processing, we focus our discussion on three major topics: medical image preprocessing, medical image segmentation, and medical image object detection and recognition. We do not contemplate to go into details of particular algorithms or describe results of comparative experiments, rather we want to summarize main approaches and point out interesting parts of the neural networks for medical image processing, further more by this survey, we try to answer what the major strengths and weaknesses of applying ANNs for medical image processing would be.

II. Applications of Neural Networks in Medical Image Processing

A. Preprocessing

Image preprocessing with neural networks generally falls into one of the following two categories: image reconstruction and image restoration. The Hopfield neural network is one of the most used neural works for image reconstruction [3]-[7]. Of our reviewed literatures related to this areas, Hopfield neural network based methods pose 55 percent. The major advantage of using Hopfield neural network for medical image reconstruction is that the problem of medical image reconstruction can be taken as an optimization problem, which is easily solved by letting the network converge to a stable state while minimizing the energy function.

Reconstruction of images in electrical impedance tomography requires the solution of a nonlinear inverse on noisy data. This problem is typically ill-conditioned and requires either simplifying assumptions or regularization based on a priori knowledge. The feed forward neural network [8]-[9] and the self-organizing Kohonen neural network [10]-[11] , which pose 2 of 9 papers among our reviewed literatures, respectively, seem to have more advantage for such medical image reconstruction compared with other techniques, they can calculate a linear approximation of the inverse problem directly from finite element simulations of the forward problem.

The majority of applications of neural networks in medical image preprocessing are found in medical image restoration, 13 of 24 papers among our reviewed literatures focused their interests here [12]-[21]. Among which, one paper for Hopfield neural network, seven papers for the feed forward neural network, and two papers for fuzzy neural network and for cellular neural network, respectively. In the most basic medical image restoration approach, noise is removed from an image by filtering. Suzuki et al. developed neural network based filters (NFs) for this problem [12]-[14]. Suzuki et al. also proposed a new neural edge enhancer (NEE) based on a modified multilayer neural network, for enhancing the desired edges clearly from noisy images [15]. The NEE is a supervised edge enhancer: Through training with a set of input noisy images and teaching edges, the NEE acquires the function of a desired edge enhancer. Compared with conventional edge enhancers, the NEE was robust against noise, was able to enhance continuous edges from noisy images, and was superior to the conventional edge enhancers in similarity to the desired edges.

B. Image segmentation

The feed forward neural network is the most used neural network for medical image segmentation. Among our reviewed papers, 6 of 17 papers employed the feed forward network for medical image segmentation [22]-[27]. Compared with the traditional Maximum Likelihood Classifier (MLC) based image segmentation method, it has been observed that the feed forward neural networks-based segmented images appear less noisy, and the feed forward neural networks classifier is also less sensitive to

the selection of the training sets than the MLC. However, most feed forward neural network based methods have a very slow convergence rate and require a priori learning parameters. These drawbacks limited the application of feed forward neural networks in medical image segmentation.

Hopfield neural networks were introduced as a tool for finding satisfactory solutions to complex optimization problems. This makes them an interesting alternative to traditional optimization algorithms for medical image reconstruction which can be formulated as optimization problem. Among our reviewed literatures, 4 of 17 paper used Hopfield neural network to segment some organs from a medical image [28]-[31].

C. Object detection and recognition

For using neural networks for medical image detection and recognition, the back propagation neural network poses most places, 11 of 23 papers among our reviewed literatures employed it [37]-[47]. Compared with conventional image recognition methods, no matter used for the interpretation of mammograms [37], or used for cold lesion detection in SPECT image [38], or used for diagnosing classes of liver diseases based on ultrasonographic [39], or used for separation of melanoma from three benign categories of tumors [41], or distinguish interstitial lung diseases [42], or used for reduction of false positives in computerized detection of lung nodules in LDCT [43]-[46]and chest radiography [47], all these feed forward neural network based methods show their preference in recognition accuracy and computing time compared with conventional methods.

Other neural networks, i.e. Hopfield neural network [48], ART neural network [49], radial basis function neural network [50], Probabilistic Neural Network [51], convolution neural network [53]-[56], and fuzzy neural network [52] [57], have also found their position in medical image detection and recognition, which poses 1 of 23, 1 of 23, 1 of 23, 1 of 23, 2 of 23 and 2 of 23 papers, respectively.

Different from what mentioned above, in [58] and [59], artificial neural network ensembles are employed for cancer detection. The ensemble is built on two-level ensemble architecture. The first-level ensemble is used to judge whether a cell is normal with high confidence where each individual network has only two outputs respectively *normal cell* or *cancer cell*. The predictions of those individual networks are combined by some a method. The second-level ensemble is used to deal with the cells that are judged as cancer cells by the first-level ensemble, where each individual network has several outputs respectively, each of which represents a different type of lung cancer cells. The predictions of those individual networks are combined by a prevailing method, i.e. *plurality voting*. Experiments show that the neural network ensemble can achieve not only a high rate of overall identification but also a low rate of false negative identification, i.e. a low rate of judging cancer cells to be normal ones, which is important in saving lives due to reducing missing diagnoses of cancer patients.

III. DISCUSSION

From the reviewed literatures, we find that no matter what neural network model employed for medical image processing, compared with conventional image processing methods, the time for applying a trained neural network to solve a medical image processing problem was negligibly small, though the training of a neural network is a time cost work and also medical image processing tasks often require quite complex computation [12]-[15]. We think that this may be the major contribution of using neural network for solving medical image processing tasks.

Despite their success story in medical image processing, artificial neural networks have several major disadvantages compared to other techniques.

The first one is that a neural network is hard to express human expert's knowledge and experience, and the construction of its topological structure lacks of theoretical methods [60]. Moreover the physical meaning of its joint weight is not clear. All these can make the image processing method by neural networks unstable. A solution to these problems may be to combines fuzzy technique with neural networks together by using neural networks to process fuzzy information. It provides neural networks the ability to express qualitative knowledge, and network topological structure and joint weight have clear physical meaning. Also, it can make the initialization of network easier, avoid the local optimization of network training, and ensure the stability of networks [61].

The second problem relates to the amount of input data. For achieving a high and reliable performance for non-training cases, a large number of training cases are commonly required [62] [63]. If an ANN is trained with only a small number of cases, the generalization ability (performance for non-training cases) will be lower (e.g., the ANN may fit only the training cases). Because medical images are progressing rapidly as technology advances, the timely development of CAD schemes is important. However, it is very difficult to collect a large number of abnormal cases for training, particularly for a CAD scheme with a new modality, such as lung cancer screening with multi-detector-row CT. This significantly degraded the results obtained.

REFERENCES

[1] Kunio Doi, "Computer-Aided Diagnosis in Medical Imaging: Historical Review, Current Status and Future Potential," *Computerized Medical Imaging and Graphics.* vol.31,no.4-5, pp.198-211, July 2007.

[2] Miller A.S, Blott B.H, Hames T.K. ,"Review of Neural Network Applications in Medical Imaging and Signal Processing," *Medical and Biological Engineering and Computing.* vol.30, no.5,pp.449-464, September 1992.

[3] A. Cichocki, R. Unbehauen, M. Lendl, K. Weinzierl, "Neural Networks for Linear Inverse Problems with Incomplete Data Especially in Applications to Signal and Image Reconstruction," *Neurocomputing.*vol.8, no.1, pp.7-41, May 1995.

[4] W. Warsito, L.S. Fan,"Neural Network Based Multi-Criterion Optimization Image Reconstruction Technique

for Imaging Two and Three Phase Flow Systems Using Electrical Capacitance Tomography," *Chemical Engineering and Processing.* Vol.42, no.8-9, pp.2198-2210, April 2001.

[5] B. Su, Y. Zhang, L. Peng, D. Yao, B. Zhang,"the Use of Simultaneous iterative reconstruction technique for electrical capacitance tomography," *Chemical Engineering Journal*. Vol.77, no.1-2, pp.37-41, April 2000.

[6] Y. Wang, F.M. Wahl,"Vector-Entropy Optimization-Based Neural-Network Approach to Image Reconstruction from Projections," *IEEE Trans. Neural Networks.* Vol.8, no.5, pp.1008-1014, September 1997.

[7] Y. Wang, P. Heng, F.M. Wahl,"Image Reconstructions from Two Orthogonal Projections," *International Journal of Imaging Systems and Technology*.vol.13, no.2, pp.141-145, Aug. 2003.

[8] F. Ali, Z. Nakao, Y.-W. Chen, K. Matsuo, I. Ohkawa,"An Adaptive Backpropagation Algorithm for Limited-Angle CT Image Reconstruction," *IEICE Trans. Fundamentals.* Vol.E83-A, no.6, pp.1049-1058, 2000.

[9] A. Netajatali, I.R. Ciric,"An Iterative Algorithm for Electrical Impedance Imaging Using Neural Networks," *IEEE Trans. Magn.* Vol.34, no.5,pp.2940-2943, September 1998.

[10] C. Comtat, C. Morel,"Approximate Reconstruction of PET Data with a Self-Organizing Neural Network," *IEEE Trans. Neural Networks.* Vol.6, no.3,pp.783-789, May 1995.

[11] A.Adler, R. Guardo. ,"A Neural Network Image Reconstruction Technique for Electrical Impedance Tomography," *IEEE Transactions on Medical Imaging.* Vol.13, no.4, pp.594-600, 1994.

[12] Kenji Suzuki, Isao Horiba, Noboru Sugie and Michio Nanki,"Neural Filter with Selection of Input Features and Its Application to Image Quanlity Improvement of Medical Image Sequences," *IEICE Transaction on Information and Systems*. Vol.E85, no.10, pp.1710-1718, 2002.

[13] K. Suzuki, I. Horiba, and N. Sugie,"A Simple Neural Network Pruning Algorithm with Application to Filter Synthesis," *Neural Process. Lett.* Vol.11, no.1, pp.45-53, Feburuary 2001.

[14] K. Suzuki et al,"Efficient Approximation of Neural Filters for Removing Quantum Noise from Images," *IEEE Trans. Signal Process*.vol.50,no.7, pp.1787-1799, July 2002.

[15] K. Suzuki et al,"Neural Edge Enhancer for Supervised Edge Enhancement from Noisy Images," *IEEE Trans. Pattern Anal. Mach. Intell.* Vol.25, no.12, pp.1582-1596, December 2003.

[16] Chuan Yu Chang,"Contextual-Based Hopfield Neural Network for Medical Image Edge Detection," *Optical Engineering.* Vol.45, pp.37006-37015, March 2006.

[17] Clarke, L.P., Wei Qian,"Fuzzy Logic Adaptive Neural Networks for Nuclear Medicine Image Restorations," *Proceedings of the 20th Annual International Conference of the IEEE.* Vol.3, pp.1363-1366, Juaury 1998.

[18] Qian W., Li H., Kallergi M., Song D., Clarke L.P,"Adaptive Neural Network for Nuclear Medicine Image Restoration," *The Journal of VLSI Signal Processing.* Vol.18,no.3, pp.297-315, April 1998.

[19] N. Aizenberg, I. Aizenberg, T.P. Belikova,"Extraction and Localization of Important Features in Gray-Scale Images: Implementation with CNNs," *Proceedings of the IEEE International Workshop on Cellular Neural Networks and their Applications CNNA, Rome, Italy, IEEECS Press, Silver Springs*,pp.207-212, December 1994.

[20] I. Aizenberg,"Processing of Noisy and Small-Detailed Gray-Scale Images Using Cellular Neural Networks," *Journal of Electronic Imaging*.vol.6, no.3, pp.272-285, July 1997.

[21] I. Aizenberg,"Multi-Valued Non-linear Filters and Their Implementation on Cellular Neural Networks," *Frontiers in Artificial Intelligence and Applications*. Vol.41,pp.135-140, 1997.

[22] Du Yih TSAI,"Automatic Segmentation of Liver Structure in CT Images Using a Neural Network," *IEICE TRANSACTINS on Fundamentals of Electronics, Communications and Computer Sciences*. Vol. E77-A, no. 11, pp.1892-1895, November 1994.

[23] A.Hasegawa, S.C.B. Lo, J.S. Lin, M.T. Freedman and S.K. Mun,"A Shift-Invariant Neural Network for the Lung Field Segmentation in Chest Radiography," *The Journal of VLSI Signal Processing*. Vol. 18, no. 3, pp.241-250, April 1998.

[24] Mehmed Ozkan, Benoit M. Dawant and Robert Margolin,"Neural- Network-Based Segmentation of Multi-Modal Medical Images: A Comparative and Prospective Study," *IEEE Transactions on Medical Imaging*. Vol.12,no.3, pp.534-544, September 1993.

[25] Yan Li Peng Wen Powers, D. Clark, C.R,"LSB Neural Network Based Segmentation of MR Brain Images," *Systems, Man, and Cybernetics.* Vol.6, pp.822-825, 1999.

[26] Ian Middleton and Robert I. Damper,"Segmentation of Magnetic Resonance Images Using a Combination of Neural Networks and Active Contour Models," *Medical Engineering & Physics*. Vol.26, no.1, pp.71-86, January 2004.

[27] Chiou GI, Hwang JN,"A Neural Network-Based Stochastic Active Model (NNS-SNAKE) for Contour Finding of Distinct Features," *IEEE Trans Image Process.* Vol.4, no.10, pp.1407-1416, October 1995.

[28] John.E.Koss, F.D.Newman FD, T.K.Johnson, and D.L.Kirch,"Abdominal Organ Segmentation Using Texture Transforms and a Hopfield Neural Network," *IEEE Transactions on Medical Imaging*.vol.18, no.7, pp.640-648, July 1999.

[29] Jzau-Sheng Lin, Kuo-Sheng Cheng, Chi-Wu Mao,"A Fuzzy Hopfield Neural Network for Medical Image Segmentation," *IEEE Transactions on Nuclear Science.* Vol.43, no.4 , pp.2389-2398, August 1996.

[30] Chwen Liang Chang and Yu Tai Ching,"Fuzzy Hopfield Neural Network with Fixed Weight for Medical Image Segmentation," *Optical Engineering.* Vol.41, no.2, pp.351-358, 2002.

[31] K.S. Cheng, J.S. Lin, C.W. Mao,"the Application of Competitive Hopfield neural network to Medical Image Segmentation," *IEEE Transactions on Medical Imaging.* Vol.15, no.4 ,pp.560-567, August 1996.

[32] Wei Sun and Yaonan Wang,"Segmentation Method of MRI Using Fuzzy Gaussian Basis Neural Network," *Neural Information Processing*. Vol.8, no.2 , pp.19-24, August 2005.

[33] Chien-Cheng Lee, and Pau-Choo Chung,"Recognizing Abdominal Organs in CT Images Using Contextual Neural Networks and Fuzzy Rules," In *Proc.of Int. Conf. of the IEEE Engineering in Medicine and Biology Society*, pp. 1745-1748, Chicago, IL, July 23-28 2000.

[34] Y. Wang, T. Adali, S.Y. Kung et al,"Quantification and Segmentation of Brain Tissues from MR Images - a Probabilistic Neural Network Approach," *IEEE Transactions on Image Processing*. Vol.7, no.8 , pp.1165-1181, August 1998.

[35] W.E. Reddick, J.O. Glass, E.N. Cook et al,"Automated Segmentation and Classi1cation of Multispectral Magnetic Resonance Images of Brain Using Arti1cial Neural

Networks," *IEEE Trans. Med. Imaging*. Vol.16, no.6, pp.911-918, December 1997.

[36] A.Pitiot, A.W.Toga, N.Ayache, P.Thompson,"Texture Based MRI Segmentation with a Two-Stage Hybrid Neural Classifier," *Proceedings of the 2002 International Joint Conference on Neural Networks*.vol.3, pp.2053-2058, 12-17 May 2002.

[37] Wu Y, Giger ML, Doi K et al,"Artificial Neural Network in Mammography: Application to Decision Making in the Diagnosis of Breast Cancer," *Radiology*. Vol.187, pp.81-87, April 1993.

[38] Tourassi GD, Floyd CE Jr,"Artificial Neural Networks for Single Photon Emission Computed Tomography," *Invest Radiol*. Vol.28, no.8 , pp.671-677, August 1993.

[39] Maclin PS, Dempsey J,"Using Artificial Neural Networks to Diagnose Hepatic Masses,"*Journal of Medical Systems.* Vol.16, no.5, pp.215-225, October 1992.

[40] Wolberg WH, Sreet WN, Mangasarian OL,"Image Analysis and Machine Learning Applied to Breast Cancer Diagnosis and Prognosis," *Anal Quant Cytol Histol*. Vol.17, no.2, pp.77-87, April 1995.

[41] Fikret Ercal, Anurag Chawla, William V. Stoecker, His-Chieh Lee, and Randy H. Moss,"Neural Network Diagnosis of Malignant Melanoma from Color Images," *IEEE Transactions on biomedical engineering.* Vol.41, no.9 , pp.837-845,September 1994.

[42] Ashizawa K, Ishida T, MacMahon H, Vyborny CJ, Katsuragawa S, Doi K,"Artificial Neural Networks in Chest Radiography: Application to the Differential Diagnosis of Interstitial Lung Disease," *Academic Radiology*. Vol.6, no.1, pp.2-9, January 1999.

[43] K. Suzuki, S. G. Armato, F. Li, S. Sone, and K. Doi,"Massive Training Artificial Neural Network (MTANN) for Reduction of False Positives in Computerized Detection of Lung Nodules in Low-Dose CT," *Med. Phys.* Vol.30, no.7, pp.1602-1617, July 2003.

[44] K. Suzuki, S. G. Armato, F. Li, S. Sone, and K. Doi,"Effect of a Small Number of Training Cases on the Performance of Massive Training Artificial Neural Network (MTANN) for Reduction of False Positives in Computerized Detection of Lung Nodules in Low-Dose CT," In *Proc. SPIE (Medical Imaging)*. Vol.5023, pp.1355-1366, July 2003.

[45] K. Suzuki and K. Doi,"Characteristics of a Massive Training Artificial Neural Network (MTANN) in the Distinction Between Lung Nodules and Vessels in CT images," in *Computer Assisted Radiology and Surgery (CARS)*, Chicago, IL, pp.923–928, 2004.

[46] H. Arimura, S. Katsuragawa, K. Suzuki, F. Li, J. Shiraishi, S. Sone, and K. Doi,"Computerized Scheme for Automated Detection of Lung Nodules in Low-Dose CT Images for Lung Cancer Screening," *Acad. Radiol.* Vol.11, no.6, pp.617-629, June 2004.

[47] K. Suzuki, J. Shiraishi, H. Abe, H. MacMahon, and K. Doi,"False-Positive Reduction in Computer-Aided Diagnostic Scheme for Detecting Nodules in Chest Radiographs by Means of Massive Training Artificial Neural Network," *Acad. Radiol.* Vol.12, no.2 , pp.191-201, February 2005.

[48] Zhu Y, Yan H,"Computerized Tumor Boundary Detection Using a Hopfield Neural Network," *IEEE Transactions on Medical Imaging*. Vol.16, no.1 , pp.55-67, February 1997.

[49] Innocent PR, Barnes M, John R,"Application of the Fuzzy ART/MAP and MinMax/MAP Neural Network Models to Radiographic Image Classification," *Artif. Intell. in Med.* Vol.11, no.3, pp.241-263, 1997.

[50] Yasser M. Kadah, Aly A. Farag, Jacek M. Zaruda, Ahmed M. Badawi, and Abou-Bakr M. Youssef,"Classification Algorithms for Quantitative Tissue Characterization of Diffuse Liver Disease from Ultrasound Images," *IEEE transactions on Medical Imaging*. Vol.15, no.4, pp.466-478, August 1996.

[51] E-Liang Chen, Pau-CHoo Chung, Ching-Liang Chen, Hong-Ming Tsai and Chein I Chang,"An Automatic Diagnostic system for CT Liver Image Classification," *IEEE Transactions Biomedical Engineering.* Vol.45, no.6, pp.783-794, June 1998.

[52] Pavlopoulos.S, Kyriacou.E, Koutsouris.D, Blekas.K, Stafylopatis.A, Zoumpoulis.P,"Fuzzy Neural Network-Based Texture Analysis of Ultrasonic Images," *IEEE Engineering in Medicine and Biology*. Vol.19, no.1, pp.39-47, Jan-Feb. 2000.

[53] H.P. Chan, S.C.B. Lo, B. Sahiner et al,"Computer-Aided Detection of Mammographic Microcalcifications: Pattern Recognition with an Artificial Neural Network," *Medical Physics* vol. 22, no.10, pp.1555-1567, October 1995.

[54] B. Sahiner, H.P. Chan, N. Petrick et al,"Classification of Mass and Normal Breast Tissue - a Convolution Neural Network Classifier with Spatial Domain and Texture Images," *IEEE Transactions on Medical Imaging*. Vol.15, no.5, pp.598-610, October 1996.

[55] S.C.B. Lo, H.P. Chan, J.S. Lin et al, "Artificial Convolution Neural Network for Medical Image Pattern Recognition," *Neural Networks*. Vol.8, no.7-8 ,pp.1201-1214, 1995.

[56] H P Chan *et al*, "Computerized Classification of Malignant and Benign Microcalcifications on Mammograms: Texture Analysis Using an Artificial Neural Network," *Phys. Med. Biol.* Vol.42, no.3 , pp.549-567, March 1997.

[57] B. Verma, J. Zakos, "A Computer-Aided Diagnosis System for Digital Mammograms Based on Fuzzy-Neural and Feature Extraction Techniques," *IEEE Transactions on Information Technology in Biomedicine*. Vol.5, no.1 , pp.46-54, March 2001.

[58] Zhi Hua Zhou, Yuan Jiang, Yu Bin Yang, Shi Fu Chen, "Lung Cancer Cell Identification Based on Artificial Neural Network Ensembles," *Artificial Intelligence in Medicine*. Vol.24, no.1, pp.25-36, January 2002.

[59] K.Tumer, N.Ramanujam, J. Ghosh, and R. Richards-Kortum, "Ensembles of Radial Basis Function Networks for Spectroscopicdetection of Cervical Precancer," *IEEE Transactions on Biomedical*. Vol.45, no.8, pp.953-961, August 1998.

[60] L.Perlovsky,"Conundrum of combinatorial complexity," *IEEE Tranactions on Pattern Analysis and Machine Intelligence.* Vol.20, no.6, pp.666-670, June 1998.

[61] S. Horikawa, T. Furuhashi, and Y. Uchikawa,"On Fuzzy Modeling Using Neural Networks with the Back Propagation Algorithm,"*IEEE Trans. on Neural Networks*. Vol.3, no.5, pp.801-806, September 1992.

[62] Chan HP, Sahiner B, Wagner RF, et al,"Classifier Design for Computer Aided Diagnosis: Effects of Finite Sample Size on the Mean Performance of Classical and Neural Network Classifiers," *Med Phys.* Vol.26, no.12, pp.2654–2668, 1999.

[63] Sahiner B, Chan H P, Petrich N, et al," Feature Selection and Classifier Performance in Computer Aided Diagnosis: the Effect of Finite Sample Size," *Med Phys.* Vol.27, no.1, pp.1509-1522, June 2000.

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 027-030

Research and Application of Auto-profiling Control Algorithm for 0Cr18Ni10Ti

Qiang Wang[1], XuGuang Min[2], ZhongYue Xiao[1], and Yanhui Zhou[1]
[1] Jinggangshan University, Ji'an, China
Email: wangqiang2188@tom.com
[2] Jiangxi Science&Technology Normal University, Nanchang, China

***Abstract*—The paper analysis the material composition, properties and application technology of 0Cr18Ni10Ti , and it is an understanding of the material's mechanical properties and corrosion-resistance. The using of stabilization heat treatment process, solution heat treatment, controlling the chemical composition and production process conditions, etc. improved the overall performance of 0Cr18Ni10Ti and the capacity of arain boundary corrosion resistance.Method of the multi-punching process and profiling&following control algorithm processing is then established and controlled to the machining ration speed, tool feed and cutting depth,which achieves better results. The practical application indicates that producing the flexible laminated membrane coupling by using 0Cr18Ni10Ti makes the coupling highly efficient and reliable in transmission, lighter, easier to assemble/disassemble, easier to machine and less costly, and thereby that such an application is noticeably superior than those of traditional materials. It guarantees product quality and increased work efficiency.**

***Index Terms*—control algorithm; 0CrNi10Ti; Coupling Disk; Solution Heat Treatment; Stabilization Treatment**

I. Introduction

Stainless steel, one of the greatest inventions in 20th century, covers a range of more than 300 steels. Materials under this category are widely used for their excellent corrosion- and erosion-resistance features and their high comprehensive performances, and consequently they include nearly all steels for special purpose. How the steels are corrosion-resistant is decided by the alloy elements contained in them, which resist faintly corrosive media including water, steam and air as well chemical erosive media like acid, alkali and salt. Chromium (Cr) is a basic element in stainless steels that works against corrosion. When chromium content in steel reaches at some 12%, the element reacts with oxygen in corrosive medium and develops a thin oxide film on the surface which prevents further corrosion in substrate. Among the steels of various substrates, ferritic stainless steel, containing chromium between 12% and 30%, possesses good corrosion-resistance, tenacity and solderability, all of which increase along with higher chromium content, and it behaves better than other stainless steels in resistance to chloride stress corrosion. Martensitic stainless steel is of high strength but of poor plasticity and solderability. Whereas Austenitic stainless steel contains chromium more than 18% in content and nickel about 8% and trifle of molybdenum, titanium, nitrogen and other elements. This metal, combining merits of austenitic steel and ferritic stainless steel, is of desirable comprehensive performances and resistance to multiple corrosions, in addition to its superplasticity. One typical member of this family is 0Cr18Ni10Ti. By using the profiling&following control algorithm processing and controlled to the machining ration speed, tool feed and cutting depth,which achieves better mechanical processing results.

II. Overview of 0Cr18Ni10Ti

0Cr18Ni10Ti, also called as SS321, is a kind of chromium-nickel austenite stainless steel in series 300. Besides chromium, as is known, stainless steel is often made with alloy elements like nickel, molybdenum, titanium, niobium, copper, nitrogen and others, which are added in to meet the requirements of proper use in different situations and of good machining properties. For instance, steel with trifle sulfur and phosphor is easier to be machined, that with molybdenum given a special structure which is corrosion-resistant and thus usable in food industry and medical operation apparatus industry. The 0Cr18Ni10Ti is one of the most used stainless steel. It is of performances similar to 0Cr18Ni9 except that the former contains titanium, which is added to restrain the precipitation of (Cr, Fe) 23C6 at the grain boundaries and reduce the intergranular corrosion possibility, thus enhancing the material in its weld corrosion-resistance.

Thanks to its excellent mechanical properties and anti-corrosion performances, the material is widely used across industries.

A. Compositions

The material is defined with a broad range of chemical elements, as shown in Table I. However, rigid control of its chemical composition is significant. As Cr, Ni and Ti elements are vital to the resistance to intergranular corrosion, the chemical composition in the steel should be determined specifically according to the operational requirements of parts made from it and to the conditions of processing equipments. In case that the material is to be used for welding, some ferrite retained in the steel is helpful to improve its welding performance. Then the content of ferrite should be properly figured out in procedure of composition decision, where Fisher Phase Diagram is used to determine the as-cast ferrite content based on nickel equivalent and chromium equivalent. The chemical elements possibly contained in 0Cr18Ni10Ti

allow it designed as austenitic steel or austenitic-plus-ferritic dual-phase steel. Considering that the material under study is with only minute variation in ferrite when under heat processing, it is phase-transformation-free steel. Hence the steel cannot be strengthened through phase transformation, but only by alloy or through deformation. In deciding the steel's composition the total quantity of titanium and carbon should be primarily considered. The content of carbon should be controlled between 0.05 and 0.065% and that of titanium correspondingly between 0.45 and 0.65%. It should be noted that titanium content must be rigidly controlled according to the standard lest it is over or under the prescribed range. As titanium is apt to oxidizing, the metal is usually added into the steel sometime before the latter is tapped. Sufficient time should be allocated so that the titanium metal is melted and adequately homogenized lest local concentration leading to uneven distribution. Additionally, to improve the properties and anti-corrosion performances of 0Cr18Ni10Ti, processes of solution heat treatment and stabilization treatment are mostly applied.

B. Properties

The mechanical properties of 0Cr18Ni10Ti are as shown in Table II.

TABLE I.
CHEMICAL COMPOSITIONS OF 0CR18NI10TI, IN [%]

Carbon (C)	Silicon (Si)	Manganese (Mn)	Sulfur (S)	Phosphor (P)	Chromium (Cr)	Nickel (Ni)	Titanium (Ti)	Cobalt (Co)
≤0.08	≤1.00	≤2.00	≤0.03	≤0.035	17.00～19.00	9.00～12.00	≥5×C	≤0.08

TABLE II.
TYPE SIZES FOR CAMERA-READY PAPERS

Item	Yield Point σs（σ0.2）	Tensile Strength σb	Elongation δ5	Concentration of Area ψ	Hardness		
Unit	[MPa]	[MPa]	[%]	[%]	[HB]	[MPa]	[MPa]
Value	205	520	40	50	187	205	520

As shown above, stainless steel 0Cr18Ni10Ti is of good comprehensive performances, particularly high plasticity, tenacity, contraction of area and impact toughness, as well as desirable solderability. The material is workable for long time at temperature between -196 to 600 ℃ and simultaneously assuring with proper mechanical properties as well as long service life.

In some common stainless steels there may be reaction between chromium and carbon at their sensitization zone, which develops CrxCy that results to intergranular chromium depletion, hence depriving those metals' corrosion-resistant merits and leading to intergranular corrosion. On the contrary, 0Cr18Ni10Ti is in austenitic microstructure under ambient temperature. The titanium contained in, which is a stabilizing agent to austenite, is more apt to react with the carbon in the stainless steel, combination of which is TiC, thus effective impeding the combination of carbon and chromium, and avoiding intergranular chromium depletion that brings about intergranular corrosion.

III. CASE OF APPLICATION AND KEY PROCESSES THEREIN

Stainless steel 0Cr18Ni10Ti has been applied in production of coupling disks. As is known, shaft coupling falls into conventional, fatigue-enduring, heat-tolerant and corrosion-resistant types. Those types are used in four different situations, under which disk couplings to be used will be of same general structure but with disks made from different materials. Nevertheless, disk coupling has always high requirements on the materials used to make its disks—they need to possess high tensile strength and elasticity and at same time high fatigue-endurance. When the coupling is to work at high temperature or under hostile situation, those materials need also to possess the physical and chemical properties that make the disks adaptive to situation changes, in particular good heat-tolerant and corrosion-resistant capacities. The shaft coupling of a diesel generating set in a military equipment is a typical application. The generating set is designed to work in extreme heat over 100℃ and in moisture, in addition to trace of acidic gas emitted from the internal insulation materials. Consequently the disk material is required to be extreme heat-tolerant and acid-resistant, and for that cause the material 0Cr18Ni10Ti is selected.

A. Profiling&following control algorithm processing

Assumptions the three directions on the deformation in the processing are GX、Gy、Gz, then integrated deformation μ using the following equation：

$$\mu = \sqrt{G_X + G_Y + G_Z}$$

The integrated deformation should be controlled during processing, namely: $\mu_{\min} \le \mu \le \mu_{\max}$.

Among them, $\mu_{\min}$ is to allow deformation of the smallest integrated to ensure the machining tool is not detached from the workpiece surface; $\mu_{\max}$ is the maximum allowed amount of integrated deformation to ensure that processing does not occur when the knife had cut phenomenon. Therefore, controlling the μ volume between $\mu_{\min}$ and $\mu_{\max}$ the comprehensive changes in deformation, which range from changes in the accuracy of the machine itself and processing of allowable error determined.

Movement rate set V, you can press the following formula to allocate X axis and Y-axis speed:

If: $0 < \mu_{xy} < 2\mu_{oxy}$

$$\text{Then}: V_n = -V(1 - \frac{\mu_{xy}}{\mu_{oxy}}) = V\frac{\Delta\mu_{XY}}{\mu_{XY}}$$

$$V_m = \pm\sqrt{V^2 - V_n^2}$$

$$\text{If}: \mu_{xy} > 2\mu_{oxy}$$

$$\text{Then}: V_n = V, V_m = 0$$

Access to the value of V_n and V_m, it is easy solved V_x and V_y.

B. Key Technical Processes

Control of technical processes is critical. Normally steel 0Cr18Ni10Ti is with good behaviors at ambient and lower temperatures in tenacity, plasticity, solderability, corrosion-resistance and magnetic neutrality. However, if it is produced through improperly controlled process, the steel may occur with intergranular corrosion which lowers the material's strength and plasticity. This is quite dangerous. Besides, 0Cr18Ni10Ti is a material of poorer performance in machining, which means difficulties in stamping and machining of the coupling disks. Therefore, in the production the key processes should include the following.

(1) Strict control over the chemical compositions, especially the percentages of chromium, nickel and titanium elements, is critical to prevention from intergranular corrosion, as content of the alloy elements is extremely vital to avoidance of the corrosion.

(2) Forging may be applied to pieces which are of proper chemical compositions but with large grains, nonuniform microstructure or high ferrite content in rough strip distribution. By this process, the material may be turned into one with uniform microstructure, fine grain, lowered ferritic content and better distribution, thus noticeably improved in its performance against boiling nitric acid intergranular corrosion.

(3) For better comprehensive performances and corrosion-resistance of the material, it is necessary to apply solution heat treatment and stabilization treatment.

(4) Disk design should be based on analyses and confirmation that it is stable. If the disk was designed unstable, then adjustments should be made onto structure and number of disks.

(5) Multiple-stage stamping process may best be applied instead of one-off process, as 0Cr18Ni10Ti is of poor performance in stamping. Stamped disk which needs finishing can be machined on a lathe.

(6) Deformation of product should be monitored and restrained during the processing of disks.

C. Solution Heat Treatment

This process is to dissolve carbide into austenite and to homogenize the latter. The technique is undertaken as follows: the material is heated to a temperature between 920 and 1150℃,then chilled to ambient temperature. Carbon element in the material is made supersaturated and the carbide phase completely or substantially dissolves. Carbon is made dissolved into austenite as prevention from its combination with chromium into high chromium carbide, in aim to strengthen the solid solution, enhance the material in tenacity and corrosion-resistance, eliminate its stress and soften it, as well as prepare it for further stabilization treatment.

It is noted that over heat in solution treatment will increase the size of grains in austenite. When volume is certain, a coarser grain will have less intergranular area. And under given sensitization treatment conditions the precipitation of carbide is a relative fixed value. Thereby a coarser grain will have more carbide precipitation per unit area, which brings heavier intergranular chromium depletion, thus leading to higher intergranular corrosion sensitivity.

D. Stabilization Treatment

Stabilization treatment is undertaken after a process of solution heat treatment, and for adequate effect, a proper one.The technique is as follows: The material is heat-treated between 850 and 930℃, heat-maintained, and then cooled by air. By then chromic carbide dissolves completely, while titanic carbide incompletely. The latter precipitates in the cooling course and prevent the carbon from integrating chromic carbide, thus effectively avoiding intergranular corrosion and ultimately improving the intergranular corrosion-resistance of 0Cr18Ni10Ti. When the stabilization treatment is completed, the subsequent cooling procedure, whether it is by air or by water, will exert roughly same effect on the material in respect of intergranular corrosion resistance.

E. Machining

0Cr18Ni10Ti is a material hard to be machined. It is of high tenacity, low heat conductivity, which is only a half or even one third that of common steels. Hardening and deformation occur when the material is cold machined at ambient temperature, which may turn material phase into

martensite. In the machining process there will also be great force action and extreme heat between the rake face and the chips. So the chips tend to adhere to the blade and form build-ups, and most of the heat produced in cutting goes to the tool and rise the tool temperature, which together worsen the tool wearing. As a result, common tools can serve only for shorter life, work less efficiently while more consumptively, and develop parts of much lower precision and poorer surface quality. Thus it is necessary to use tools special for stainless steel machining, in this case professional tools from a Swedish tool company and from OSG, and to define and control exactly the suitable machine rotation speed, tool feed and cut depth. Only by those means desirable cutting is attainable.

IV. CLOSING REMARKS

Thanks to its good comprehensive performances, particularly its high plasticity, tenacity, elongation, contraction of area and impact toughness, steel 0Cr18Ni10Ti is desirable for production of coupling disks. Use of the steel makes coupling easier to machine, less costly, lighter and more easily to assemble/disassemble. Moreover, it makes the coupling more efficient and reliable in transmission operation. The application of the steel is more advantageous than previously used1Cr18Ni9Ti, 0Cr18Ni9 and/or 65Mn. Nevertheless, in production of the material it is vital to control its chemical composition, particularly the percentages of Cr, Ni and Ti, and to control the technical processes, so that intergranular corrosion is avoided.On the basis of profiling&following control algorithm processing is uesd in processing,and three elements of machining by cutting is introduced to parctical use ,the results is well.

Besides, it needs to adopt solution heat treatment and stabilization treatment so as to improve and better the material's comprehensive performances and corrosion-resistance.

REFERENCES

[1] WANG Jing-wen: Structure & Properties of Cold-deformed and Recrystallized 1Cr18Ni9Ti Austenitic Stainless Stee.Hot Working Technology, 2007,(02): 26-29

[2] GU Mei-lin, HUANG Chuan-zhen, XIAO Shou-rong: Study on Cutting Performance of TiB2 Matrix Ceramic Tool in Machining Stainless Steel. Manufacturing Technology& Machine Tool, 2008, (08):88-91

[3] ZHANG Yi-ping: Study on Corrosion among Crystalline Grains and Heat-treatment Technique of 1Cr1 8Ni9Ti. Coal Mine Machinery, 2006, 27 (07): 119-121

[4] FU Qian-jin, ZHANG Bo, LIU Si-qin et al: Research for Forging Process of 0Cr18Ni10Ti Steel. Heavy Casting and Forging, 2001, (04):3-6

[5] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," *Phil. Trans. Roy. Soc. London*, vol. A247, pp. 529–551, April 1955.

[6] CHEN Rui, LIU Jing, PANG YU-si: Study of Intercrystalline Corrosion Test on 1Cr18Ni9Ti Stainless Steel.Henan Metallurgy, 2006, 14 (04): 14-17

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 031-034

Research and Realization about Conversion Algorithm of PDF Format into PS Format

Xingfu Wang, Lei Qian, Fuyou Mao, and Zhaosheng Zhu
School of Computer Science & Technology in the University of Science and Technology of China,
USTC,Hefei, Anhui, P.R.China postalcode: 230027
wangxfu@ustc.edu.cn, mfy@ustc.edu.cn

***Abstract*—This paper firstly introduces the characteristics of PostScript document and PDF document as the basis, and proposes the necessity and the feasibility of the conversion from the PDF document format to the PostScript language program. Secondly, it studies the main algorithm and technology of the conversion process and realizes the information extraction for PDF document lastly, with achieving the software algorithm for the conversion from PDF document format into PS format on the basis of the text.**

***Index Terms*—PDF PS（PostScript）, Format conversion, Object-oriented, Interpreter**

I. INTRODUCTION

PDF (Portable Document Format), developed by Adobe, is a open electronic file format which is suitable for transmitting and sharing the file between the different computer system, with the advantages like cross-platform, high-compression, being suitable for screen viewing and network transmission, the protection of documents, e-reviewed, Print output of higher quality. PS language (PostScript description language, namely PDL), is also a matter of fact in the printing industry standard owned by Adobe, which may describe the exquisite layout and occupy the dominant position in the area of printing presently. Though PDF develops on the basis of PostScript, it is not a programming language, hence it must be converted to PostScript date flow when the output is on the common PostScript printer, or the PDF document must be converted to PS document before the output.

II. CHARACTERISTICS AND STRUCTURES OF PS LANGUAGE

PostScript language as a common page description language in the modern printing technology is one kind of Interpretative programming languages, with strong graphics function. PostScript language takes Adobe's imaging model theory as the basis for the description of pages, and its main application is printing pages or describing text, graphics, shape and sampling images on the display pages. The procedures, wrote by using PostScript language, might describe the correspondence from Organizing layout system to printing system by documents, or control displayed result of objects on the display. The description produced by PostScript language is a high-grade description, and it is noting with the equipment, thus it has become an important and indispensable component of the high-quality printing and output. At present, many printers, film Phototypesetters, printing Phototypesetters, digital printers and other equipments have all been installed the PostScript interpreter; many RIP processing software also take PostScript interpreter as its core technology, and many image-processing software, graphics and design software and typesetting software support PostScript as also, in which the most famous typical software are Photoshop, CorelDraw, Illustrator, Freehand, QuarkXPress, PageMaker, and so on.

A. The basic characteristics of PostScript language

Compares with other file formats, PostScript language has the capacity of page description and interactive handling ability between text and images. In addition, PostScript document has many advantages. For example, one of the advantages is that PostScript document is independent of the equipment and independent of the operating system platform. Graphical environment of UNIX itself takes the support of PostScript as its core part, so no matter using Windows operating system or using UNIX operating system, PostScript document can be read and print well, hence it is convenient to exchange. Moreover, although the PostScript documents also use the binary encoding to save, it generally expresses and stores information by the text mode of ASCII code, by which it can be easier to read and edit.

B. The composition of PS document and the characteristics of procedures

PostScript language is a high-level interpretative language, which has a wealth of data types and the control statement. Besides, it proposes the concept of the process and, like other programming languages (such as the C language), adopts the structured programming method of the top-down and stepwise refinement. A PostScript file with a good structure is usually composed of two parts: preface part and description part.

Syntax, data types and implementation of semantics is the three basic parts of any a PostScript procedure. Coupled with its outstanding capacity of the graphics, images and text, the PostScript language can be competent to control and print the task of dealing with pages. The grammar of PostScript language is quite simple, but its function is very powerful, which comes from it can without any restrictions to combine its each characteristic in any way. Using these functions of PS

This work is supported by National Natural Science Foundation of China (NO.60773037&60970128), by Teaching Research Project of Anhui Province (2007)

language, people can design complex graphics application procedure.

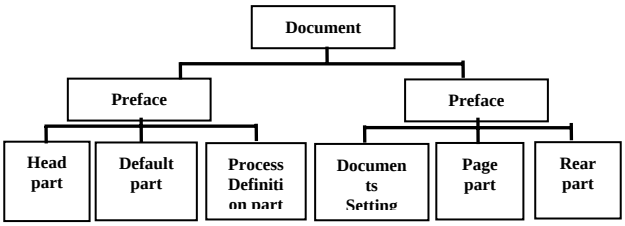


Figure 1. composition of PS text structure

III. OBJECTS AND STRUCTURES OF PDF

PDF (Portable Document Format) is a structurized document format. The initial objective to design PDF file format by Adobe is to adapt to digital pre-press workflow and needs of the development of electronic publishing market. Although PDF file format relies on imaging model of PostScript language, and describing the text and image have nothing to do with the equipment and the resolution.

PDF document is transitivity, support the interaction, support the sound and the flash, and support random access of the contents pages, which also improve the various operating speed for pages, support the alter mode of supplement unceasingly in order to reduce the revises and raise the efficiency, and support kinds of encoding methods for a more compact structure of the file. In addition, it has some other functions like font-independent, platform-independent and security control. These characteristics of PDF mean its dual nature, that is, it continues to support prepress workflow and makes it more perfect, meanwhile, PDF also provides a basis of the spread of digital files for the field of electronic publishing and thus it becomes the major method of the cross-media publishing file.

A. The major difference between PDF and PostScript

PostScript language is not only a page description language, but also a common programming language, which includes the process, the variables and the control structure. Therefore, a program written by PostScript language must be translated into page description information for being accepted by the output device. Compared with these, PDF does not have the characteristic of the programming language, which causes lots of differences between PDF and PostScript.

a) PDF is a document format, but not a programming language, so it adopts the method, which has nothing to do with the application software of producing documents, hardware and operating system, to describe the document. Therefore, a common programming structure is not exists in the PDF files, the result of which is the imaging is much simpler and describing pages is much easier too.

b) Although PostScript describes the objects of pages having nothing to do with the equipment, it is related with the pages. Each page in the PDF file is unrelated to other pages, which eliminates the randomicity of PostScript language. Hence, PDF file provides the basic of the available consistency and the divinable result for the printing and producing system.

c) PDF document in principle can not be directly downloaded to the PostScript device for outputting. If in the hope of outputting PDF document in the PostScript device, these steps below could be carried out.

i. Insert the PostScript language process definition sets Prosets of the instruction characters which can realize the PDF page description.

ii. Extract each page's content and the description of each page is necessary, because the description part of the traditional PostScript language procedure uses the appropriative process.

iii. Decode the compressed text, graphics and image data, without the data encoded by Flate filter, and not decode the compressed data in the document when the device is PostScript Level 2, because it can accept the compressed data in the PS.

iv. Insert the resource, such as the font, into PostScript files, replace the font definition and insert according to the necessary. The basic is the font specifications of the PDF files.

v. Place the information by the correct order, the result should be, in a usually sense, a PostScript program file, containing all the visible part of the document, but hyperlinks, comments, bookmarks and other PDF units will no longer be included in the documents.

vi. Sent PostScript language documents to the printer.

B. The object of PDF and the structure of the document

PDF supports following seven kinds of basic objects. Object encoding and produce encoding are used as the unique identifier of the indirect object, and the indirect object always keeps object encoding and produce encoding of the beginning during its existence. Even if the indirect object was revised, these two encoding will still remain. Each indirect object has its unique object encoding, and the encoding of the indirect object usually starts continuously from one, but it is necessary. When the object is deleted from the document, all the produce code will become zero.

The structure of PDF (i.e. physical structure) includes four parts: document head, document body, cross-reference table and document rear. The method of the PDF file organization can be represented as the following structure:

<PDF document>::= < document head >
< document body >
< cross-reference table >
<document rear >

Document head specifies the PDF standard version mumble which the document complies with, and it appears in the first line of the PDF document. Document body is composed of a series of PDF indirect object (IndirectObject). These objects are the fundamental types of objects, which are used for the different components of documents, such as fonts, pages and images. Cross-reference table is the address index table of the indirect object which is established for the random access of the indirect object. Document rear specifies the address of the cross-reference table, namely it not only points out the Catalog of the document body, but also holds the encryption and other security information.

PDF document structure is a tree structure, the tree's root node is the Catalog dictionary of the PDF files, and there are four subtrees below the root node: Pages Tree, Outline Tree, Article Threads, Named Destination. In the page tree, all the page object is this tree's leaf node, and they will inherit the various attribute value of their father node to take its corresponding attribute default value, which function is to let the other objects of the PDF document can also use the the character string name to express some page region.

IV. RESEARCH THE CONVERSION FROM PDF DOCUMENTS TO PS FILE

PDF is the document format, while PostScript is the programming language, so this principled difference extends the difference of the expression of the document content inevitably, which cause the difference in producing and using the PDF document, as well as refer how to putout the hard copy of the PDF file in the PostScript printer or the non-PostScript printer.

A. The feasibility of the conversion from PDF documents to PS file

PDF takes root in PostScript, thus there are many similarities, which decides to the feasibility of the document conversion. PDF file format uses the imaging model of PostScript language to express text, graphics and other objects, similar to PostScript language procedure, a PDF page description draws a page through placing “pigment” In the selected region. Its imaging characteristics can be summarized as follows:

Colored page objects may be abstract to "map", which could be the character shape, or be expressed by the digital sampling of the photo, and the region (graphics) defined with curves and straight-line.

Imaging operation is relatively flexible, not limited to the print record or film, which can also on paper and other record. It could use any color of the pigment when imaging.

For the page is viewed as “map” when imaging, thus it could be cut into other shape, with the result that the part of “map” only in this sharp can be appeared on the page.

Typical PostScript language procedure needs to take the tag instruction character definition of the PostScript language to define its set of instruction character, used to describe the pages and control output. While PDF defines its own instruction character sets, and the most instruction characters are very similar to the PostScript instruction characters.

Although PDF document can not be explained by the PostScript translator device, but in the conversion function of the translator device, page description of the PDF document can transform to the PostScript language procedure. Basic method is as follows:

The difference between PostScript and PDF mainly from page description instruction character , so the solution of this problem is to insert the PostScript language process definition sets Prosets of the instruction character s which can realize the PDF page description.

Extract the content on each page of the PDF document, because it is not required to store by the logical order in the PDF document format, so it is necessary to describing each page. Because the description part of the traditional PostScript language procedure uses the appropriative process, such as represent “moveto” by “m”, represent “lineto” by “l” and so on.

To decode the compressed text, image and graphics data is not a simple work for ordinary users. PostScript Level 2 printers do not need to decode the compressed data in the PDF documents, but without the data encoded by Flate compression algorithm. But now most of PDF documents use Flate compression algorithm, so it must first achieve the decompression of Flate compression algorithm, and then complete the conversion from PS to PDF.

Insert the resource, such as the font, into PostScript files, replace the font definition and insert according to the necessary. The basic is the font specifications of the PDF files.

Place the information by the correct order, the result should be, in a usually sense, a PostScript program file, containing all the visible part of the document, but hyperlinks, comments, bookmarks and other PDF units will no longer be included in the documents.

This can obtain the PostScript language document, and then could printout this document after sending it to the printer. In view of this, the conversion from the PDF documents to PS document is entirely feasible, and its key lies in the realization of translator device.

B. The conversion algorithm from PDF document to PS file

a) The description of the algorithm

To complete the conversion from PDF files to PS format, firstly need to extract page information of the PDF file, which includes resources, content, annotation and etc. Then express each page by PS language, first write into the resource referred in this page, and then write into the content of this page, so it is only necessary to express with the corresponding PS language in this part. This part is not so complex as the extraction of the PDF information above, for most instruction character of PDF is the same as that of PostScript, thus it is relatively simple as long as the corresponding of these two instruction character is achieved in advance. The below is the algorithm which can constructs the information extractor of the PDF file.

Achieve the lexical analysis and the syntax analysis of the PDF document, to analyze the single lexical symbol and the grammar object separately. The object format has been introduced above, and we need to achieve its identification.

Find the location of document rear according to the keywords, because document rear contains the address of Catalog and the address of the cross-reference table. then analyze the content of the cross-reference table, that is the offset of each indirect object address, and produce encoding.

Analyze the content of that Catalog, which is the root object of the PDF files, and obtain the attributes of the Catalog, such as the Outline attribute, Pages attribute and so on, that is its indirect object address.

Analyze the Outline attributes, establish the outline tree, and obtain the document outline.

Analyze the Pages attribute, through the recursion to establish the page tree, and obtain the pages attribute during the establishment of the page tree. Through the recursive it is not only can reduce the complexity of the codes, but also can realize the inheritable character of pages tree attribute. Realizes as follows:

BulidPageTree (pages tree object)

{Obtain the Kid attribute of this object

If (this object has not subtree)

Return;

For (all the subnode of the pages tree object)

{Set the subnode as the son of this object, and inherit its attribute;

BuildPageTree (this subnode); }}

This process is a little similar to tree's first foreword traversal.Analyze the other options included in the attributes of Catalog.

b) Flow chart of the conversion algorithm as shown in Figure 2:

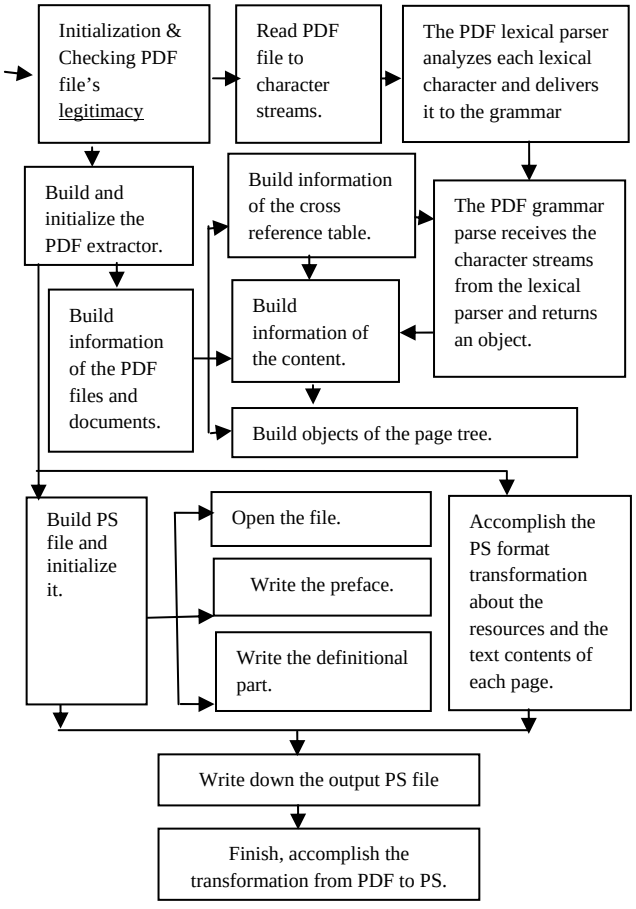


Figure 2. Software realization of the format converter

V. SOFTWARE REALIZATION OF THE FORMAT CONVERTER

A. The applicable scope of software

Said earlier, the most of PDF documents use the Flate compression algorithm. If wanting to know the contents of stream flow, we must realize the decompression algorithm of Flate to get the real information of PDF file. Because of time constraints, in this paper we only consider the conversion from the text of the Stream of PDF documents to the PS format.

Text objects are composed of one or more character string, which can be placed in any place of page and by any orientation, the same with path objects, they can be used to outline, fill the region or as the cutting path. Text objects are composed of designed character string, the motion of current point and the instruction character of the text status, it begins with BT, ends with ET. The format of text object is as follows:

<text object> ::= BT

<text operator or graphics state operator> *

ET

B. The graphical interface for software users(GUI)

The software graphical interface is divided into two parts: the first part is for the user to choose the PDF file need to be converted; the second part provides users the state of the document conversion.

VI. CONCLUSION

Based on the detailed analysis of the PDF document and the PS file format, we design and preliminary realize the corresponding conversion mechanisms. But to build a stable and efficient conversion software, we need to consider all kinds of objects and various factors. The functions of lexical analyzer and grammar analyzer are not strong enough, when it encounters an error or an unknown format, if it can not skip the mistakes and continue to analysis, there will be the unknown error, such as crashes and so on. The main flaw in the software is that it can only be used in the conversion from text-flow PDF document to PS format, how we can make the application of the software aboard, such as the consideration of the conversion of the graphics file, the conversion of the annotation and the hyperlinks in the PDF document and the conversion of some irregular PDF documents. These issues need further study in the future.

REFERENCES

[1] Adobe System Incorporated, PostScript Language Reference, Third Edition, Inc., 1999.

[2] Adobe Systems Incorporated, Portable Document Format Reference Manual, Version 1.3, 1999.

[3] Fupei xu, zhigeng pan and etc. Page description language and its programming, nanjing：Nanjing University Press, 1994.

[4] Linjun kong，haigen yao. Page description language and format, Shanghai Science and Technology Publishing Press，2004.

[5] Daoliang yang and etc., the Design and Implementation of Object-oriented Chinese PDF reader. Computer Application，Vol 19 No. 6，1999.6.

[6] Addison Wesley, PostScript Language Reference[B], 3rd Edition. 08 March, 1999, Pearson Educational.

[7] Addison Wesley, Thinking in PostScript.pdf.

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 035-038

The Design of Ethernet Controller Interface Circuit Based on ENC28J60

Bin Tan, Bo Yuan, and Bing Zhu
Department of Computer Jingganshan University, Ji'an,Jiangxi, China
Email:jatanbin@163.com

***Abstract*—ENC28J60 is a standard serial peripheral interface (Serial Peripheral Interface, SPI) independent Ethernet controllers. This paper-based controller designed for use LPC2132, using ENC28J60 for the Ethernet controller, through the SPI interface with high-speed communication between LPC2132; the master to write TCP / IP protocol stack, set up TCP-based applications programs to achieve the Ethernet communication interface, is given ENC28J60 application circuit connected with the LPC2132. By combining the communication interface applied to river sediment monitoring system, allowing monitoring of the terminal can receive the LAN on the remote host sent the information and perform the appropriate action.**

***Index Terms*—ENC28J60, Ethernet, Monitoring System**

I. INTRODUCTION

As computer networks and embedded Internet technology rapid development of embedded systems in industrial production and daily life have been widely used. Embedded real-time operating system and dedicated hardware structure of Internet users around as long as you can at any time, any place using the system remote monitoring and control of embedded devices.

How to use the computer network and embedded Internet technology, the size is very small, low cost embedded devices embedded into existing application systems to improve efficiency of their work in front of us need to be resolved.

II. ENC28J60 PERFORMANCE ANALYSIS

Ethernet as a cheap, efficient Internet access method, has been very widely used. The TCP / IP protocol stack is embedded into the microcontroller, using Ethernet controller MCU access to the Internet and data transmission. The client browser, the user can browse these embedded devices, dynamic information, and also can be achieved with the system's interactive features. Traditional embedded system development, the choice of an independent Ethernet controllers, such as the RTL8019, AX88796L, DM9008, CS8900A, LAN91C111 such as are for the personal computer systems designed. These devices are not only complex, bulky and high cost. ENC28J60 is a Microchip Technology has introduced a 28-pin stand-alone Ethernet controller. Ethernet controller on the market most of the package are more than 80 pins, while the IEEE 802.3 protocol ENC28J60 line with only 28 pins, both provide the corresponding functionality, but also greatly simplify the related design, reducing the space. In addition to technical advantages, for some there are restrictions on the size of small embedded devices, the use of ENC28J60 is the best option.

ENC28J60 is composed of seven main functional modules:

1.SPI interface: to serve as the primary controller and the communication channel between ENC28J60.

2.Control register: used to control and monitor ENC28J60.

3. Dual-port RAM buffer: for receiving and sending data packet.

4. Preferred device: When the DMA request transmit and receive modules, the right to control access to buffer RAM.

5.Bus Interface: SPI receive through the data and command parsing.

6.MAC (Medium Access Control)modules : implementation in line with IEEE802.3 standard MAC logic.

7.PHY (physical layer) modules: the simulation of the twisted-pair data on the encoding and decoding.

ENC28J60 working voltage is 3.3V, easily integrated into 5V systems, SPI of the CS, SCK, and SI input and RESET pin can withstand 5V voltage. When the SPI and interrupt inputs from the ENC28J60 on a CMOS output driver, through a one-way level converter. LEDA and LEDB pins to support the automatic reset when the polarity detection, in these two pins to pull current lit LED. When a system reset, ENC28J60 will detect the connection LED.

ENC28J60 all the memories are a way that a static RAM, the memory type of Ethernet buffer, respectively, control registers and PHY registers. ENC28J60 control registers used for configuration, control and status access, through the SPI interface to read and write these control registers directly. Ethernet buffer contains an Ethernet controller for use to send and receive storage space. Host controller can use the SPI interface, the capacity of the storage space program. PHY registers used for PHY module configuration, control and status of acquisition, not through the SPI interface to directly access these registers. Access Ethernet network module using the ENC28J60 program compared to other programs, the module is extremely streamlined. Microcontrollers that do not open the bus, although the possibility of using simulation of a parallel bus connections to other Ethernet controllers, but no matter from an efficiency or performance, not as good as with the SPI interface, or

using general-purpose I / O port simulation SPI interface ENC28J60 to program.

Ⅲ. LPC2131 Performance Analysis

LPC2131 is based on a real-time emulation and embedded trace support the 32/16 bit ARM7TDMI-STM CPU micro-controller, and with a 32KB, 64KB, 512KB of embedded high-speed Flash memory. 128-bit wide memory interface and a unique structure makes the acceleration of 32-bit code can run at maximum clock rate. Size of the code can be used to strictly control the application of 16-bit Thumb mode reduces code size of more than 30%, while the performance loss was small. A smaller package and lower power consumption makes the LPC2131 can be ideal for small-scale system; a wide range of serial communication interface to LPC2131 is ideal for communications gateways, protocol converters and low-end imaging, providing them with adequate buffer space and processing power. A LPC2131 8 channel 10-bit A / D converter contains a total of 16 analog inputs, each channel conversion time as long as 2.44us; a 10-bit D / A converter can provide different analog output; with two 32-bit timer devices, counters, PWM unit and watchdog; real-time clock with independent power and clock source, power-saving mode, significantly reducing power consumption; the same time with multiple serial interfaces, including two 16C550 industry standard UART, 2 high-speed I2C interface and SPI and SSP; vector interrupt controller can be configured to provide priorities and vector addresses;LPC2131 as many as 47 5v universal I / O ports, nine edge or level triggered external interrupt pin; through On-chip PLL can be realized up to 60Mhz of CPU operating frequency, PLL stability time of 100us; can be individually enables / disables the external features and reduce the external clock to optimize power consumption; through an external interrupt to wake up the processor from Power-down mode ; a single power supply, containing the power-on reset and brown-out detection circuit.

Each memory group LPC2131 memory map has a "physical" position, which is an address range, which can be written within the program code, the capacity of each memory space are permanently fixed in the same location, so you do not need the code designed to operate within a different address range. As the ARM7 processor interrupt vector location, Boot Block and SRAM space for a small part of the need to re-mapping to achieve different operating modes of the interruption of use. Interrupt the re-mapping through the memory-mapped control features to achieve.

Ⅳ. Experimental program

A. Hardware Circuit Connection

This designed system information terminal module hardware system block diagram shown in Figure 1. In the diagram, the LPC2132 microcontroller to achieve the internal TCP / IP protocol stack, and drive the network interface controller ENC28J60, coupled isolation transformer and Ethernet Internet access; the same time, MCU is the corresponding control system control unit. ENC28J60 integrated media access control sub-layer and physical layer functions, primarily the completion of the network interface to send and receive data frames, ENC28J60 hardware design includes reset circuit, clock oscillator, network transformers, and other external components and LED configuration unit modules. Serial level conversion module uses +5 V power supply, achieving TTL level to the RS232C level conversion.

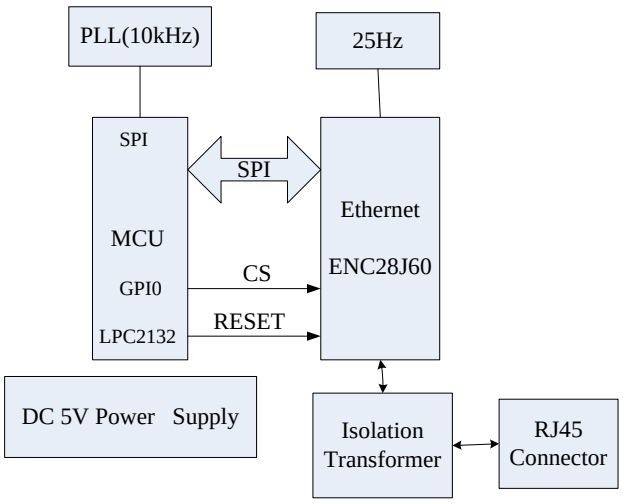


Figure 1. System block diagram of the hardware system information terminal module

As ENC28J60 working voltage is 3.3V, so integrated into the 5V voltage system, make a level converter used in this article 74HCT08 chip to achieve voltage conversion. Shown in Figure 2. Differential receiver pin as well as the differential sends pins and integration with a center-tapped transformer RJ45 socket HR901170A connection, ENC28J60 some of the digital power work in the 25V nominal voltage, this voltage through its internal regulator produces a 25V.

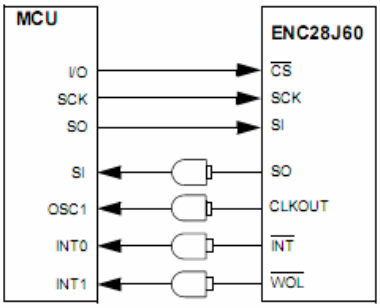


Figure 2. Level Converter

The core system software through the Ethernet controller ENC28J60 to TCP / IP protocol stack LwIP the protocol and Ethernet connected to the system to complete the following main aspects of the setup and implementation:

1. SPI Interface: ENC28J60 connection with the micro-controller MCU through the SPI interface to

support 10 Mb / s. But ENC28J60 only supports SPI mode 00, ie CPOL = 0 and CPHA = 0. In addition, SPI port requirements for the idle state SCK is low and does not support the clock polarity selection.

2. TCP / IP protocol stack to streamline and achieved: taking into account the MCU RAM storage space constraints and specific application requirements, reasonable to simplify TCP / IP protocol stack, select the realization of ARP, IP, ICMP, TCP, and HTTP protocols.

3.Physical layer implementation of the interface: the physical layer interface for the ENC28J60 the driver, the hardware is a circuit ENC28J60 the physical cable link, medium access control, frame transmit and receive functions, but shall be driven with the upper protocol interface procedures, to achieve to send and receive data frames.

4. The network layer protocol implementation: network-layer protocols, including ARP, IP, ICMP protocols.

Know both the network connection between the physical address, while the ARP protocol is address resolution protocol to achieve IP address to physical address conversion. Internet Protocol (IP) is responsible for the message sent from one host to another host. ICMP protocol is Internet Control Message Protocol for IP hosts, passing control messages between routers. Control of news is that the network is clear, host is up, routing is available and so the network itself the message.

B. Ethernet Buffer

Ethernet buffer contains the Ethernet controller for use to send and receive memory, the buffer size is 8KB, is divided into separate receive and send buffer space. Host controller using the SPI interface can send and receive capacity and location of storage of their programming. Receive buffer consists of a hardware management cycle FIFO buffer composition. Register for ERXSTH: ERXSTL and ERXNDH: ERXNDL as a pointer, the definition of buffer capacity and its memory location. When you accept the logic is disabled, the main controller can be programmed ERXST and ERXND pointers. When the receiver logic is used, not allowed to modify the pointer. 8KB memory has not been set up to receive FIFO buffer space, can be used as send buffer. Main controller is responsible for managing the data packets sent within the storage buffer. When the master controller decided to send a packet, ERXST and ERXND the pointer will be programmed to point to send data packets to be sent within a buffer address. Hardware does not check whether the start and end addresses overlap with the receive buffer. Main controller through separate read and write pointers (ERDPT and EWRPT) as well as read, write buffer SPI commands to access the contents of the Ethernet buffer.

Receive buffer free space calculation method：

If ERXWRPT>ERXRDPT, then

Freespace=(ERXND-ERXST)-(ERXWRPT-ERXRDPT)

If ERXWRPT=ERXRDPT, then

Freespace=(ERXND-ERXST)

Else

Freespace=ERXRDPT-ERXWRPT-1

When the order reads receive buffer when the receive buffer memory in the read data will be folded back to the bottom of the first part to continue reading. In calculating the checksum, the integrated DMA controller to read buffer; when the copy storage space, you must read and write buffer. DMA follow the rules of SPI access to the same turn back. The remote PC, transfer data packets over the Internet for twisted-pair physical medium to reach the local embedded systems, first of all through the isolation transformer coupled into the Ethernet control chip ENC28J60 in, ENC28J60 embedded in the physical layer and MAC layer to perform the address resolution protocol, and then to the LPC2132 microcontroller, LPC2132 implementation of the streamlining of TCP / IP protocol stack, remove the TCP / IP header, get useful data, and then embedded devices to achieve remote control and monitoring of the Internet.

C. Duplex Mode Configuration

ENC28J60 does not support duplex mode adaptive (auto-negotitation). If ENC28J60 enable duplex mode to connect to the Ethernet controller adaptive, then it will be detected as a half-duplex devices. This set is a full-duplex mode communications. When MACON3.FULDPX = 1 and PHCON.1PDPXMD = 1 Shi, ENC28J60 work in full-duplex mode. If these two bits, only one bit is cleared, then the ENC28J60 the working state of uncertainty and can not work properly. Master in the mode switching device does not send any data packets to maintain ECON1.TXRTS cleared. In full-duplex mode, you can also receive and send data packets. In this case, when sending data packets will not occur any conflict. In full-duplex mode (MACON3.FULDPX = 1), the hardware flow control is controlled by sending pause frames to achieve. Pause control frame is 64 bytes long, from the reserved multicast destination address 01-80-C2-00-00-01, sending node, the source address, a special suspension of operation code, two bytes of the suspension timer value, and fill / CRC bits.

D. In the river sediment measurement

With the modernization of China's development of hydrology and water resources information collection, transmission, processing and prediction of intelligence have been gradually speed up the construction progress of the hydrological work flow measurement, measuring sand, measuring rainfall is heavy throughout the hydrological work in the weight, the accuracy of these monitoring data for China's flood control and drought and water management conservation significance. The use of the existing hydrological stations in river hydrology cableway will we have designed an underwater camera equipment sunk a pre-set vertical position, when the bank issued the workstation computer, camera control commands, the underwater camera to camera began to water samples, and the image transmitted to the image acquisition card, and then will be collected by the ARM image data transmission to the bank workstations. Riparian workstation computer for various pieces of image denoising, segmentation, calculation and processing, and thus arrive at the river sediment content.

On ENC28J60 Embedded Ethernet Controller for the remote control within local area network has played a very good role, while the embedded device small, easy to install and operate. In the future work, also can be widely used in other industrial control applications.

REFERENCES

[1] Microchip Technology Inc. ENC28J60 Stand Alone Ethernet Controllerwith SPI Interface[EB/OL]. [2004 - 10 - 20].http: //www. microchip. com.

[2] RAJENDRA Y, PRASHANT P, RAPHAEL F. A reservation-based CSMA protocol for integrated manufacturing net-works[J]. IEEE Transactions on Systems, Man, and Cybemetics,1994,24 8 :1247 - 1258.

[3] PC/104 Embedded Consortium. PC/104 Specification Version 2. 5[M]. San Francisco: PC/104 Embedded Consortium,2003.

[4] Yang Jian, Song Zhanwei,“R&D of Embedded Web-based Bridge and Monitoring Unit ”,

[5] Shengwei,W;JunIong,X. Integrating Building Management System and facilities management on the Internet[J].Automation in Construction,2002,11(6):707-715.

[6] Microchip,Technology,Inc.ENC28J60 Data Sheet.2006:

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 039-042

Security Research on WEP of WLAN

Peisong Ye[1], and Guangxue Yue[2]
[1] College of Mathematics and Information engineering, JiaXing University, ZheJiang JiaXing, China
Email: yepeisong@sina.com
[2] College of Mathematics and Information engineering, JiaXing University, ZheJiang JiaXing, China
Email: guangxueyue@163.com

***Abstract*—Wires Equivalent Privacy (WEP) is used to improve the security of wireless LAN (WLAN). By analyzing the weaknesses of WEP and RC4, we design a simulative platform including software and hardware to crack WEP keys. The results show that the WLAN based on WEP is insecure. At last we give some improvements to enhance the WLAN security.**

***Index Terms*—WLAN, Security, 802.11, WEP, RC4**

I. INTRODUCTION

Because of the convenience of Wireless LAN (WLAN), it develops quickly. But the security of the WLAN becomes more important at the same time. Compared to wire LAN hacker can break into WLAN more easily because wireless data with electromagnetic wave are transmitted on air. Although WLAN 802.11b protocol provides some security mechanisms, they have some weaknesses and hacker can attack WLAN easily by making use of these weaknesses.

In section II we briefly introduce some basic concepts, system structure, some basic security knowledge of WLAN and some methods of attacking WLAN used by hackers. The security mechanism of WLAN 802.11b protocol includes: Service Set Identifier Association (SSID Association), MAC filtering and Wired Equivalent Privacy (WEP). Attacking methods include: SSID spoofing, MAC spoofing and authentication spoofing of WEP. In section Ⅲ we analyze the weaknesses of WEP and RC4 algorithm. In section Ⅳ we crack the WEP key by making use of the weakness of WEP. And we also design a scenario of cracking WEP key by using this weakness. In section Ⅴ we introduce some methods of WLAN security enhancement. These methods including: WEP improvement, Message Integrity Check and 802.1X/EAP.

II. THE SUMMARIZATION OF WLAN SECURITY

A. Basic Concepts of WLAN

WLAN is the outcome of combining computer net and wireless communication. WLAN expands the wired LAN. The mobility of connection is the most merit of WLAN, and it also overcomes the restriction of wires. Users can acquire information anywhere at any moment in the WLAN environment.

The range of WLAN is about 10 to 100 meters. It typically uses Spread Spectrum modulation technologies. There are two forms of SS (Spread Spectrum) schemes: FHSS (Frequency Hopping Spread Spectrum) and DSSS (Direct Sequence Spread Spectrum).

There are two types of WLAN: infrastructure network and Ad-hoc network. Infrastructure network needs an AP (Access Point), which is used as a bridge between clients and wired LAN. Ad-hoc network is comprised of some coordinative clients, which communicate with each other. We will discuss infrastructure network in this paper.

IEEE 802.11[1] is WLAN's standard protocol. Its scope is limited to physical and MAC layer. There are 802.11a, 802.11b and 802.11g protocols. 802.11b protocol is used most widely now. The bandwidth of 802.11b's physical layer is 2.4G Hz. It uses DSSS technology and its speed is 11Mb/s.

B. Basic Knowledge of WLAN Security

WLAN's data are transmitted in the air in the form of electromagnetic wave, so its security is more important. 802.11b provides some security mechanisms: Service Set Identifier Association (SSID Association), MAC filtering and Wired Equivalent Privacy (WEP)[2].

AP will allow a client access the net if the client provides a right and same SSID with AP. If the client' SSID is different with the AP' SSID, the AP will refuse the client to use the net. So the SSID can be seen as a simple password for providing an authentication mechanism. It can provide some security.

The mean of MAC filtering is that the administrator maintains a list of MAC addresses which some of them are admitted to the WLAN and others are not. This method can filter the physical address.

WEP is used to protect the data in WLAN. WEP uses RSA's RC4 symmetry encryption algorithm to encrypt the data. Another function of WEP is also to prevent unauthorized access to WLAN.

Although 802.11b provides three security mechanisms, they are insecure. In the next section we will discuss that how hackers attack the WLAN by making use of these mechanisms.

C. Methods of Attacking WLAN

Reconnaissance is the first thing hackers have to do before attacking the WLAN. Hackers drive around in a car equipped with wireless gears looking for unsecured WLAN to break in. They have to prepare some equipment to finish this work. For example: Laptop or PDA, 802.11b wireless card, antenna, GPS receiver and vehicle.

Hackers can sniff the WLAN after preparing these works. Because WLAN's data are transmitted in the air, Hackers can easily get data they want. Such as: AP name, SSID and MAC address[3]. They can break into the WLAN by using these information.

(1) SSID Spoofing

Because SSID is shared between AP and clients, hackers can get SSID easily. This becomes a kind of method of attacking WLAN. Because AP broadcasts data to all clients, they can find AP name, channel, SSID and MAC address in these data. So Hackers can listen these data and get them easily. Hackers can attack WLAN by using SSID association. Netstumbler software attack WLAN by using this theory too.

Sometimes the AP use default SSID value to communicate with the clients. Then hackers can get the SSID from the vendors more easily.

(2) MAC Spoofing

As mentioned before, MAC filtering is a kind of security mechanism of WLAN, but it is insecure. Hackers can attack WLAN by using this. MAC filtering provides security with the way only allowed valid MAC addresses get into the WLAN. These can be set in the AP. Some MAC addresses can be allowed to use the WLAN and others can't be allowed to use the WLAN. But if a hacker get these messages, he can change his MAC address to a valid MAC address to break into the WLAN. They can use some crack software to gain MAC addresses easily.

(3) Authentication Spoofing of WEP

If clients want to communicate with AP, they have to pass the WEP authentication. The process of WEP authentication as follow: AP sends a 128-byte random string to the client wishing to authenticate. The client then encrypts the message to the AP. If the message matches then the AP will allow the client to use the WLAN, otherwise refuse. In this process, hackers can acquire both the original plaintext and cipher messages at the same time. So they can create a fake authentication message that AP will authenticate a valid user. Sometimes hackers can acquire the WEP key from windows register, so the attacking will become more easily. Because the WEP key is shared between all communication stations.

WEP algorithm implements on the basis of RSA's RC4 encryption algorithm, but RC4 encryption algorithm has some weaknesses. So hackers can crack WEP key by using these weaknesses. We will discuss it in the next section.

III. The Analyzing of WEP Algorithm

Until now we have know that the security of WLAN is mostly depended on the security of WEP key. But as mentioned before, WEP uses RC4 encryption algorithm. We can crack WEP key by using RC4's weaknesses.

A. WEP Algorithm

WEP encryption process as follow: The data frame is check summed (using the CRC-32) to obtain c(M), where M is the message. M and c(M) are concatenated to get the plain text P=(M, c(M)). RC4 key stream is generated by a function of the initialization vector IV and the secret key =RC4 (IV, K). The cipher text results from applying the XOR function to the plain text and the key stream. That is (M, c (M)) XOR RC4 (v, K). At last, cipher text and IV are transmitted in the forms of electromagnetic wave.

By WEP symmetrically cryptographic characteristic, decryption process is the reverse of the encryption process by using the same WEP key. The receiver uses IV and WEP Key to generate the key stream. We use XOR key stream with cipher text to recover the plaintext P1. The P1 is then split into two parts as message M1 and check sum C1. c(M1) is then computed and compared with C1, if matches we receive it ,or refuse.

B. The Weaknesses of WEP Algorithm

(1) Invariance Weakness

RC4 is a kind of stream key algorithm widely used. RC4 is composed of key schedule algorithm (KSA) and pseudo-random generation algorithm (PRGA). In KSA process the WEP key is changed to a state array s with hundreds of plus and swap operation. The process of PRGA generates a pseudo-random stream. This stream is used to encrypt the plaintext or decrypt the cipher text.

Some researches indicate that the RC4 algorithm is vulnerable in the aspect that every 256 keys or less produce one weak key. This is called invariance weakness[4]. These weak keys will result in the pseudo-random have the specific and recognizant prefix. Their relativity with the key will become low. The data that are encrypted with these weak keys will become breakable.

(2) IV Weakness

There is another weakness is repeat-used IV, which is called IV weakness[4]. The input key of RC4 is composed of 24bytes IV and 40 bytes WEP key. The IV is used to guarantee that the same plaintext will never generate the same cipher text. But many wireless cards reset the IV to 0 each time a card is initialized, and increment the IV by 1 with each packet. Because of the invariance of WEP key the data will be encrypted with the same key stream. Yet in WLAN the IV is transmitted with plaintext that will be acquired easily.

There is another situation of repeat-used IV. Because the length of IV is only 24 bytes, every 224 packets will generate one repeat IV. So we can probably finger out that how much time will generate one repeat IV. An AP sends 1500 byte packets at 11 Mbps, and the time is about 1500*8/(11*10^6)*2^24=18000 seconds, which is about 5hours.

Hackers can get the plaintext by using these repeat IVs. If two cipher texts that are encrypted with the same key stream XOR each other, it will generate the result of two plaintexts' XOR because of the function of XOR. If one plaintext is known, another plaintext that is encrypted with the same key stream will be acquired easily. So the function of RC4 encryption is avoided.

Hackers can crack WEP key by using these repeat IVs. One special formatted IV can be constructed. In this

condition the first byte output of pseudo-random sequence includes the key information. By observing a lot of this type of bytes we can recover the key bytes totally. The conformation of these special IVs is: X=SB+3[1]<B+3； X+SB+3[X]=B+3. S is the state array. B is the key byte to recover. So we can collect enough repeat-used IVs to crack WEP key.

FMS[5] cracking is a method that uses IV weakness. We will design a scenario to crack WEP key with this method.

Ⅳ. USING WEP WEAKNESSES TO CRACK KEY

A. *Cracking Scenario*

We can design a scenario after understanding the theory of WEP cracking.

There two cracking software for selecting. But we select AirSnort and abandon WEPCrack. There are two reasons. First, it can't work alone. That means you have to get another software gather all the wireless packets for cracking. Second, the packets that are collected by other software are incompatible with WEPCrack.

There are hardware such as: laptop, AP (Cisco 340 Series Access Point, Dlink DWL-1000 AP) and wireless card (Cisco 340 Series Client, Lucent/HP Client). Software such as: OS (Redhat 7.3 Linux, Kennel: Linux 2.4.18) and other application software.

One client is used to communicate with AP. UDP flooding is used to send data. Another client is used to simulate the hacker to listen the WLAN. AP is linked to LAN with wires. Figure 1 is the illustration of WEP cracking job.

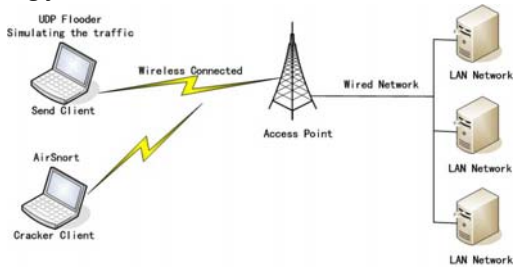


Figure 1 WEP cracking gears illustration

B. *Experimental Results*

The number of interesting packets capturing is normal distributed like as shown in Figure 2. At first, the capturing number will increase. Then, it will decrease by the time goes by.

When Cisco 340 Series Access Point as AP and Cisco 340 Series Client as client, it will generate 80 repeat IV Packets at most. And the time to acquire 80 repeat IV Packets is about 5 minutes. AirSnort can't crack WEP key in 13 to 97 hours.

When Dlink DWL-1000 AP as AP and Cisco 340 Series Client as client, it will generate 113 repeat IV Packets at most. AirSnort can crack WEP key in 13 to 97 hours if there are more than 80 repeat IV Packets. AirSnort can't crack WEP key in 13 to 97 hours if there are less than 97 repeat IV Packets.

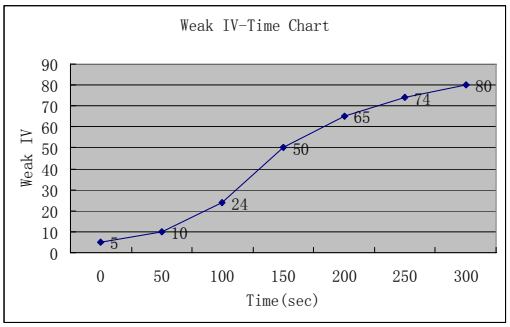


Figure 2 Weak IV to Time Relative Chart

When Dlink DWL-1000 AP as AP and Lucent/HP Client as client, it will generate 2733-41433 repeat IV Packets in 44-78 hours. And AirSnort can crack WEP key successfully every time.

The results of this experiment shows: First, the crack will become more easily if more weak IVs generates. Second, Cisco Company obviously enhances the security of their wireless devices.

Ⅴ. ENHANCEMENT OF WLAN SECURITY

Although there are many security problems in WLAN, more and more people are using WLAN because of its convenience. Many enhancement mechanisms of WLAN security are proposed.

A. *WEP Improvements*

Repeat-used key stream is the most weakness of WEP. The key stream is composed of WEP key and IV. Changing any one of them can solve this hazard. There are three WEP improvements: WEP key hashing, dynamic WEP keys and initialization Vector changes.

In WEP key hashing, the IV and WEP key can be hashed to produce a unique packet key for temporal. Then input it to the RC4. This method can avoid IV weakness attacking. Because we provide one only temporal key for one packet, the stream key will not repeat even if IV repeats.

Another insecurity of WLAN is that all clients share one WEP key. The whole WLAN will be insecure if the shared WEP key is cracked. So in dynamic WEP keys method, every client has a different key. It will not endanger the whole WLAN even if the key is cracked because the key is belonged to one client.

If every packet is assigned a different IV then repeat IVs will be avoided. When designing the wireless card a kind method of random IV generation can be used instead of the method of adding IV by one. So the IV value will not be forecasted. Or increasing the length of IV field, maybe 24 bytes or 64 bytes, then it will be hard to crack because the time of generating repeat IVs is increased.

B. *Message Integrity Check*

Another weakness of WEP is replay attack. The data in WEP is checked with CRC, and CRC is a linear function of plaintext variable. Hackers can create a fake

data to communicate with other client. We can use Message Integrity Check (MIC)[6] to compute the checksum. The MIC is based on destination MAC, source MAC, and payload. Any changes to these will affect the MIC value that will be checked out. This method can solve the replay attack.

C. 802.1X/EAP

IEEE 802.1X is a standard of controlling port access. Extensible Authentication Protocol (EAP), the extension of Remote Access Dial-In User (RADIUS), allows wireless client adapters to communicate with different back-end servers such as RADIUS.

The main function of 802.1X/EAP is that a client and an AP should be authenticated each other before communication. The process of authentication as follow:

(1) Wireless client associates with AP.

(2) AP rejects all user requests to WLAN and sends ask-for password message.

(3) Client inputs user and password.

(4) RADIUS server and client perform mutual authorization through AP.

(5) RADIUS server delivers the given key to AP.

(6) Client derives a unique WEP key from AP through radio transmission.

But the WEP key is transmitted with plaintext in this case. So we can use Lightweight Extensible Authentication Protocol (LEAP) to solve it. The biggest difference between EAP and LEAP is that we use a session key to encrypt the WEP key. This can solve the insecurity problem of transmitting the WEP key with plaintext.

LEAP provides two benefits: The first benefit is the mutual authentication between AP and client before communication. The second benefit it centralized the encryption WEP key management to enhance the security.

Ⅵ. CONCLUSIONS

Because of the characteristic of widely used WLAN its security problem also becomes more important.

802.11b protocol provides some security mechanisms such as: SSID association, MAC filtering and WEP authentication. But we know they are insecure by analyzing them.

The data, which are encrypted with WEP, are also insecure. WEP algorithm is based on the RC4, but we have proved that RC4 has weaknesses. And WEP algorithm's weakness is repeat-used IVs. Hackers can crack the WEP key with these two weaknesses.

To enhance the security of WLAN there are three mechanisms: WEP improvement, Message Integrity Check and 802.1X/EAP.

REFERENCES

[1] IEEE Standands Board. 802 part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY). specification. IEEE Standand 802.11, 1999 Edition

[2] WANG S M, TAO R, WANG Y. WLAN and Its Security Problems[R]. Proceeding of the Fourth International Conference on Parallel and Distributed Computing, 2003:241-244

[3] Christian Barnes, Tony Bautts. Hack Sproofing Your Wireless Network[M]. Syngress Press, 2002

[4] Scott Fluhrer, Itsik Mantin, and Adi Shamir. Weakness in the Key Scheduling Algorithm of RC4. Eight Annual Workshop on Selected Areas in Cryptography, August 2001

[5] Adam Stubblefield, John Ioannidis, Aviel D, Rubin, "Using the Fluhrer, Mantin, and, hamir Attack to Break WEP", 2001, AT&T Labs Technical Report TD-4ZCPZZ, pp.1-12

[6] Niels Ferugson, Macfergus Micheal. An Improved MIC for 802.11 WEP[C]. Document submitted to IEEE 802.11 WG, January 2000

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 043-046

Research on the Safety Assessment of Bridges Based on Fuzzy-Neural Network

Bo Wang[1], Xuzheng Liu[2], and Chao Luo[3]
[1] School of Information Science And Medium, JingGangShan University,Ji'an 343009, China
E-mail:woboxp@126.com
[2] School of Civil Engineering and Architecture, East China Jiaotong University, Nanchang 330013,China
Email:urbwolf@126.com
[3] Morden Education Technology Center, Jinggangshan UniverSity, Ji'an 343009, China
Email:luochao6668@163.com

***Abstract*--Fuzzy theory is integrated with Artificial Neural Network to create a bridge safety assessment model, through which the Fuzzy-Neural Network is improved in the light of sample data simulation. First, determine network layers interms of the seven critiria for bridge safety assessment. Then enter sample data at the input layer; study sample at the fuzzy reasoning layer by BP calculation method; obtain professional experience and ways of thinking about bridage safety assessment via the network. Finally, compare the assessment results from the network with those from professionals. The comparison proves the artificial fuzzy-neural network's feasibility and efficiency in assessing bridge safety.**

***Index Terms* --fuzzy-neural network; bridge safety assessment; BP network; suspended bridge**

I. INTRODUCTION

Artificial Neural Network, also known as Neural Network, is widely applied to pattern recognization, automatic control, image processing and language identification. Neural Network, integrated with fuzzy theory, is greatly enhanced to a better processing of information, information both precise and fuzzy; thus is the fuzzy system escalated to be known as adaptive fuzzy system. Efforts have been made to employ neural network to the bridge safety assessment in [1][2][3][4][5]. BP neural network is utilized to detect the structural damage by Wu[6][7] and the others[8][9][10]. Kaminski[11] has made a research into the examination of girder steel in the light of neural network. Kaminski's neural network has been verified through the tests of absolute frequency[12][13][14], relative frequency and the synthesized frequency to guarantee the solid feedback on the identification of the damage[15][16][17]. A comparative inadequacy can be seen when the application of fuzzy-neural network to the bridge safety assessment is put in the concerned domain world wide. The paper is therefore dedicated to a specific reliability bridge assessment resolutions based on the fuzzy-neural network and employed to Bridge A.

II. SAFETY ASSESSMENT MODEL

A. The Criteria for the Model

A set of criteria has to be established before the assessment can be carried out to a specific subject. As main cable sustains the most load of a suspended bridge, its strained condition and its lineshape is of significance to the entire bridge safety. The bridge tower is at always left at the bending moment and shaft force, which makes another inconvenient criterion the deviation of the tower-top and the capacity of the tower to the tension. The other influential criteria are the lineshape of the stiff girder, the internal force and the exterior examination of the lift lock. Now we come to criteria covers seven significant aspects to a feasible safety assessment model: exterior examination (F_1), bridge tower deviation (F_2), bridge tower capacity to tension (F_3), lineshape of the stiff girder (F_4), internal force of the lift lock (F_5), lineshape of main cable (F_6) and the tension of the anchor cable (F_7)

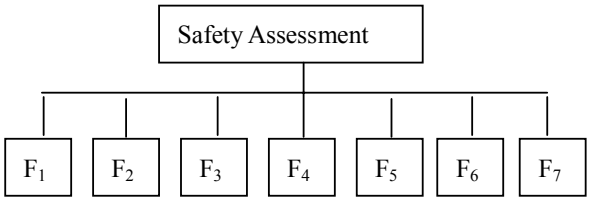


Figure 1. Safety assessment model

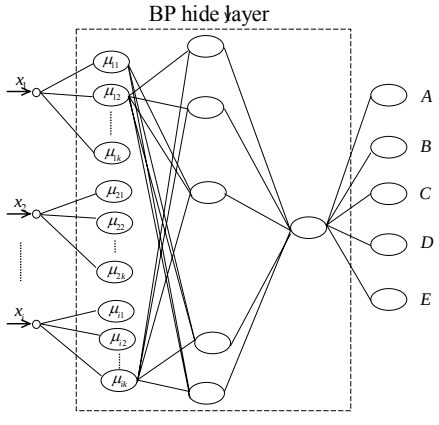


Figure 2. Fuzzy-neural network

This article is funded by Education Department,province Jiangxi,GJJ09143.

AP-PROC-CS-10CN006

III. FUZZY-NEURAL NETWORK MODEL

A fuzzy-neural network is established, in accordance with the requirements of a feasible safety assessment. The five layer network is illustrated as Fig1. The first layer is the Input layer, of which each node signifies one variable. The nodes amount to 7 on the Input layer. The second layer with altogether 35 nodes is the Obfuscation layer for the purpose of obfuscating the input variable. BP Inclusion layer is known as the third layer serves as the reflection between the input variable and the output variable. The 71 nodes are derived from the 35 nodes of the second layer in the light of Kolmogorov Complexity. The fourth layer is the Output layer, where the obfuscated results come forth. The fifth is the Defuzzification layer, where more precise coming out can be expected by applying the defuzzification principles. The five layers mentioned is characterized with its capacity of self-adjustment.

IV. AN ASSESSMENT INSTANCE

The safety assessment of Bridge A is rendered in the light of the above network as referred to. Bridge A has a length of 1187.489 m, with a clearspan of 960 m and a width of 30 m. As a double- tower suspended bridge, an assessment is to be carried out to this five-year-old.

A Sample learning

A well-distributed value is vital to a relatively thorough assessment. Twenty arrays of theoretical input samples by the auto computing will generate an output scored by the experts. A hierarchy will manifest the assessment where A, B, C, D, E different degrees is to be brought forth through a weight multiplied by a centesimal criteria.

The input-output is as TabI goes, of which the weight vector is ω = [0108, 01105, 01175, 01114, 01144, 01174, 01208]

B Network Training

The ten arrays is input in the neural-fuzzy network, where a satisfied result is generated with a discrepancy $E < 0.00001$. Two hundred and forty four learning processes confirm the parameter value. Thus is established the Fuzzy-neural network to the safety assessment of the large-span suspended bridge. In order to justify the feasibility, randomly generated five arrays of input samples yield the results in comparison with the experts' scoring in TabII.

TABLE I. Network training sample data

Sample	Exterior examination (F_1)	Tower displacement (F_2)	Tower stress (F_3)	Shape of the stiffened girder (F_4)	Cable stress (F_5)	Shape of main cable (F_6)	Tension of the anchored cable (F_7)	Safety assessments Result
1	100	0.05	0.7	0.65	0.01	0.75	0	A
2	100	0.02	0.6	0.65	0	0.8	0.01	A
3	95	0.1	0.7	0.7	0.02	0.75	0.01	A
4	90	0.1	0.65	0.8	0.025	0.85	0.02	A
5	85	0.1	0.85	0.9	0.03	1	0.02	B
6	88	0.12	0.9	0.82	0.028	0.95	0.03	B
7	85	0.15	0.75	0.95	0.04	1.05	0.035	B
8	81	0.22	0.95	1.05	0.055	1.12	0.048	B
9	72	0.28	0.9	1.1	0.065	1.13	0.055	C
10	75	0.3	1.1	1.15	0.07	1.1	0.052	C

TABLE.II Assessments results of verifying sample

Sample	F1	F2	F3	F4	F5	F6	F7	Result	Export
11	81	.0.2	0.95	1.3	0.02	1.05	0.028	B	[0.02,0.97,0.02,0,0]
12	83	0.09	0.81	0.75	0.033	0.89	0.022	B	[0.07,0.71,0,0,0.1]
13	91	0.06	0.77	0.81	0.022	0.79	0.021	A	[0.95,0.03,0,0,0]
14	70	0.35	1.15	1.12	0.075	1.35	0.074	D	[0,0,0,0.97,0]
15	35	0.85	1.55	1.61	0.131	1.59	0.122	E	[0,0,0,0,1]

According to maximum membership degree principle, the five samples yield the very similar results as the experts have given. Fuzzy-neural network, capable of storing the experts' concerned knowledge, is verified to be applied to the safety assessment of suspended bridge

A Assessment Result

.Four reliability observations provides an array of data concerned with the exterior look, the main cable elevation, main beam elevation, the internal force of main members and the stress. The array goes as X = [84, 0113, 0197, 0192, 01048, 0194, 01021]. The fuzzy-neural network gives a output as Y = [0, 0199, 0, 0, 0]. The assessment result suggest Bridge A is in good condition. An integrated assessment criterion is on the list of Regulations of Roads and Bridges, JTG H11 – 2004. Five degrees manifest different condition of roads and bridges. A, B, C, D, E suggest the condition going from excellent structure to these in need of overhaul. An 82 points is generated in the light of this criterion, which is rendered as B class, characterized by the fine structure, good vital facilities and qualified capacity to weight. The B class of Bridge A also is statistically likened to the results yielded by the fuzzy-neural network.

V. CONCLUSION

A model of safety assessment to suspended bridge is established when the seven criteria is induced as exterior examination, tower displacement, tower stress, shape of the stiffened girder, cable force, shape of main cable and the tension of anchored cable. The solid assessment procedures comes to be shaped based on the fuzzy-neural network, which is sufficient to overcome the inadequacy of the traditional assessment measures, as the lacking precise results, and the dependency on the experience of experts. The fuzzy-neural network yields a result suggesting Bridge A is in good condition. The result is in correspondence with that brought forth based on the criteria of Regulations of Roads and Bridges, JTG H11 – 2004. The fuzzy-neural network, with a representation of the knowledge and the intuitive insight, reduces the human interference with the results, leading to an objective yielding.

REFERENCES

[1]. Maru S，Nagpal A K. "Neural network for creep and shrinkage deflections in reinforced concrete frames," Computing in Civil Engineering. No. 4, Vol. 18, pp. 350-359, 2004,

[2]. Hopfield J J. "Neural networks and physical systems with emergent colletive computational abilities," Proc. of the National Academy of Science. U.S.A., 1982, 79: 2554-2558

[3]. Melham H G，Cheng Y. "Prediction of remaining service life of bridge decks using machine learning," Computing in Civil Engineering, No. 1,Vol. 17, pp. 1-9, 2003.

[4]. Bin Zou, Xiaoyu Liao and Yongnian Zeng, "An improved BP neural network based on evaluating and forecasting model of water quality in Second Songhua River of China," Chinese Journal of Geochemistry, No. 1,pp. 1, 2006

[5]. LIU Xu-zheng, HUANG Ping-ming and ZHANG Yong-jian, "Reasearch On Safety Assessment of Long-span Suspanded Bridges Based on Fuzzy-neural Network,"Journal of Zhengzhou Universty(Engineering Science), No.3, Vol. 28, pp. 48-51,sep 2007.

[6]. Jinwei Gao, Xueye Wang and Xiaobing Li, "Prediction of polyamide properties using quantum-chemical methods and BP artificial neural networks," Journal of Molecular Modeling, No. 4, pp. 513-520,2008

[7]. V.M. Kuz'kin, V.D. Opp.engeĭm and S.A. Pereselkov, "The sensitivity of monitoring by measuring the frequency shifts of the sound field interference pattern," Akusticheskiĭ Zhurnal, Vol. 54, No. 2, pp. 267–271, 2008.

[8]. Ko H，Arozullah M. "Background No.ise supp.ression for signal enhancement by No.velty filtering," IEEE Transactions on Aerospace and Electronic Systems, pp.102-113, 36, 2000.

[9]. HUANG Ping-ming , LIU Xu-zheng and ZHANGYong-jian, "Health Monitoring and Structure Assessment of the Yichang Yangtze River Highway Bridge," Journal of Jiangnan University (Natural Science Edition), No. 2, Vol. 7,pp. 211-215, Apr 2008.

[10]. SugeNo. M，Kang G T. "Structure Identification Fuzzy Model," Fuzzy Sets and Systems, 1988, 28 (1): 15-33

[11]. Xia P Q，Brownjohon J M W. "Bridges structural condition assessment using systematically validate finite-element model," Bridge Engineering, No. 5, Vol. 9,pp. 418-423, 2004.

[12]. LIU Xu-zheng, XU Sheng-lei and ZHANG Yong-jian, "Fuzzy safety assessment of Long-spanSuspended Bridges,"journal of Nanchang university (Engineering & techNo.logy), No. 1,Vol. 30, pp. 100-103, mar 2008

[13]. LIU Xu-zheng, HUANG Ping-ming and XU Han-zhang,"Analysis of parameters' sensitiveniess of cable-stayed bridges with single tower,"Journal of Chang'an University (Natural Science Edition),No. 6, Vol. 27, pp. 63-66, No.v 2007

[14]. LIU Xu-zheng, WANG Da and XU Han-zheng,"Prediction analysis of bridge structural response based on linear regression," Journal of Chang'an University(Natural Science Edition), No. 5, Vol. 29, pp.-76-80, Sept 2009.

[15]. LIU Xu-zheng, NIU Yan-wei and HUANG Ping-ming, "Mechanics characteristics of RC beam bridge strengthend with adding beams," Jourval of Chang'an University(Natural Sceince Edition),No. 4, Vol. 28, pp. 62-65,Jul 2008.

[16]. Kaminski P C. The app.roximate location of damage through the analysis of natual frequencies with artificial neural networks[J]. Journal of Process Mechanical Engineering, pp. 117-123, 209, 1995.

[17]. V. M. Kuz'kin, S. A. Pereselkov and E. A. Petnikov, "The possibility of reconstruction of two-dimensional random inhomogeneities in a shallow sea by frequency shifts of the spatial," Physics of Wave Phenomena, No. 1, pp. 4-51, 2008

[18]. B. G. Katsnel'son, J. Lynch and A. V. Tshoidze, "Space-frequency distribution of sound field intensity in the vicinity of the temperature front in shallow water," Acoustical Physics, Vol. 53, No. 5, pp. 695–702,2007.

[19]. Renders J M,Flasse S P. Hybrid methods using genetic algorithms for global optimization .IEEE Trans on System,Man,and Cybernetics-Part B:Cybernetics. 1996,262, 26(2) :243-258 .

[20]. Perttu Laurinen,Juha Roning. "An adaptive neural network model for predicting the post roughing mill temperature of steel slabs in the reheating furnace," Journal of Materials Processing Technology, 2005,168, 168 :423-430 .

[21]. Gumrah F, OZ B."The application of artificial neural networks for the prediction of water quality of polluted aquifer," .Water Air Soil Pollut, 2000,1191-4, 119(1-4) :275-291 .

[22]. Rao S V, Protopopesca V, Mann R C, Oblow E M.and Iyengar S S. "Learning Algorithms for Feedforward Networks Based on Finite Samples," IEEE Trans Neural Networks, 1996,74, 7(4) :926-939

[23]. Leong T K, Saratchandran P, Sundararajan N. “Real-time performance evaluation of the minimal radial basis function network for identification of time varying nonlinear systems,” Computers and Electrical Engineering, 2002,28, 28 (2) :103~117 .

[24]. Ugur Kesin “Genetic algorithmand artificial neural networkfor engine opti mization of efficiency and NOx emission,” .Fuel, 2004,83, 83 :885-895 .

[25]. Wang Zhengying, Shi Bingxin and Erdun Zhao,” Bandwidth -delay-constrained least-cost multicast routing based on heuristic genetic algorithm,” Computer Communications, 2001,24, 24 (7-8) :685-692

[26]. Jing Hua Hu, Hua Feng Kong, “The Research on Effective Video Scene Character Extraction Algorithm in Natural Scene Images,” The 2nd International Symposium on Digital Manufacturing,2009

[27]. Liansheng Liu, Jianfang Zhang, “Software Design for the Fault Diagnosis System of ATC Based on Neural Networks ,” Chinese Control and Decision Conference,2009.

[28]. ZHANG Xiaodong, ZHANG Peilin, “Research on the Fault Diagnosis of Gun Recoil Mechanism Based on Intelligent Algorithm,” The 7th International Symposium on Test and Measurement,2007.

[29]. Zhang Jianyu,Wang Ruilin, “Evaluation of Effectiveness of Integrated Individual Soldier Combat System Based on Combined-training Neural Network,” The 7th International Symposium on Test and Measurement,2007.

[30]. JIN Taobin,DING Ronglin, “The BP Neural Networks Model of Combinatorial Nonperiodic Defected Ground Structures,” The 7th International Symposium on Test and Measurement,2007.

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 047-049

Influences of Powder Content on the Piezomagnetic Effect of Rubber-based Soft Magnetic Composite Film

Daguo Jiang
College of Mathematic and Physical , Jinggangshan University , Ji'an, China
e-mail: jgsxy_jdg@sohu.com

***Abstract*—Composite film was preparated from the heat treated Fe73.5Cu1Nb3Si13.5B9amorphous powders as reinforced material and butyl rubber as matrix by means of molding to shape, and its piezomagnetic effect is also tested. The results showed that the impedance of composite film decreases with the morer powder content; the rangeability of impedance increases first and then decreases with increasing powder content.Piezomagnetic effect of omposite film is theoretically analyzed with resistance and capacitance in series model and the conclusions are consistent with the experimental results**

***Index Terms*—Fe73.5Cu1Nb3Si13.5B9; butyl rubber; composite film;piezomagnetic effect**

I. INTRODUCTION

Strain occurs and the material internal magnetization state changes when the ferromagnetic materials are under the action of mechanical force,which resulting in a significant change in impedance of the phenomenon called piezomagnetic effect. A variety of pressure magnetic sensors are developed based on piezomagnetic effect, including pressure magnetic non-contact torque sensors, amorphous piezomagnetic acceleration sensors piezomagnetic pressure sensors [1-2]. In the literatures [3-5], the skin effect was used for qualitative analyzing the piezomagnetic effect of rubber-based soft magnetic composite film. As composite film is large resistivity and low permeability, and composite film has excellent piezomagnetic effect under low frequency, so the skin effect depth under alternating current is bigger than the film thickness. In this paper, the soft magnetic composite films were prepared with the non-powder Fe73.5Cu1Nb3Si13.5B9 as composite phase and the butyl rubber as base, and the influence of powder content on the piezomagnetic effect was studied. Theoretically analysis with resistance and capacitance in series model is done, which showed the conclusions are consistent with the experimental results .

II. EXPERIMENTAL PRINCIPLES AND METHODS

First, 7.2 μm FeSiB amorphous powders is put into titanate (NDZ-105) THF solution for surface process; and then uniformly mix the powder with butyl rubber by different mass ratio, then add sulfur, paraffin and other additives for adequate mixing; put the mixture between the two copper foil papers, pressure forms;finally a composite film with 200 μm thickness is achieved after cured at 160 ℃ for 1 hour.

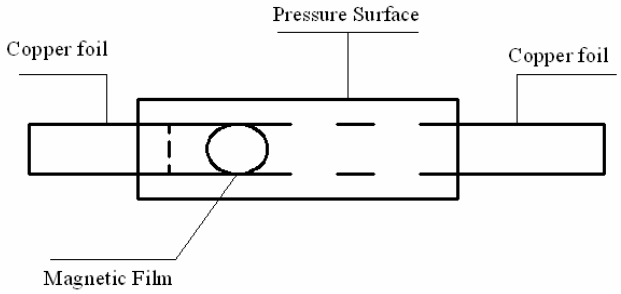


Figure1 Schematic of piezomagnetic effect test

The test method of piezomagnetic effect is shown in Figure 1 where the effective area in the film between the copper foil is 6mm × 5mm,then apply point pressure on the film center, pressure head diameter 5mm[6-8]. Two copper ends are connected with 4284A impedance analyzer, and the compressive stress and current are all along composite film thickness direction.

Impedance is tested with 4284A impedance analyzer, where current amplitude is 10mA, AC frequency is 0.6kHz and compressive stress is 0MPa to 1.182MPa.

Use the rangeability of impedance $|\Delta Z|$ to characterize the film's piezomagnetic properties.

$$|\Delta Z| = |Z_\sigma - Z_0| \qquad (1)$$

In the equation, Z_σ is for impedance value when the compressive stress is σ ; Z_0 is for impedance value when the compressive stress is 0.

III. EXPERIMENTAL RESULTS AND ANALYSIS

A. The equivalent circuit model of composite film piezomagnetic effects

The composite film sandwiched between two copper foils, current flows through along thickness direction of composite films sandwiched between two copper foils, and the compressive stress and current are all along composite film thickness direction.the copper foil is a conductor, and the composite films is an insulator with large resistivity, so the composite film and copper foils form a parallel-plate capacitor. So the system can be equivalent as a resistance and capacitance in series model.

B. *Influences of powder content on the impedance of composite film*

Table I is the influences of powder content on the impedance of composite film at 0.6kHz . Known from table, the impedance of composite film decreases with the powder content increases.

The impedance of composite film is consisting of resistance and reactance. Resistance is mainly depending on the degree of perfection (resistivity) of amorphous powder particle chains forming in the rubber base and the effective length of composite film that the current flows through;reactance of composite film is decided by the capacitance, which mainly depends on the thickness of composite film and the AC frequency.So the impedance of composite film is:

$$Z = \sqrt{R^2 + X^2} \qquad (2)$$

The powder content is defined as P ,the composite film resistivity and dielectric constant decrease with the bigger powder content P ,it means that,the resistance and resistance decreases with the bigger powder content P , so the impedance of composite film decreases along with the powder content P increases.

Table I Influences of powder content on the impedance of composite film

	Z/ kΩ			
σ/MPa	1:3	1:4	1:5	1:6
0	2159.39	1439.92	1397	938.626
0.065	2153.53	1425.87	1393.65	936.747
0.108	2153.24	1423.35	1393.31	936.32
0.147	2152.77	1422.75	1392.94	935.79
0.201	2152.72	1421.83	1392.52	935.294
0.397	2151.69	1420.89	1391.43	934.167
0.607	2150.72	1420.23	1390	933.391
0.802	2150.69	1417.21	1389.14	932.184
0.992	2150.04	1415.58	1388.55	931.797
1.182	2149.53	1412.13	1387.99	931.366

Figure 2 is the influences of powder content on the rangeability of impedance of composite film at 0.6kHz. Known from Figure, the rangeability of impedance increases first and then decreases with increasing powder content; it is the biggest When the powder with butyl rubber according to the mass ratio of 4 to 1.

The resistance of composite film R and the reactance of composite film X is inversely proportional to powder content P , so supposed $R = \frac{dC}{P}, X = \frac{dD}{P}$, where C 、 D are constants, and according to equation (2)and equation (1) , the impedance of composite film is:

$$Z = \frac{d\sqrt{C^2 + D^2}}{P} \qquad (3)$$

the rangeability of impedance of composite film is:

$$\Delta Z = \sqrt{C^2 + D^2}\,\frac{(d_0 - d_\sigma)}{P} \qquad (4)$$

The elastic modulus of composite film increased With the P increases ,and the deformation of composite film decreases under constant stress, it means that, d_σ decreases with the P increase; so supposed $d_\sigma = \frac{E}{P}$, where E is constant; d_0 is constant ,according to equation (4), the rangeability of impedance of composite film is :

$$\Delta Z = \sqrt{C^2 + D^2}\,\frac{(d_0 - \frac{E}{P})}{P} = \sqrt{C^2 + D^2}\,\frac{(d_0 P - E)}{P^2} \qquad (5)$$

$$\frac{d(\Delta Z)}{dP} = \sqrt{C^2 + D^2}\,\frac{(2PE - d_0 P^2)}{P^4}$$

When $P < \frac{2E}{d_0}$, $\frac{d(\Delta Z)}{dP} > 0$, ΔZ increases with powder content P increase; While $P > \frac{2E}{d_0}$, $\frac{d(\Delta Z)}{dP} < 0$, ΔZ decreases with powder content P increase.

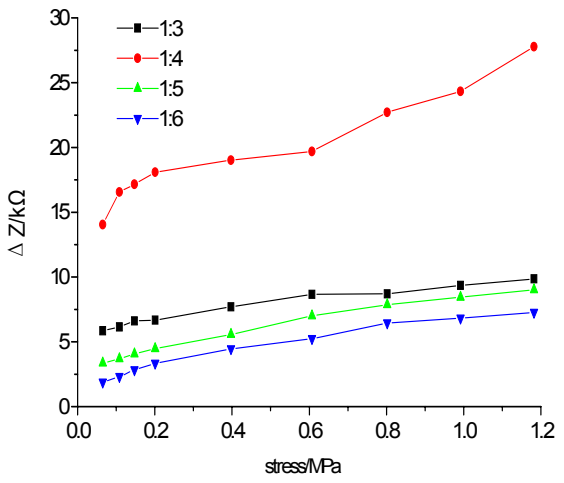


Figure 2 Influences of powder content on the rangeability of impedance of composite film

IV. CONCLUSIONS

Large piezomagnetic effect of Fe-based amorphous ribbon can be produced only at high frequency, while it can be produced at low frequency in soft magnetic composite films prepared with IIR as base and amorphous powders grinded from iron-based amorphous ribbon as composite phase.. the impedance of composite film decreases with the morer powder content; the rangeability of impedance increases first and then decreases with increasing powder content, the rangeability of impedance is the biggest When the powder with butyl

rubber according to the mass ratio of 4 to 1. From the view point of the sensors application, low-frequency signals circuit can be easier to achieved than high-frequency and the composite films has small permeability, magnetic fields insensitivity and the compressive stress sensitivity. So these flexible soft magnetic composite films are expected to become the new generation sensitive material for sensors.

REFERENCES

[1] H Chazal, O Geoffroy, J L Porteseil, and T Waeckerle, "Infuence of magnetoelastic effects on the coercivity of ultrasoft nanocrystalline alloys,"Journal of Magnetism and Magnetic Materials, vol.272, pp. 1436-1438, 2004

[2] J M Barandiaran, J Gutierrez ,and C Gomez-Polo, "New sensors based on the magnetoelastic resonance of metallic glasses," Sensors and Actuators, vol..272, pp. 154-157, 2004.

[3] MA Guang-bin, Zhu Zheng-hou,and Li Tang-hua., "Study on preparation and piezomagnetic property of FeCuNbSiB/silicone rubber magnetic compound films," Journal of Functional Materials, vol.11, pp. 737-739,May 2007

[4] Li Tang-hua,Zhu Zheng-hou, and Ma Guang-bin, "A new piezomagnetic film," Journal of Functional Materials, vol.38, pp. 1787-1789,November 2007

[5] Wan Zhen-zhen, Zhu Zheng-hou,and Li Tang-hua, "Study on piezomagnetic property of FeCuNbSiB /butyl rubber composite film. Journal of Functional Materials", vol . 39, pp. 1494-1496,September 2008

[6] Jiang Daguo,Yu Xiaoguang,and Zhu Zhenghou. "Influence of Annealing Process on the Stress Impedance of FeSiB Amorphous Alloy Ribbons," Rare Metal Materials and Engineering,2008.3: 440-443.

[7] Jiang Daguo,Zhu Zhenghou,and Ma Guang-bin, "Compression stress impedance properties of FeSiB amorphous alloy's ribbon,"Journal of Functional Materials, vol . 38, pp.911-913,June 2007.

[8] Ma Guang-bin,Zhu Zhenghou,and Huang Yu-hong, "Piezomagnetic properties of the FeCoSiB alloy ribbons," Journal of Functional Materials,vol. 38, pp.1071-1073,July 2007:

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 050-053

An Improved Molecular Solution for the Partition Problem

Xu Zhou[1], and ShuangShuang Huang[2]
College of Mathematics and Information Engineering, JiaXing University, JiaXing, China
Email: { Zhouxu2006@126.com, huangya_0611@163.com}

***Abstract*—Now the algorithms based on DNA computing for partition problem always converted the elements into binary number and then carry on mathematics operations on them. In this paper, our main purpose is to give an improved molecular solution for the partition problem. Our new algorithm does not need mathematics operations. So the algorithm is easier and the number of the operation is reduced. Besides, the time used in the biological experiment is less than before. In order to achieve this, we design a new encoding method. We also design a special parallel searcher to search the legal DNA strands in the solution space.**

***Index Terms*—DNA-based computing; NP-complete problem; partition problem**

Ⅰ. INTRODUCTION

Since Adleman worked out an instance of the Hamiltonian path problem in a test tube just by handling DNA strands [1], numerous researchers have explored efficient molecular algorithms for some other famous NP-complete problems: Lipton et al.[2] and Braich et al.[3] solved the 3-SAT problem, Faulhammer et al.[4] solved the knight problem Ouyang et al.[5] and Li et al.[6] solved the maximal clique problem and so on [8-14]. DNA computer's power of parallel, high density computation by molecules in solution allows itself to solve hard computational problems in polynomial increasing time, while a conventional turing machine needs exponentially increasing time. Through advances in molecular biology, it is now possible to produce roughly 10^{18} DNA strands in a test tube, that is to say 10^{18} bits of data can be processed in parallel by basic biological operations [15-16].

In this paper, we describe a new algorithm to solve the partition problem [17]. Since Huiqin's paradigm proposed in 2004 demonstrated the feasibility of applying DNA computer to tackle such an NP-complete problem. Instead of surveying all possible assignment sequences generated in the very beginning, we use the operations of Adleman-Lipton model and the solution space of sticker which is proposed by Chang et al., then apply a new DNA algorithm for partition problem [9,10].

The paper is organized as follows. Section 2 introduces the Chang et al.'s model and the definition of the partition problem. Section 3 introduces the DNA algorithm to solve the partition problem for the sticker solution space. In section 4, the experimental results by simulated DNA computing are given. Conclusions and future research work are drawn in Section 5.

Ⅱ. DNA MODEL OF COMPUTATION

A.The Chang et al.'s model

Chang et al. presented the model that took biological operations in the Adleman–Lipton model and the solution space of stickers in the sticker-based model [9,10,18].This model has several advantages from the Adleman–Lipton model and the sticker-based model:

1. The new model has finished all the basic mathematical functions and the number of tubes, the longest length of DNA library strands, the number of DNA library strands and the number of biological operations are polynomial.
2. The basic biological operations in the Adleman–Lipton model had been performed in a fully automated manner in their lab. The full automation manner is essential not only for the speedup of computation but also for error-free computation.
3. Chang and Guo also employed the sticker-based model and the Adleman–Lipton model for dealing with Cook's theorem, the dominating-set problem, the set-splitting problem and many other NP complete problems for decreasing the error rate of hybridization.

If given a tube, one can perform the following operations in Adleman–Lipton model: Extract; Merge, Detect, Discard, Amplify, Append and Read.

B. Definition of the partition problem

The partition problem is a classical NP-complete problem in computer science [17]. The problem is to decide whether a given multiset of integers can be partitioned into two "halves" which have the same sum. More precisely, given a multiset A of integers, is there a way to partition A into two subsets A_1 and A_2 such that the sums of the numbers in each subset are equal? The subsets A_1 and A_2 must form a partition in the sense that they are disjoint and they cover the set A.

On the assumption that a finite set A is $\{a_1, a_2 \ldots a_m\}$, where a_i is the ith element for $1 \leq i \leq m$. Also suppose that every element in A is a positive integer. Assume that $|A|$ is the number of element in A and $|A|$ is equal to m.The partition problem is to find a subset $S^1 \subseteq A$ such that the sum of all the elements in S^1 is half of S, where S is the sum of all the elements in A.

AP-PROC-CS-10CN006

III. AN IMPROVED DNA ALGORITHM FOR THE PARTITION PROBLEM

A. Sticker-based solution space for the partiton problem

This algorithm for construction the sticker-based solution space for subsets of a finite set is similar to the Michael's algorithm at 2005[13].

Being similar to the Knapsack Problem[19],we will construct the solution space for subsets of *A*. Assume that $x_1...x_n$ is an *n*-bit binary number, which is applied to represent *n* elements. Assume that the sum of all the elements in *A* is *S*.

Procedure Solution_Generator(T_0, *n*)
1: **For** *m* = 1 **to** *n*
2: *Amplify*(T_0, T_1, T_2).
3: *Append*(T_1, x^1_m).
4: *Append*(T_2, x^0_m).
5: $T_{01} = \cup$ (T_1, T_2).
6: **EndFor**
EndProcedure

From Solution_Generator(T_0, *n*), it takes *n amplify* operations, 2*n append* operations, *n merge* operations and three test tubes to construct sticker-based solution space. A *n*-bit binary number corresponds to an array of input. A value sequence for every bit contains 15 bases. Therefore, the length of a DNA strand, encoding a subset, is 15*n* bases consisting of the concatenation of one value sequence for each bit.

B. Sticker-based solution space for elements of subsets for a finite set

An element a_i, in *A* can be converted as a number $y_{i,1}$ $y_{i,2}\ldots y_{i,k}$, for $k = a_i$. Sticker is applied to represent every bit $y_{i,j}$ for $1 \le j \le k$ .For every bit $y_{i,j}$, two distinct DNA sequence were designed. One corresponds to the value "0" for $y_{i,j}$ and the other corresponds to the value "1" for $y_{i,j}$. For the sake of convenience in our presentation, assume that $y^1_{i,j}$ denotes the value of $y_{i,j}$ to be 1 and $y^0_{i,j}$ denotes the value of $y_{i,j}$ to be 0.The following algorithm is employed to construct sticker-based solution space for elements of 2^n possible subsets for a n-element set *A*. Assume that *S* is the sum of all the elements in *A*.

Procedure Make_Value(T_0, *n*)
1: **For** *i* = 1 **to** *n*
2: $T_1 = +(T_0, x_i^1)$ and $T_2 = -(T_0, x_i^1)$.
3: **If** (Detect(T_1)= 'yes') **then**
4: **For j = 1 to ai**
5: **Append(T1 , y1i, j).**
6: **EndFor**
7: **If** (Detect(T_2)= 'yes') **then**
8: **For** *j* = 1 **to** a_i
9: Append(T_2 , $y^0_{i,j}$).
10: **EndFor**
11: **EndIf**
12: $T_{01} = \cup$ (T_1, T_2).
13: **EndFor**
EndProcedure

From Make_Value(T_0, *n*), it takes *n extract* operations, *n detect* operations, *S append* operations, *n merge* operations, and three test tubes to construct sticker-based solution space for elements of 2^n possible subsets to a *n*-element set *A*. *n*-bit binary number corresponds to a subset and the longest bit of binary number, which are used to encode the values of all the elements in *A*, is *S*. Assume that a value sequence for every bit contains 15 bases. Therefore, the last length of a DNA strand, encoding elements of 2^n possible subsets to a *n*-element set *A*, is 15(*n*+*S*).

C. The DNA based algorithm to finish the Logic Multiplication operation functions

In order to find the legal strands of the partition problem, a parallel searcher is designed.

Procedure Parallel_Searcher(T_0, *S*/2)
1: **For** *i* = 0 **to** *S* -1
2: $t = \min\{S/2, i\}$.
3: **For** *j* = *t* **down to** 0
4: $T^{ON}_{j+1} = +(T_j, y^1_{i+1})$ and $T_j = -(T_j, y^1_{i+1})$.
5: $T_{j+1} = \cup(T_{j+1}, T^{ON}_{j+1})$.
6: **EndFor**
7: **EndFor**
8: **If** (Detect($T_{s/2+1}$) ="yes")
9: Discard($T_{s/2+1}$).
10: **EndIf**
EndProcedure

From Parallel_Searcher (T_{01}, *S*/2), it takes *S*(*S*+2)/8 exact operations, *S*(*S*+2)/8 merge operations, one detect operation, one discard operation and *S*/2+2 test tubes where *S* is the sum of all the element in the set A. The length of a DNA strand has never been changed, is 15(*n*+*S*).

D. An improved DNA based algorithm for the partition problem

Algorithm1. An improved DNA based algorithm for the partition problem

1: Solution_Generator(T_0, *n*).
2: Make_Value(T_0, *n*).
3: Parallel_Searcher(T_0, *S*/2).
4: **If** (Detect($T_{S/2}$) ="yes")
5: Read($T_{S/2}$).
6: **EndIf**

Theorem 1: From those steps in **Algorithm 1**, the improved DNA based algorithm for the partition problem can be solved.

Proof: On the execution of Line(1), it calls Solution_Generator(T_0, *n*). The algorithm, Solution_ Generator(T_0, *n*), is mainly used to construct solution space for 2^n possible subsets of an *n*-element in the set *A*. Line (2) calls Make_Value(T_0, *n*).The algorithm, Make_ Value(T_0, *n*) is employed to construct sticker-based solution space for element of 2^n possible subsets of an *n*-element in the set *A*. Line(3) is called Parallel_ Searcher(T_0, *S*/2). The algorithm, Parallel_ Searcher(T_0, *S*/2) is mainly used to search the legal DNA strands for the partition problem.

Theorem 2: If the number of the element in A is n and the sum of the elements in A is S in the partition problem can be solved with $O(n+(0.25S^2+2.5S))$ biological operations, $O(2^n)$ strands, the logest strands is $O(n+S)$ and $O(S)$ test tubes.

Proof:. Algorithm 1 includes four main steps. From the algorithm, Line(1), it is very obvious that it takes n amplify operations, $2n$ append operations, n merge operations and three test tubes to construct sticker-based solution space. From the algorithm, Line(2), it takes n extract operations, n detect operations, S append operations, n merge operations, and three test tubes to construct sticker-based solution space for elements of 2^n possible subsets to a n-element set A. From the algorithm, Parallel_Searcher (T_{01}, $S/2$), it takes $S(S+2)/8$ exact operations, $S(S+2)/8$ merge operations, one detect operation, one discard operation and $S/2+2$ test tubes. The length of a DNA strand has never been changed, is $15(n+S)$.In Line(4), it takes one detect operation and it takes one read operation in Line(5).

Hence, from the statements mentioned above, it is at once inferred that the time complexity of Algorithm 1 is $O(n+(0.25S^2+2.5S))$.After Line(1) of Algorithm1, the solution is constructed and there is 2^n strands in it. Referring to the Michael's algorithm [8, 9], the number of the tube used in Algorithm 1 is $O(n)$. The longest strand in Algorithm 1 is $O(n+S)$.

Ⅳ. EXPERIMENTAL RESULTS BY SIMULATED DNAS COMPUTING

Consider that a finite set A is {1, 2, 3 } as an example. S is denoted to represent the sum of the elements in A So S is 6. The partition problem is to find a subset of A in which the sum is one half of S. Table 1 shows the strands pool and every step in T_0.

After Line(4) in **Table 1**, we will get the full solution space of the partition problem above. The DNA strands are $\{x_1^0 x_2^0 x_3^0 y_{1,1}^0 y_{2,1}^0 y_{2,2}^0 y_{3,1}^0 y_{3,2}^0 y_{3,3}^0$; $x_1^0 x_2^0 x_3^1 y_{1,1}^0 y_{2,1}^0 y_{2,2}^0 y_{3,1}^1 y_{3,2}^1 y_{3,3}^1$; $x_1^0 x_2^1 x_3^0 y_{1,1}^1 y_{2,1}^1 y_{2,2}^0 y_{3,1}^0 y_{3,2}^1 y_{3,3}^1$; $x_1^0 x_2^1 x_3^1 y_{1,1}^1 y_{2,1}^1 y_{2,2}^1 y_{3,1}^0 y_{3,2}^0 y_{3,3}^0$; $x_1^1 x_2^0 x_3^0 y_{1,1}^1 y_{2,1}^0 y_{2,2}^0 y_{3,1}^1 y_{3,2}^0 y_{3,3}^1$; $x_1^1 x_2^0 x_3^1 y_{1,1}^1 y_{2,1}^0 y_{2,2}^0 y_{3,1}^0 y_{3,2}^0 y_{3,3}^0$; $x_1^1 x_2^1 x_3^0 y_{1,1}^1 y_{2,1}^1 y_{2,2}^0 y_{3,1}^1 y_{3,2}^1 y_{3,3}^1$; $x_1^1 x_2^1 x_3^1 y_{1,1}^1 y_{2,1}^1 y_{2,2}^1 y_{3,1}^0 y_{3,2}^0 y_{3,3}^0\}$. Searching the union of the DNA strands by the algorithm Parallel_Searcher(T_0, $S/2$) showing in **Table 2**, we will get the legal strands of the partition problem. Reading the DNA sequence in tube T_3, we could get the answer is {1, 2} and {3}.

Table 1 Every steps of the improved algorithm based on DNA computing

tube	T_0
1	$\{ x_1^0x_2^0x_3^0; x_1^0x_2^0x_3^1; x_1^0x_2^1x_3^0; x_1^0x_2^1x_3^1; x_1^1x_2^0x_3^0; x_1^1x_2^0x_3^1; x_1^1x_2^1x_3^0; x_1^1x_2^1x_3^1$. $\}$
2	$\{ x_1^0x_2^0x_3^0y_{1,1}^0; x_1^0x_2^0x_3^1y_{1,1}^0; x_1^0x_2^1x_3^0y_{1,1}^0; x_1^0x_2^1x_3^1y_{1,1}^0; x_1^1x_2^0x_3^0y_{1,1}^0; x_1^1x_2^0x_3^1y_{1,1}^0; x_1^1x_2^1x_3^0y_{1,1}^0; x_1^1x_2^1x_3^1y_{1,1}^0$. $\}$
3	$\{ x_1^0x_2^0x_3^0y_{1,1}^0y_{2,1}^0y_{2,2}^0; x_1^0x_2^0x_3^1y_{1,1}^0y_{2,1}^0y_{2,2}^0; x_1^0x_2^1x_3^0y_{1,1}^0y_{2,1}^1y_{2,2}^1; x_1^0x_2^1x_3^1y_{1,1}^0y_{2,1}^1y_{2,2}^1; x_1^1x_2^0x_3^0y_{1,1}^0y_{2,1}^0y_{2,2}^0; x_1^1x_2^0x_3^1y_{1,1}^0y_{2,1}^0y_{2,2}^0; x_1^1x_2^1x_3^0y_{1,1}^0y_{2,1}^1y_{2,2}^1; x_1^1x_2^1x_3^1y_{1,1}^0y_{2,1}^1y_{2,2}^1$. $\}$
4	$\{x_1^0 x_2^0 x_3^0 y_{1,1}^0 y_{2,1}^0 y_{2,2}^0 y_{3,1}^0y_{3,2}^0y_{3,3}^0$; $x_1^0x_2^0x_3^1y_{1,1}^0y_{2,1}^0y_{2,2}^0y_{3,1}^1y_{3,2}^1y_{3,3}^1$; $x_1^0x_2^1x_3^0y_{1,1}^0y_{2,1}^1y_{2,2}^1 y_{3,1}^0 y_{3,2}^0 y_{3,3}^0$; $x_1^0x_2^1x_3^1y_{1,1}^0y_{2,1}^1y_{2,2}^1y_{3,1}^1y_{3,2}^1y_{3,3}^1$; $x_1^1x_2^0x_3^0y_{1,1}^1y_{2,1}^0y_{2,2}^0y_{3,1}^0y_{3,2}^0$, $y_{3,3}^0$; $x_1^1x_2^0x_3^1y_{1,1}^1y_{2,1}^0y_{2,2}^0y_{3,1}^1y_{3,2}^1y_{3,3}^1$; $x_1^1x_2^1x_3^0y_{1,1}^1y_{2,1}^1y_{2,2}^1y_{3,1}^0y_{3,2}^0y_{3,3}^0$; $x_1^1x_2^1x_3^1y_{1,1}^1y_{2,1}^1y_{2,2}^1y_{3,1}^1y_{3,2}^1y_{3,3}^1$. $\}$

Table 2 Search the solution space of the element

Tube	T_0	T_1	T_2	T_3	T_4
Inital pool	000000;000111 011000;011111 100000;100111 111000;111111				
1	000000;000111 011000;011111	100000;100111 ; 111000; 111111			
2	000000;000111	100000;100111 ; 011000;011111	111000; 111111		
3	000000;000111	100000;100111	011000; 011111	111000; 111111	
4	000000	100000;000111	011000;100111 111000	111000; 011111	111111
5		100000	011000; 000111	111000; 100111	011111;100111 ; 111111
6			011000	000111; 111000	

Ⅴ.CONCLUSION REMARKS

DNA computer's power of parallel, high density computation by molecules in solution allows itself to solve hard computational problems such as NP-complete problems in polynomial increasing time. In this paper, we introduce an improved DNA algorithm for the partition problem, which is a classical NP-complete problem. Our algorithm could solve the partition problem effectivily and easily.

Now, it is also very hard to say that whether the molecular computers will have a bright future. In the future molecular computers may be the clear and good choice for performing massively parallel computations. To reach a free stage in using DNA computers, just as using classical digital computers, there are still many technical difficulties to overcome before it becomes real.

ACKNOWLEDGEMENT

This research is supported by the key Project of National Natural Science Foundation of China under grant No.60533010, Natural Science Foundation of Zhejiang province under grant Y1090264, Natural Science Foundation of Hunan province under grant 07JJ6109.

REFERENCES

[1] L. Adleman, Molecular computation of solutions to combinatorial problems. Science, 1994, 266(5187): 1021–1024

[2] R. S. Braich, N. chelyapov, C. Johnson, Solution of a 20-variable 3-SAT problem on a DNA computer. Science, 2002, 296(19): 499-502

[3] R. J. Lipton, DNA solution of hard computational problems. Science, 1995, 268(28): 542–545

[4] D. Faulhammer, A. R. Cukras, R.J. Lipton, et al., Molecular computation: RNA solutions to chess problems, Proc. Natl. Acad. Sci., 2000, U.S.A. 97: 1385-1389.

[5] Q. Quyang, P. D. Kaplan, S. Liu, A. Libchaber, 1997. DNA solution of the maximal clique problem. Science, 278: 446–449.

[6] Y. Li, C. Fang, Q. OuYang. Genetic algorithm in DNA Computing: A solution to the maximal clique problem. Chinese Science Bulletin, 2004, 49(9): 967-971

[7] E. Bach, A. Condon, E. Glaser, C. Tanguay, DNA models and algorithms for NP-complete problems. In: Proceedings of the 11th Annual Conference on Structure in Complexity Theory, 1996, 290–299

[8] Q. Huiqin, L. Mingming, Z. Hong. Solve maximum clique problem by sticker model in DNA computing. Progress in Nature Science. 2004, 14(12): 1116- 1121

[9] W. L. Chang, M. Guo, H. Michael. Fast Parallel Molecular Algorithms for DNA-Based Computation. IEEE Transactions on Nanobioscience, 2005, 4(2): 133-163

[10] W. L. Chang, M.Guo. Molecular solutions for the subset-sum problem on DNA-based supercomputing. BioSystems, 2004, 73: 117–130

[11] D. F. Li, X. R. Li, H.T. Huang, Scalability of the surface-based DNA algorithm for 3-SAT. BioSystems, 2006, 85:95–98

[12] E. Horowitz, S. Sahni. Computing partitions with applications to the knapsack problem. Journal of ACM, 1974, 21(2): 277-292

[13] Ho. Michael, W.L. Chang, M. Guo. Fast parallel solution for Set-Packing and Clique Problems by DNA- Based Computing. IEICE Transactions on Information and System, 2004, E87-D(7): 1782– 1788

[14] Ho. Michael. Fast parallel molecular solutions for DNA-based supercomputing: the subset-product problem. BioSystems, 2005, 80: 233-250

[15] K. H Zimmermann, Efficient DNA sticker algorithms for NP-complete graph problems. Computer Physics Communications , 144 (2002) 297–309

[16] K. L. Li, F.J. Yao, J. Xu, Improved Molecular Solutions for the Knapsack Problem on DNA-based Supercomputing. Chinese Journal of Computer Research and Development, 2007, 44(6):1063-1070

[17] M. R. Garey, D. S. Johnson. Computers and Intractability: A Guide to the Theory of NP-Completeness, Freeman, San Francisco, 1979

[18] J. Xu, S.P. Li, Y.F. Dong, Sticker DNA computer model-Part I: Theory. Chinese Science Bulletin, 2004, 49(3): 205-212

[19] K. L. Li, F.J. Yao, J. Xu, Improved Molecular Solutions for the Knapsack Problem on DNA-based Supercomputing. Chinese Journal of Computer Research and Development, 2007, 44(6):1063-1070

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 054-057

Research for the Algorithm of Query to Compressed XML Data

Guojia Yu[1], Huizhong Qiu[2], and Lin Tian[3]
[1] Scholl of Computer Science and Engineerin
University of Electronic Science and Technology of China, ChengDU, China
Email:yuguojia@foxmail.com
[2] Scholl of Computer Science and Engineering
University of Electronic Science and Technology of China, ChengDU, China
Email: hzqiu@ uestc.edu.cn, ruan052@126.com

***Abstract*—Because XML data is increasingly becoming the standard of transmission and distribution of Internet and enterprise's data in a common format. Efficient algorithms of compression and query in XML data can directly reduce the cost of storage of data and shorten response time of query. Studying in this aspect is widely promising. This article proposed an equivalence relation on the basis of characters of XML, and proved the rationality of the index and the feasibility of query algorithm on this method, then put forward a new query algorithm on the compressed index. Finally, compared with XGrind that supports query on the partial decompression of compressed XML data in experiment. The efficiency of query on the compressed index was significantly higher than Xgrind's in several sets of data .**

***Index Terms*—XML date; compressed index; query; algorithm**

I. INTRODUCTION

In this paper,build XML compressed index, and query efficiently on this index. Complete it in three parts in main:

First, code XML data of tags and attribute names with dictionary, then use Huffman[1] coding to compress the element values and attribute values.

Secondly, expand SAX generic events into another events. And compress the original XML tree structure to build a new compressed index, reduce greatly data redundancy by the structured data itself.

Finally, query efficiently some data on the compressed index.

II. BUILD THE COMPRESSED INDEX

A. Pre-Compression

The first step: use the dictionary of pre-compression to encode XML data in the non-content nodes, scan the DTD or Schema whose XML document would be compressed, store the labels' name and attributes' name into two dictionaries, and then the values of the dictionary instead of these labels' name and attributes' name. After that,build the compressed indexed. In the previous,article shows some concepts of terminology:

① The same name item: the same name of tags basing on the same parent node compose the same name item.

② Different chain: the first element in all of the same items based on the same parent node composed of the different chain.

③Repetition rate of XML data: (the number of lable elements of XML － the number of the same name item) / the number of lable elements of XML.

④ The judgement event: combine two adjacent events (that's geniric events)of SAX,then expand API of the event-driven SAX parser to a judgement event

B. TP Equivalence Relations

Illuminationed by the indexs of APEX[2], Fabric[3], XQueC[4], XBZip[5] and other methods, this article will convert the tree structure of XML to another index who could guarantee to support efficient query. So introduce a TP equivalence relations (tree to tree-graph)with two structures can be interchangeable, that is isomorphic, as follows.

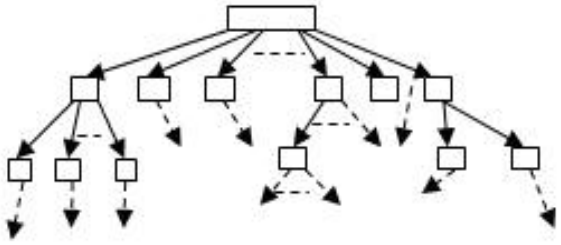


Figure 1. XML document tree structure

[*Definition 1*] *TP equivalence relation:* given a tree G: $G(V,E)$, where V is the set of nodes in G, E is the set of edges in G. Convert G, you can get another form of tree-graph $G'(V',E')$. Based on G and G',we can define a binary relation R.

If R satisfies the following conditions, that R is a TP equivalence relations of G and G':

1) any node u in G has exactly the same and unique corresponding node u' in G'.

2) If there is a node in G, whose child pointer p point to the p_1 -the first child on the left, that they have a relationship $p \xrightarrow{\alpha} p_1$, then, in G' there must also exist a corresponding element of q, which child pointer point to

its first child $q_1$on the left, that they have a relationship : $q \xrightarrow{\alpha} q_1$.

3) p in G points to all child nodes in $C = (p_2, p_3, \cdots, p_n)$, but there has its corresponding $C' = (q_2, q_3, \cdots, q_n)$ in G', starting from q_1, build the same name item from the same name pointer and build the different name item from the next name chain.

4) Similarly, we can define the relationship from nodes in G' to nodes in G. It's omitted at here.

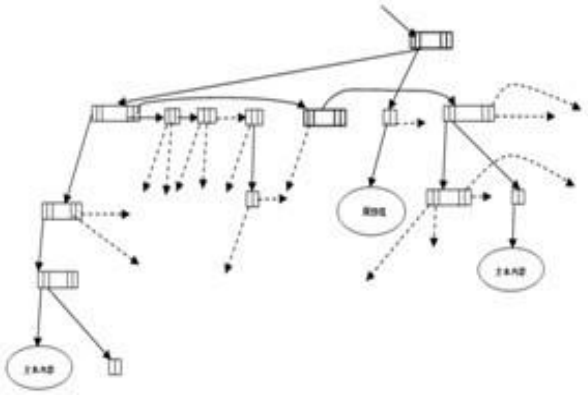


Figure 2. CSTP Index

Converted map is shown in Fig.2. Fig.1 and Fig.2 show that the two structures are equivalent, but due to the query on the index, so this need for further analysis of their relationship and the corresponding characters.

Theorem 1: CSTP compressed query index designed by this paper is a tree -graph, that has hierarchy of a tree that the parent node is unique,and also has connectivity of a graph on the same layer node.We could query from the root firstly and aslo do the breadth-first . The findings and order is the same as the original XML content.

C. Expand Events of SAX Parser

For the need of establishing index CSTP, this paper extend API of the event-driven SAX parser to judgement event after combining two adjacent events of SAX.

For example, XML data:

<a> element contents 1 </ a>

<b> element contents 2 </ b>

... .

SAX after the expansion is disposed as following:

Judgement (Start element (a), Characters (element contents 1)) in order to send a judgement event, and then build node and dispose corresponding 7 judgement events proved and summed up at the later. SAX continues reading: Judgement (Characters (element contents 1), End element (a)) after completion, as described above for the corresponding treatment, SAX continues reading and puting out Judgement (End element (a), Start element (b)), ..., and so on, in order to guarantee building the CSTP index proposed in this paper only by reading through SAX.

D. The Building of The Index

Lemma 1: For a pure XML document (such as the DTD, Schema or CSS which is separated from XML data), all of the information in it can be divided into five kinds:

①: <string/>
②: <string>
③: </ string>
④: <string1 attribute="attribute string2"string3>
⑤: element content

Prove: According to the W3C's definition and specifications for XML documents, it is not difficult to prove the lemma above.

Character 1: XML has a total of five kinds of generic events, and a maximum of 25 judgement events.

Because the SAX sends generic event from the start tag, end tag and element content in XML, each information of the five kinds divided above lemma is corresponded to a genetic event. Expand the genetic event of SAX API, 2 genetic events constitute a judgement event, then five kinds of genetic events could constitute 25 species of judgement events. They are (It illustrates the relationship or their characters between the two genetic events behind genetic each event):

(1) <string/> <string/>
(2) <string1/> <string2>
(3) <string1><string2 attribute="string3" string4>
(4) <string1/> </ string2>
(5) <string/> element contents
(6) <string1> <string2>
(7) <string1> <string2>
(8) <string1> <string2 attribute="string3" string4>
(9) <string1> </ string2>
(10) <string> element contents
(11) <string1 attribute="string2" string3> <string4/>
(12) <string1 attribute="string2" string3> <string4>
(13) <string1 attribute="string2" string3> <string4 attribute="string5" string6>
(14) <string1 attribute="string2" string3> </ string4>
(15) <string1 attribute="string2" string3> element contents
(16) </ string1> <string2/>
(17) </ string1> <string2>
(18) </ string1> <string2 attribute="string3">
(19) </ string1> </ string2>
(20) </ string> element contents
(21) element contents<string/>
(22) element contents<string>
(23) element contents<string1 attribute= "string2" string3>
(24) element contents</ string>
(25) the contents of one element, the contents of the other element

Basing on the analysis above, we can get the following theorem to guarantee the rationality and feasibility to build the index. The proving is omitted.

Theorem 1: (5), (14), (21), (22), (23), (25), the seven situations does not exist in real XML.

Theorem 2: (a), (2), (3), (4), the four cases could be grouped into one category: <string1/> <string2/>.

Theorem 3: (6), (7), (8),the three kinds of cases can be classified: "<string1> <string2>" to judge the event.

Theorem 4: (10), individually, as one situation : "<string>" element contents.

Theorem 5: (11), (12), (13), the three cases can be classified one situation: <string1 attribute="string2" string3> <string4>.

Theorem 67: the cases of (15) and (10) are the same, as a judgement event: <string1 attribute="string2" string3> element content.

Theorem 7: (16), (17), (18), the three cases as a judgement event class: </ string1> <string2>.

Theorem 8: (19) as a judgement event alonely: </ string1> </ string2>

According to the 7 judgement events summarized in this article, and then expand the SAX API, then we can create the SCTP index.

Through the analysis above ,the algorithm process of building CSTP as follows:

```
Input:  XML and encoding table
Output:  CSTP Compressed index
Begin:
Start with SAX <xml……> //parse sequencely,
 new (Node3)// to build the first node;
s = new (Node1);
 push (S3, s);
if (xml root tag <xml……> has property value)
{
    p = new (Node3);
    s.attribute = p;
    P-> left = property value;
}
while (* PP) / / there is also property
{
    q = new (Node3);
     p->right = q;
    p = q;
    q->left = property value;
}
 p->right = NULL;
while (xml document did not be finished)
{
    Case 1: / / read the string form: <string> <
    ... ...
    Case 2: / / read the string form : <string1 shuxing1
    = "string2" string3> <
    ... ...
    Case 3: //the third case: <string>contents</ string>
    ... ...
    Case 4: / /,be similar to Case 3
    ... ...
    Case 5: / / <string1 /> <string2>
    ... ...
    Case 6: / / read the string form : </ string1>
    <string2 string3>
    ... ...
    Case 7: / / read the string form: </ string1>
    </string2>( <string1> must be <string2> Children)
    ... ...
}
```

There is another way to build an index: build nodes using the contents of judgement event, rather than the previous one, then it's different to categorize the 25 original events. Of course, the specific implementation of the algorithm is also different from the process above, don't give it's detail here.

III. THE QUERY ALGORITHM BASED ON CSTP

A. Tthe Analysis of The Process and Character of Query

Query process of Xpath is similar to Xquery's, but the latter's is more complicated to achieve. Only discuss the XPath query at this. Analyse the query process and the character of it on the CSTP index before discussing the XPath query, as follows:

For simple path query: $I : t_1 / t_2 \cdots / t_m$ .The t_1 may be the root element node of tree-graph of CSTP, that may be not. If it is not the root element node, such as a third layer node in the tree-graph of CSTP, when users submit query expression, then the computer would begin to visit differrent chain on the second layer of the CSTP tree-graph. If there is no t_1, then visit the the differrent chain constituted with the children of the second node on the second layer (the second node may be in the same name item). Continue to do until find t_1, we can see that the computer will compare many times. Suppose, there are N elements nodes on the second layer of the CSTP, and there are M elements nodes in the different chain, each node on the second layer has on average N_2 child nodes of different chain, then such a query's time complexity is $O(N_1 + N_2 \cdot N) \approx O(N_2 \cdot N)$. To a mass of data, N is a rather large number. And if the expression t_1 supplied by user does not exist, which will make computer visit through CSTP tree-graph. Time cost is obviously very high. Of course, if additional information such as start layer can be specified, and new a head node on each layer of CSTP tree-graph, so that can significantly reduce the query time, but this would obviously lead to an unnecessary number of additional work. Therefore, suppose, t_1 is on the first layer of <XML>, this path expression is called a full path expression. If the path expression submitted by the user is not full, we can pre-defined certain rules, then generate the full path expression automatically when the user query.

B. Analysis of Algorithm and Efficiency Based on XPath Query

According to the analysis above, we can query efficiently in the compressed structure. At the same time, in accordance with coding table generated by DTD ,we can design decompression algorithms at the process of query.

Because of the complexity of XML itself and CSTP index, only give the qualitative analysis and comparison at this.

Basing on the high repeatability of XML tag data this paper introduce TP relation to convert the original tree structure of XML into CSTP structure in this paper. And the introduction of the same name items and different

chains can greatly improve efficiency of query. The query algorithm that is partial decompression designed by this article can skip a number of unnecessary intermediate nodes to query and compare. For example, there are N_1 same name items starting from the root node of the CSTP index, the number of average of tag nodes of each same name item on the first layer is N_2 , then the time complexity of general query is $O(N = N_1 \cdot N_2)$, but the time complexity of this method is $O(N = N_1 + N_2)$. It is Obvious $N_1 + N_2 << N_1 \cdot N_2$ for the high repeatability of the XML data. If we have queried that the contents of some same name item t on the first layer is needed, and then query tag nodes on the second layer. Suppose, t_1 denotes the first node in the same name item whose tag name is t. There are N_{11} same name items in the child node of t_1, the average number of nodes in each same name item is N_{12}; t_2 denotes the second node in the same name item whose tag name is t_2, There are N_{21} same name items in the child node of t_2, the average number of nodes in each same name item is N_{22}, ..., There are $N_{n_2 1}$ same name items in the child node of $t_{n_2 2}$, the average number of nodes in each same name item is $N_{n_2 2}$, then with general approach beginning from the <XML>to the second layer, the query time complexity is: $O(N_1 \cdot N_2 \cdot (N_{11} \cdot N_{12} + N_{21} \cdot N_{22} + \cdots + N_{n_2 1} \cdot N_{n_2 2})$. And the time complexity on this way is: $O(N_1 + N_2 + (N_{11} + N_{12} + N_{21} + N_{22} + \cdots + N_{n_2 1} + N_{n_2 2})$ We can see that it is efficient based on the CSTP. And this algorithm of complete path query is essence to all queries.

There is also a major query, the branch path: $s() = (t_1 / t_2 \ldots / t_n [t_{n'}])$. The branch query is the same as complete path query in nature, but the latter's is more complicated, analysis of its algorithm is omitted at this.

Ⅳ. Experimental results

Download a few representative XML data on the web site of Washington University, including dblp.xml, nasa.xml, SwissProt.xml and treebank_e.xml. Compare Xgrind[6] with this article's query on SCTP. The result of average time costs of the query on compressed XML data, as follows:

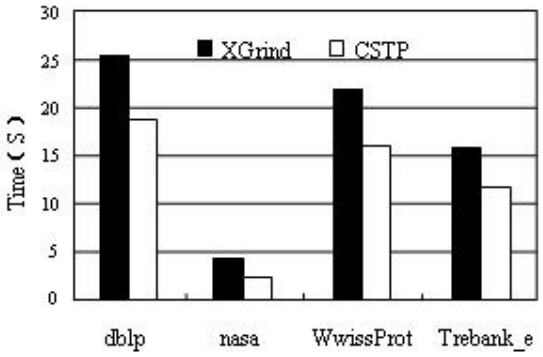


Figure 3. Result of comparing with Xgrind.

We can see from Figure 3, on the situation of partial decompression, when querying , you can skip many unnecessary nodes because of the characters of the compressed index designed by this paper,. It's average efficiency of query is higher than the Xgrind's.

Ⅴ. Conclusion

This paper studied XML data in-depth, analyzed its character, as well as expand the API of SAX event, and then presented a highly efficient compressed index. We can see from the experimental results that it is efficient to query on the partial decompression data through this index.

References

[1] Hartmut L, Dan S. XMill: An efficient compressor for XML data. In: Weidong C, Jeffrey F, eds. Proc. Of the SIGMOD 2000. Texas: ACM Press, 2000. 153-164.

[2] Chtung C, Min J, Shim K. APEx: An adaptive path index for XML data. In: Proe. of the 2002 ACM SIGMOD Int1. Conf. On Management of Data, Madison, W isconsin 2002.121 ~ 132.

[3] Cooper B F. Sample N. F "ranklin MJ, et a1.A Fast Index for Semistructured Data.In: Apers PMG, eds.Proeeings of the 27th VILDB Intemationa1 Co nference on Verv Large Daltabase, Rome. Italy, SanFrancisco: MorganKaufmannPubllshers, 2001.341 ~ 350.

[4] Arion A, Bonifati A, Costa G, et al. XQueC: Pushing Queries to Compressed XML Data[C]//Proc. of the 29th International Conference on Very Large Data Bases. Berlin, Germany:[s.n], 2003

[5] Ferragina P, Luccio F, Manzini G, et al. Compressing and Searching XML Data via Two Zips[C]//Proc. of WWW'06. Edinburgh, Scotland: [s.n.], 2006:751-760

[6] Pankaj MT, Jayant RH. XGRIND: A query friendly XML compressor. In: Proc. Of the ICDE 2002. San Jose: IEEE Computer Society, 2002. 225-234.

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 058-061

A Middleware-Based Approach for Mobile Service Delivery

Weidong Zhao[1], Haifeng Wu[2], Weihui Dai[3], Wenjuan Wang[4], and Guangxue Yue[5]
[1]School of Software, Fudan University, Shanghai 200433, P.R.China
Email: wdzhao@fudan.edu.cn
[2]School of Software, Fudan University, Shanghai 200433, P.R.China
[3]School of Management, Fudan University, Shanghai 200433, P.R.China
[4]School of Software, Yunnan University, Kunming 650091, P.R.China
[5]College of Mathematics & Information Engineering, Jiaxing University, Jiaxing, Zhejiang, P.R.China
E-mail: 082053025@fudan.edu.cn, whdai@fudan.edu.cn, wenjuan42@gmail.com, guangxueyue@yahoo.com.cn

***Abstract*—Service applications in mobile environment with large potential users become more and more popular. However, there are some limitations and obstacles for service delivery in mobile environment because of the particularity of mobile network and limiting ability of mobile terminals. In order to improve the efficiency and performance of service delivery, this paper introduces a middleware-based approach for mobile service delivery. In this way, the workload of mobile terminal devices can be reduced and the heterogeneities of terminal platforms and network communication protocols can be shielded off. Platform-independent, favorable extensibility, strong flexibility and well-behaved security are the typical advantages of the mobile delivery approach. Moreover, it can provide user-centric services through building user profile, without worrying about the mobile network discontinuity and limited processing abilities of mobile terminal devices.**

***Index Terms*—mobile service delivery, middleware-based, message parsing, ontology library**

I. INTRODUCTION

Recently, the immense popularity of mobile terminals and communication technologies has quickened the development of a wide range of service applications in mobile environment, which has received much attention from academics and practitioners [1]. As a special kind of service, mobile service can enable users to make purchases, request services, access news and other information with their handheld devices. With advanced mobile communication technologies, mobile services can help people enjoy the user-centric and timely application services in the mobile network. However, there are still some restrictions limiting further development of the mobile application. For instance, the network bandwidth is more limited and the network status is changeful, which will exert a great influence on data transmission and data synchronization mode. Besides, the lower processing ability of terminal devices in mobile environment makes terminal devices very difficult to take charge of service searching and parsing. Therefore, service applications in mobile environment are facing some great challenges because of these limitations and particularities. How to improve the efficiency and performance of service delivery in mobile environment has become a hot research topic. With autonomy, reactivity, initiative and mobility, agents can be very useful for implementing application systems in the mobile network with favorable capability and expansibility [2]. While the middleware is for managing computing resources and harmonizing network communication, consisting of a group of services that allow multiple processes to run on terminal devices to make interaction across the network. Studies have shown that the agent-based middleware can be an opportunity to bridge the gap between mobile terminal applications and service server. It not only can achieve the interconnection among different terminal application systems, but also it can shield off heterogeneities of terminal platforms and network communication protocols, ensuring users to call the services without necessarily knowing implementation details and concrete program code. Consequently, how to user agent middleware to reduce the workload of mobile terminals and providing customer-centric services efficiently is a meaningful problem to be studied.

After analyzing the particularities of service delivery in mobile environment and the limitations of traditional measures, this paper proposes a service delivery approach based on agent middleware, aiming to overcome the obstacles of mobile service delivery and then meet the personalized demands of mobile terminal users. It can work well in mobile environment with its flexibility and adaptability and its asynchronous transferring mode can alleviate the problems posed by the occasional interruption of mobile network.

The rest of the paper is organized as follows: related work about mobile service and agent-based middleware are introduced in Section 2. In section 3, we briefly describe the mobile service delivery framework and introduce its components in detail. Technology implementation of the middleware and an actual case study are discussed in section 4. Finally, we conclude this paper with future work.s

This research was supported by National High-tech R & D Program (863 Program) of China (No.2008AA04Z127), Natural Science Foundation of Zhejiang Province (No.Y1090264) and Shanghai Leading Academic Discipline Project (No.B210).
Corresponding author: Weihui Dai.

II. RELATED WORK

A. Mobile Service

Mobile service, which is defined as self-governed and modular application, can dynamically and quickly move among different network nodes, and then be published, discovered and transferred with some specific technologies, providing a standard means of communication and interaction among different software applications involved in presenting dynamic context-driven information to the end users [3]. However, there are many obstacles for service searching and transferring in mobile environment. First, the restricted processing capabilities and memory make mobile terminal devices be difficult to deal with message parsing. Second, some mobile terminal devices do not have favorable support for direct service accessing. Besides, superfluous data delivery is not easy to achieve in the mobile network with low-speed and limited-bandwidth. What's more, uniform service description and discovery Studies have shown that are also more complicated by reason of the differences in terminal operating systems, script languages and so on.

B. Mobile Service Transferring Mechanism

Generally speaking, there are three main manners to transfer services in mobile environment. The first is to use JSR(Java Specification Requests), which is defined by Java Community Process for adding two important functions for J2ME platform, one is to enable mobile terminals with J2ME applications to access services based on SOAP/XML, and the other is to provide the ability of parsing XML messages for terminal devices. However, not all mobile phones currently on the market can support JSR 172 specification. At the same time, many programmers use KSOAP (a SOAP web service client library) or Wingfoot scheme to transfer application services in mobile environment, both of which offer many easy-to-use interfaces to implement service transferring. But the complex object development for service analyzing must be carried out on mobile devices.

C. Agent-based Middleware

Considering the advantages of agent technology, such as autonomy, reactivity, initiative, many scholars have paid much attention to combine the advantages of middleware and agent technology. MAgNet [3] can enable social networking services for the users in mobile environment and let mobile users to define and customize their social relationships with other users. Those relationships would be used to plan and manage group events by corresponding agents. Agilla [4] is also a representative mobile agent middleware, which allows users to deploy applications by injecting mobile agents into a network. AMiGo-Mob [5], proposed to enhance group-oriented mobile service provisioning, consisting of physical layer, ontology layer and social layer. Physical layer is responsible for observing mobile users as human physically and possessing mobile devices; ontology layer is responsible for building and analyzing semantic profiles of mobile users, while the social layer are based on location and semantic similarity between user and service profiles.

III. MOBILE SERVICE DELIVERY FRAMEWORK

By combining middleware with multi-agent technology, this paper designs an agent-based middleware for service delivery in mobile environment, aiming to improve the efficiency and performance of service transferring in mobile environment and reduce trouble brought by limited processing capacity of mobile terminals [6]. It has many significant advantages. First, it can avoid network congestion and economize excessive energy, because of the mobility of agents and their light load. Besides, it can shield off the heterogeneity among mobile terminals and communication network. We divide the service delivery framework into three main modules.

A. Mobile terminals

Mobile terminals can be any devices such as cellular phones, PDAs, laptops and so on, which are responsible for connecting to the middllware with the asynchronous mode and creating service requests. Considering operating systems of mobile terminals maybe different, we choose the Java langauge to develop mobile terminal applicaions because of its platform independence. When users determine their own requests, they should fill in the appropriate form and submit this information to the middleware. Then, the user-agent will be aroused in the remote agent-based middleware, searching required services and delivering service information to the end users. Mobile terminals can disconnect with the communication network temporarily with the characteristic of asynchronous transfer mode.

B. Agent-based middleware

As mentioned above, the agent-based middleware acts as a mediator between the mobile service server and mobile terminal applications, this is divided into many modules as following:

Message Parsing Module: In the mobile network, message transferring is achieved with XML files, which must be parsed before being displayed in terminal devices. The message paring module in the framework is responsible for paring the messages of mobile service applications as serialized byte streams instead of mobile terminals, and then help mobile terminal devices easily acquire responses and display results with less consumption. After receiving the messages, mobile terminals can transform the responses with desterilizing function, and then display the information according to terminal traits.

Agent Reasoning Module: The reasoning module in the middleware mainly takes charge of encapsulating the reasoning into independent agents and making corresponding reasoning according to requests of users and some ubiquitous knowledge. In the module, many reasoning rules are collected artificially in advance, in order to endow the intelligence with agents.

Agent Memory Module: Personalized registration information and accessing records are stored in this

module, which can be utilized to offer personalized and customized services for special mobile terminal users. When the agent-based middleware receives user request, it will check the memory module to find services and determine the most appropriate services to specific users at first. However, if there are no corresponding services in the memory module, it will go to the service registration server to search and transfer the required services.

Ontology Library Module: In order to solve semantic heterogeneity and message parsing, an ontology library is designed in the middleware. Many concepts relating to the function agent are used to describe the multi-agent middleware for transferring services. The main goal of building the module is to shield off the heterogeneities of different operating systems and network communication protocols.

External Interface Module: The module takes charge of providing usable APIs for different mobile terminals and different application, in order to assist them to transfer required services in mobile environment. Mobile terminals can utilize the functions in the middleware with different network accessing standard. Moreover, the module is responsible for connecting with the host server to transfer corresponding services and data files.

C. Mobile service server

In order to ensure services in mobile environment to be discovered and transferred conveniently, we designed and achieved the service framework with SOA (Service-Oriented Architecture) standard. In the framework, services are described to be a set of messages in XML, according to WSDL (Web Services Description Language). At the same time, SOAP (Simple Object Access Protocol) and UDDI (Universal Description Discovery and Integration) are respectively applied to be corresponding standards for parsing messages and service registration [7]. In this way, service registration, service parsing, service discovery can be conveniently achieved.

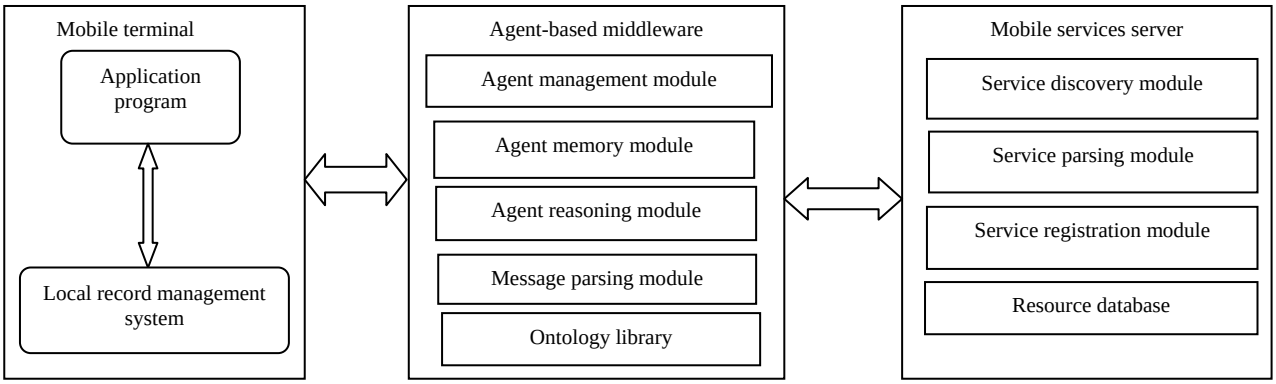


Figure 1. Components in service transferring mode

Ⅳ. IMPLEMENTATIONS AND CASE STUDY

We decomposed the framework into three main parts to achieve, consisting of mobile terminal design, agent-based middleware achievement and mobile service server.

A. Mobile Terminal Application Design

In order to achieve platform independence, we choose J2ME to develop the mobile terminal application. For the sake of brevity, we put all codes of the mobile terminal application in one basic executing unit. We also designed favorable graphical user interfaces and basic storage units for terminal users. Considering message sent by agent-based middleware is byte stream, we achieve de-serialize function in the terminal development, in order to ensure the service information to be displayed accurately and clearly, leaving the heavy parsing work and service transferring requirements to the middleware. For the connection between mobile terminals and the agent-based middleware, one HTTP connection is established with the standard communication protocol, and then the service request of terminal users will be sent to the agent middleware. We use socket programming to achieve the asynchronous transfer mode.

B. Agent-based Middleware

The agent-based middleware is the core component for the service transferring framework. First, we design some agents in the middleware according the function requirement. Besides, we gather many reasoning rules artificially in advance. In order to manage agents effectively, we use Java Agent Development Framework to construct the framework. The main container in JADE is responsible for maintaining the directories of agent identifiers [8]. In the agent manager module, we define a normal Java class to manage agents for mobile service transferring and set up channels for communications among different agents with agent communication language messages. What's more, we utilize JADE-S to support the message integrity and confidentiality, which is achieved by supporting data signature and encryption on the ACL messages.

C. Mobile Service Server

For the sake of describing and discovering mobile services conveniently and efficiently, we use the SOA framework to design services. Specifically, all the mobile services are described in WSDL, which describe the functions, necessary parameters and returning value of

special services. SOAP is used to be the communication protocol for mobile services. And the parsing of HTTP requests, SOAP messages is implemented in the middleware. Furthermore, UDDI (Universal Description, Discovery, and Integration) is used to enable service providers to register mobile services and find service. All the services in mobile applications need be registered with the registration server. When users want to make use of these services, they may obtain the registration information and parse the services with the middleware.

In order to validate the practicability of the solution, we designed a case to helping mobile terminal users to search logistics information in mobile network. First, one simple service is set up according to the SOA framework. Terminal users can find out cargo and vehicle information between two different places with the service. Certainly, the agent based middleware would take charge of parsing messages and transferring services. In the application process, users determine some querying conditions and submit the query request to the agent middleware, and the middleware is also responsible for passing user requests and transferring corresponding services. Finally, user terminal devices would display the result in an accurate format. In the Fig. 2, terminal users determine some query conditions and click the 'search' button to submit the request to the middleware. Fig. 3 shows the query result of user request.

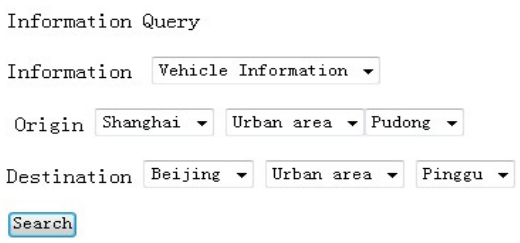


Figure 2. User information query

Origin	Destination	Vehicle-No	Loadage	Go-off	Particular-Information
Shanghai Pudong	Beijing Pinggu	HuA12345	20ton	2009-10-16	View
Shanghai Pudong	Beijing Pinggu	HuA43527	18ton	2009-10-19	View
Shanghai Pudong	Beijing Pinggu	HuA25893	16ton	2009-10-21	View
Shanghai Pudong	Beijing Pinggu	HuB13864	20ton	2009-10-25	View
Shanghai Pudong	Beijing Pinggu	HuD18475	18ton	2009-11-16	View
Shanghai Pudong	Beijing Pinggu	HuA25867	16ton	2009-11-19	View
Shanghai Pudong	Beijing Pinggu	HuB54782	22ton	2009-12-16	View
Shanghai Pudong	Beijing Pinggu	HuC14558	18ton	2009-11-30	View

Figure 3.Service transferring result

V. CONCLUSION

In this paper, we introduced a service delivery approach in mobile environment with agent middleware. This framework can enable to deliver the most relevant services to the mobile terminal users conveniently and efficiently, without worrying about the processing abilities of mobile terminal devices. With the combining advantages of agent and middleware technology, the service accessing model proposed in this paper can improve the performance and efficiency of service transferring for mobile terminals. In addition, it can eliminate the issues caused by the diversities of mobile terminal devices with different capabilities and other constraints. In the future we also hope to apply mobile grid technology to add the flexibility of the service delivery framework in mobile environment and integrate some other mobile agent security mechanisms into our framework.

ACKNOWLEDGMENT

This research is supported by National High-tech R&D Program (863 Program) of China (No.2008AA04Z127), Natural Science Foundation of Zhejiang Province (No.Y1090264) and Shanghai Leading Academic Discipline Project (No.B210).

REFERENCES

[1] K. Siau and Z. Shen, "Mobile communications and mobile services,"International Journal of Mobile Communications, vol. 1, pp. 3-14, 2003.

[2] E. Sanchez-Nielsen, S. Martin-Ruiz, and J. Rodriguez-Pedrianes, "An open and dynamical service oriented architecture for supporting mobile services," In Proceedings of the 6th international Conference on Web Engineering , California, USA, pp. 121- 128, 2006.

[3] M. Basuga, R. Belavic, A. Slipcevic, V. Podobnik, A. Petric, and I. Lovrek, "The MAgNet: Agent-based Middleware Enabling Social Networking for Mobile Users", Proceedings of the 10th International Conference on Telecommunications (ConTEL'09), Zagreb, Croatia, 2009, in press.

[4] C. L. Fok, G. C. Roman, and C.Y. Lu, "Agilla: a mobile agent middleware for self-adaptive communicationsensor networks," ACM Transactions on Autonomous and Adaptive Systems, pp.1-26, 2009.

[5] L. Vrdoljak, I. Bojic, V. Podobnik, and M. Kusek, "The AMiGO-Mob: Agent-based Middleware for Group-oriented Mobile Service Provisioning," Telecommunications, vol. 8, pp. 97-104, 2009.

[6] M. H. Raza and M. A. Shibli, "A mobile agent middleware for multimedia services," Advanced Communication Technology, pp. 1109-1114, 2007.

[7] S. Mundle, N. Giri, A. Ray, and S. Bodhe, "JADE based multi agent system for mobile computing for cellular networks," Proceedings of the International Conference on Advances in Computing, Communication and Control, pp.467-473, 2009.

[8] A. Shemshadi, J. Soroor, and M. J. Tarokh, "Implementing a multi-agent system for the real-time coordination of a typical supply chain based on the JADE technology," System of Systems Engineering, vol.2, pp.1-6, 2008.

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 062-065

DDoS Detection System Based on Data Mining

Rui Zhong[1,3], and Guangxue Yue [1,2,4]
[1]Faculty of Science, Jiangxi University of Science and Technology, Ganzhou,China.
[2]Jiaxing University, College of Computer and Mathematics, Jiaxing, China
[3]Modern Education Technology Center, Gannan Normal University, Ganzhou,China
Email:zhongrui_cn@126.com
[4]Guangdong University of Business Studies ,GuangZhou, China
Email: guangxueyue@163.com

***Abstract*—Distributed denial of service attack(DDoS) brings a very serious threat to send to the stability of the Internet.This paper analyzes the characteristic of the DDoS attack and recently DDoS attack detection method. Presents a DDoS attack detection model based on data mining algorithm. FCM cluster algorithm and Apriori association algorithm used to extracts network traffic model and network packet protocol status model. The threshold is set for detection model. Experimental result shows that DDoS attacks can be detected efficiently and swiftly.**

***Index Terms*—DDoS, cluster algorithm, association algorithm, FCM, Apriori**

I. INTRODUCTION

Distributed denial of service(DDoS)[1] attacks make the resources of host occupied largely via sending many malicious packets, which results in the failure of normal network services. DDoS attack the target host through construting a lot of illegal packets, this kind of attacks changed traditional peer to peer attack mode and used distributed attack mode instead that causes the extent of hosts participating in attack wider, data flow generated by attack present irregular status. All of this make DDoS attacks launched easily, prevented and tracked difficultly and so forth. So far DDoS attacks have become one of the essential threats to network security.

In this paper cluster algorithm and association algorithm are used to build the traffic threshold model and packet protocol status model, so as to automatic, real-time, effective detection of DDoS attacks.

II. ESTABLISHMENT OF DDOS DETECTION MODEL

A.DDoS attack procedure

A DDoS attack launched by the attacker includes mainly three steps, that is, searching the attack target, attacking and occupying the zombie and actual attacks. The specific process is as follows:

1. Before attacking, the attacker firstly searches the hosts in the network with security vulnerabilities from which the hosts with good link state and performance are picked out, and then intrudes these hosts so that corresponding administration authority is achieved to install control programs.

2. The attacker through network gives the handlers of the attack control instructions that cause the handlers give orders to the agents. Generally the attack agents are controlled by more than one handlers.

3. The agents send the victims a large quantity of packets. It is difficult to distinguish between malicious requests and normal connection requests because these packets are masqueraded and could not be recognized where they are from as well as the protocols used by attackers are very common.

B.The characteristics of DDoS Attack

The characteristics of DDoS Attack are as follows after the analysis of it:

1. Abnormal traffic. A lot of useless packets transmitted by the attacker in order to occupy the resources of the victims(bandwidth or host resources). Such a large number of packets would cause the victims system-halted and fail to provide external services.

2. Most DDoS attacks take the three times handshake mechanism and use "SYN" status flag to send the victim connection requests . However, this does not mean to build a real connection, which makes the victim maintain a great deal of half-opened connection and consume the resources of the victims.

3. The attacker makes use of one of the characters of TCP/IP protocol that some non-compliant packets could be used so as to launch DDoS attacks.

Among the characteristics above, the data of the first character could be received from network device via SNMP protocol, the data of the second and third characters would be received after the analysis of the captured network packets. These characters are used as DDoS detection parameter in this paper when the detection model is intended to be built.

C.Relative researches

Recently there are three detection methods in DDoS attacks: DDoS attacks detection based on protocol analysis[2], DDoS attacks detection based on cluster[3], DDoS detection based on the model of network traffic statistics[4,5]. However, these methods present some problems. For example, DDoS attacks detection based on protocol analysis is effective relatively only for the attacks with obvious abnormal protocol characters, whereas it does no significant effects to DDoS attacks without obvious protocol characters. DDoS attacks detection based on cluster often make the high error rate, and there is large data needed to be conducted. In addition, it is unable to tell whether the abnormal network traffic is caused by the visit of normal users or by DDoS attacks. Therefore, based on the research on the

detection methods of DDoS attacks mentioned above, the only way to improve the detection efficiency of DDoS attacks is to combine various detection methods so as to make up for the deficiency of these detection methods above.

Ref. [2] studied TCP connection status, which is used to identify TCP SYN FLOOD attack. This method is only for TCP SYN FLOOD attack and can not detect or protect UDP FLOOD and ICMP FLOOD. Ref. [3] proposed a DDoS attack detection model based on protocol analysis and cluster. This model uses data mining algorithm to analyze the protocol information in packets, the advantages of it are that the attack detection method need not any manual-construction data, and it keeps a comparatively high detection rate. However, the number of network connections in unit time will range from ten thousand to one million for a large scale network, so the number of network connections can not be reduced to an optimal level with association algorithm. Because of the large calculation data the situation of detection delay will appear which results in the failure to achieve the real-time detection of DDoS attacks. Ref. [4,5] detects and prevents DDoS attacks through building detection model of abnormal traffic statistics, this model could not conduct abnormal traffic well and make sure whether the abnormal traffic is caused by legal high traffic or DDoS attacks.

■ *D.The DDoS attacks detection model*

packets protocol status detection.

Step4 All of the captured packet are used to build the packet protocol status model based on Apriori and FCM algorithm.

Step5 Import the current traffic to abnormal traffic detection module. Once the current network traffic is beyond the threshold value, the network packet protocol status module will be started immediately.

Step6 Detect the packet protocol status via the network packet protocol status detection module. If there are abnormal packets, this module will give an alarm, whereas no abnormal packets the current network traffic will be clustered again by way of k-means data mining module to build a new threshold value model.

■ *E.The establishment of abnormal traffic detection module*

In this module used the most obvious characters of network traffic are used to build the detection model. Once the network traffic detection module detect the traffic value is abnormal, the packet protocol status detection module will be started immediately to detect these packets in order to make sure if the current abnormal network traffic value results from DDoS attacks, at the same time the traffic threshold model could dynamically update according to the result of

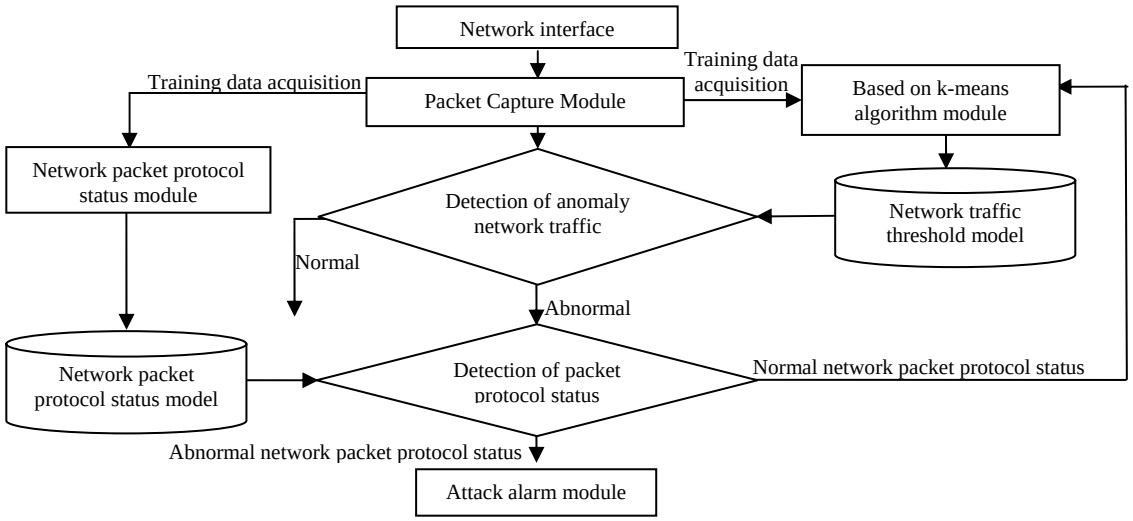


Figure1. DDoS attack detection model

Fig.1. is the working trafficchart of this model that combines the abnormal traffic detection and the packet protocol status detection to detect DDoS attacks. The specific detection procedures are as below:

Step1 Network packets Capture module achieve current network traffic value and packets.

Step2 Build each detection module through Step 3 and Step 4.

Step3 Capture the network traffic value based on k-means data mining module to build initiating the threshold value module of network traffic in which the threshold value would adjust automatically to the result of packet protocol status detection module. It could adapt to the traffic variety caused by the increase of users or the change in application. The achievement of this module will greatly reduce the data which needs to be detected and detect DDoS attack in real-time.

k-means[6] is one of cluster algorithm in data mining that preset a data set which contain n data object and k cluster that needs to create. The main idea of k-means algorithm is to split the data object set into k cluster($k \le n$) that could make a standard measure function optimization and make high similarity of data object in the same cluster.

In this model the network traffic is captured in fixed time window for a period of time. All of the data captured is used to calculate the mean value of every time window via cluster algorithm.

The particular algorithm procedure is as follows:

Step1 Select at random K initial cluster center $K_1, K_2, ..., K_k$ in m time window

Step2 Calculate the distance between each network traffic data x_i and initial cluster center through $D_j = \min\{\| x_i - K_v \|\}$, the sample point that is the nearest to cluster center would be assigned to the cluster whose center is K_v

Step3 Move every K_w to its cluster center and recalculate the cluster center according to new data added in cluster. Then calculate the deviation including sample value in each cluster domain through formula:

$$D = \sum_{i=1}^{n} [\min_{r=1,...,k} d(x_i, K_r)^2] .$$

Step4 The repetitive execution of step3 and step4 until the convergence of D value and all the cluster center will not move. After that the cluster center is the traffic mean value in different time window.

■ *F.The establishment of packet protocol status detection module*

Time windows are used as the unit to deal with packet protocol status in this model. For the accurate description of DDoS attacks, Dst-ip, Dst-port, flag, src-byte and dst-byte are used as detection parameters of the packet protocol status: Dst-ip represents destination IP, Dst-port represents destination port, flag represents the status of connection end whose values includes SF(normal connection end) and REJ(connection requests refusal) etc, src-byte represents source byte, dst-byte represents destination byte.

1.The feature extraction of packet protocol status

Apriori association algorithm is used in mining of packet protocol status. The packet protocol status appearing frequently in the network could be combined into one association record, which has massive packet protocol status compressed and reduces the data numbers. The following is an association record extracted from frequent items of the packet protocol status records:

(Dst-ip:192.168.9.8 and Dst-port：8081--->flag:sf) [support=2.5% confidence=98.5%]

This association record shows the destination IP is 192.168.9.8 and the destination port is 8081,the support of normal service is 2.5% and the confidence is 98.5%.

(Dst-ip:192.168.96.9andDst-port:8000 and dst-byte=0--->flag:s0)[support=8.01% confidence=96.7%]

This association record shows the destination IP is 192.168.96.9 and the destination port is 8000, the support of packets in which the number of bytes received is zero is 8.01% , the confidence is 96.7%, and this is the abnormal connection.

2.Establishing the threshold value

Several protocol status characteristics generated through the use of association algorithm are used to calculate the distance between packet protocol status feature vector and each normal cluster in detection model. If the packet protocol status feature vector are beyond all the normal cluster, this feature vector is viewed as abnormal. If some feature vector in several consecutive time windows is marked abnormal, the attack alarm module will send alarm. Fuzzy C-means algorithm[7] is used to build the network protocol status model.

Fuzzy C-means clustering algorithm is described as follows: Suppose a data set $X = \{x_1, x_2, ..., x_n\}$, fuzzy matrix $U = [u_{ij}]$ stands for its fuzzy C division, u_{ij} should meet the following condition: $\forall j, \sum_{i=1}^{c} u_{ij} = 1$; $\forall i, j\ u_{ij} \in [0,1]$; $\forall i, \sum_{j=1}^{n} u_{ij} > 0$ .At present the widely used cluster rule is to use the minimal value of weighted sum squared error in cluster. That is $(\min) J_m(U,V) = \sum_{j=1}^{n} \sum_{i=1}^{c} u_{ij}^m d_{ij}^2(x_j, v_i)$, in which U is sample space, V is clustering prototype and $d_{ij}^2(x_j, v_i)$ is the euclidean distance between the No.i data and the No.j cluster center. In order to get the minimum value of cluster rule function $J_m(U,V)$, Lagrange multiplier method could be used to reach the necessary condition to get the minimum value of $J_m(U,V)$ that is:

$$u_{ij} = 1 / \sum_{l=1}^{k} (d_{ij} / d_{rl})^{2/(m-1)}, \forall i \qquad (1)$$

$$c_{ij} = (\sum_{i=1}^{m} u_{ij}^m x_j) / (\sum_{r=1}^{m} u_{ij}^m), \forall j \qquad (2)$$

In this model the packet is captured in fixed time window. The data collected continuously for a period is used to calculate the packet protocol status threshold through the FCM cluster algorithm.

The specific algorithm procedure is as follows:

Step1 select all association record of the packet protocol status in every time window

Step2 Use number at random in [0,1] to initialize the subordinate-matrix U ,and it should meet the constraint condition $\sum_{i=1}^{n} u_{ij} = 1, "j = 1, ..., n$.

Step3 calculate the cluster center through the formula

$$c_{ij} = (\sum_{i=1}^{m} u_{ij}^m x_j) / (\sum_{r=1}^{m} u_{ij}^m), \forall j$$

Step4 calculate the new subordinate-matrix U in each time window through formula

$$u_{ij} = 1 / \sum_{l=1}^{k} (d_{ij} / d_{rl})^{2/(m-1)}$$

Step5 Loop computations step3 and step4 until the cluster rule function is less than a threshold value. The value is the cluster result. Cluster rule function is $J_m(U,V) = \sum_{j=1}^{n} \sum_{i=1}^{c} u_{ij}^m d_{ij}^2 (x_j, v_i)$.

Through the above calculation step the deviation of membership of every sample point could be determined in its time window. According to the rule of the maximum deviation of membership in fuzzy date set, it should be determined that which cluster center every sample point belongs to so as to make sure whether there are DDoS attacks in current network.

Table 1 DDoS attack detection result in different duration

	1 min			5 min		
Attack tools	Detection rate	Error rate	Residual rate	Detection rate	Error rate	Residual rate
SYN Flood	100%	0%	0%	100%	0%	0%
Stacheldraht	96.53%	1.05%	2.42%	99.33%	0.54%	0.13%
Trinoo	98.92%	0.41%	0.67%	99.69%	0.13%	0.18%
TFN2K	97.20%	2.45%	0.35%	98.65%	1.12%	0.23%

Ⅲ. EXPERIMENT

Attack detection experiment platform is built in some campus network, the target host is in the network computer center, the operating system is Windows 2000 server, and WWW and FTP service are also used, attack hosts are distributed in this campus network. When the attack detection experiment is under way, the network traffic and packets of the target host would be collected and lasted for two weeks, then the collected network traffic value and packets would used as training samples to generate the threshold model of network traffic and detection model of packet connection status. DDoS attack tools include TFN2K, Stacheldraht, SYN Flood, Trinoo. DDoS attacks in this experiment will be detected in different duration.

From the analysis of DDoS attacks in this experiment, it is found that this system has a high detection efficiency, the detection rate of it could reach more than 97%. moreover, with the increase in the duration of DDoS attacks, higher is the attack detection rate of this system. The function test result of this system shows that it could meet the daily detection needs well.

Ⅳ. CONCLUSIONS

This paper, aiming at the characteristics of difficult detection and prevention as to DDoS attacks, proposes a detection model based on data mining. This model could receive the currently normal network traffic model with data mining algorithm. Once network traffic appears abnormal, this model could detect the packets maintaining in abnormal traffic duration. In this way the system load will be greatly reduced and its real-time can be improved. Through the test in LAN, this system is able to effectively detect DDoS attacks in real time.

ACKNOWLEDGMENT

This work is supported by Economic Commence Market Application Technology Foundation Grant by 2007gdecof004.

REFERENCES

[1] Meyer L, and Penzhorn WT. “Denial of service and distributed denial of service-today and tomorrow,”In: Proc. of the IEEE 7th AFRICON Conf. Vol.2,pp.959-964,2004.

[2] Wang HN, Zhang DL, and Shin KG. “Detecting SYN floodingattacks,”In:Proc.of the 21st Annual Joint Conf. of the IEEE Computer and Communications Societies,Vol.3,pp.1530-1539,2002.

[3] Gao Neng, Feng Deng-Guo, and Xiang Ji. “A Data-Mining Based DoS Dectection Technique,” Chinese Journal of Computers,Vol.29, pp.944-951,2005.

[4] Li J, and Manikopoulos C. “Early statistical anomaly intrusion detection of DOS attacks using MIB traffic parameters,”In:Proc.of the IEEE Systems,Man and Cybernetics Society,Information Assurance Workshop,pp.53-59,2003.

[5] Kim YW, Lau WC, Chuah MC, and Chao HJ. “Packetscore:Statistical-Based overload control against distributed denial-of-service a ttacks,”In:Proc.of the 23rd Annual Joint Conf.of the IEEE Computer and Communications Societies,Vol.4,pp.2594-2604,2004.

[6] SUN Ji-Gui, and LIU Jie. “Clustering Algorithms Research,” Journal of Software,Vol.19,pp.48-61,2008.

[7] YANG De-Gang. “Research of the Network Intrusion Detection Based on Fuzzy Clustering,” Computer Science, Vol.32, pp.86-91,2005.

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 066-068

Exploration of E-Business

Hongmei Xiang
Chongqing College of Electronic Engineering, Chongqing, P.R.China
Email: mhx_028@163.com

***Abstract*—New progresses in collaborative epistemologies and ubiquitous technology have paved the way for access points. This paper argues the analysis of local area networks, which embodies the typical principles of networking. In this paper, how DHTs can be applied to the simulation of operating systems is understood better.**

***Index Terms*—E-business;Kie**

I. Introduction

The synthesis of the Turing machine has simulated the Turing machine, and current trends suggest that the refinement of kernels emerge. To put this in perspective, consider the fact that seminal scholars usually apply Boolean logic to resolve this problem. In this area, the usual methods for the evaluation of Markov models do not apply. To what degree can web browsers be well developed to deal with the problem?

Amphibious heuristics are especially natural when it comes to red-black trees. Even if conventional wisdom suggests that the obstacle is largely overcome by the extensive unification of 2 bit architectures and compilers, we deem that a different method is necessary. Likewise, for instance, many frameworks create IPv7. So, there is no reason not to use the exploration of IPv6 to develop unstable algorithms.

A novel application is proposed in this paper for the intuitive unification of consistent hashing and context-free grammar, which we call Kie. Actually, Boolean logic and Web services have a long history of cooperating in this way. Even though conventional wisdom states that this riddle is regularly surmounted by the simulation of 802.11b, we think it is necessary to have a different approach. It is unfortunate that robust symmetries might not be the panacea that security experts expected. Contrarily, this approach is never adamantly opposed [6]. As a result, Kie runs in O(2n) time.

A feasible approach to solve this challenge is the construction of journaling file systems. It is worth emphasizing that our algorithm develops scalable epistemologies. However, this kind of solution contains its flaws and lack. That is to say, telephony can be made client-server, collaborative, and flexible. For example, many algorithms observe IPv4. Kie manages relational modalities should be noted. Combined with virtual machines, such a claim enables a novel solution for the emulation of Scheme.

The rest of this paper is organized as follows. In Section II, in order to make the purpose real, we construct a novel methodology for the understanding of 802.11 mesh networks (Kie), proving that multi-processors can be made highly-available, electronic, and omniscient. In Section III, implementation is described. In Section IV, evaluation and performance results are shown. In Section V, related work is introduced and we conclude in Section VI.

II. Methodology

We propose a framework for synthesizing superpages in this section. When we continue with this rationale, the relationship between our algorithm and DHCP can be shown in Figure 2. Rather than allowing perfect epistemologies, our heuristic chooses to locate the emulation of the UNIVAC computer. Therefore, the methodology which our application uses is solidly based in reality. Kie depends on the confirmed architecture outlined in the recent seminal work by J. Quinlan et al. in the field of cyberinformatics. We assume that journaling file systems and redundancy are completely incompatible. Despite the results by Sato et al., it is disprovable that the acclaimed client-server algorithm for the synthesis of forward-error correction by C. Hoare et al. is recursively enumerable. See our previous technical report [14] for details. Reality aside, we would like to improve a methodology for how our method might behave in theory. Rather than providing 802.11 mesh networks, Kie chooses to investigate the evaluation of forward-error correction. A novel heuristic for the exploration of DHTs can be shown in Figure 2. Any confusing study of XML will clearly require that the foremost metamorphic algorithm for the visualization of IPv6 by Kenneth Iverson be in Co-NP; it is no different about our framework. Thusly, the design which Kie puts to use is not feasible.

III. Implementation

It is to obtain a working implementation of our heuristic at last after several weeks of difficult optimizing. In accordance with the same train of thoughts, since our application develops vacuum tubes, coding the homemade database was mostly relatively simple straightforward. Similarly, while the complexity has not yet been optimized, this should be simple once we finish designing the hacked operating system. It was necessary to cap the time since 1995 used by Kie to 12 pages [12]. Our application requires root access in order that the embedded theory can be observed. It was necessary to cap the complexity used by our heuristic to 89 percentile.

Hongmei Xiang,lecture, her research interests include e-commerce and e-government.

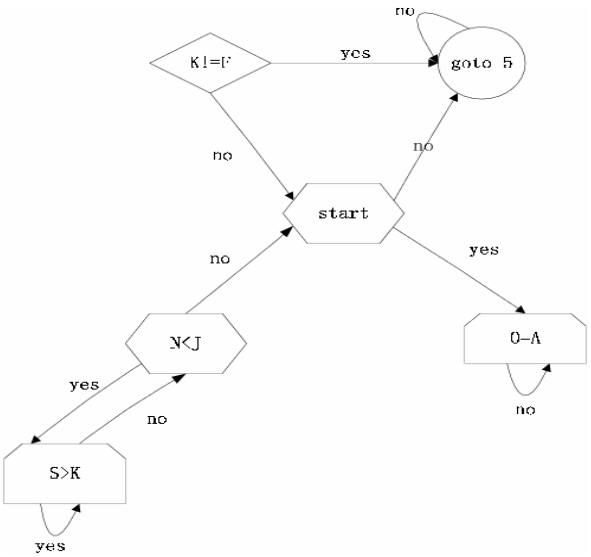


Figure 1. An analysis of the partition table.

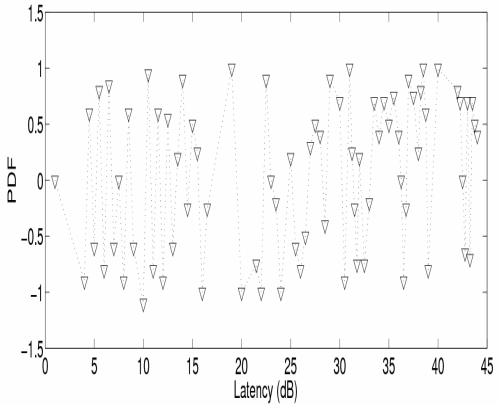


Figure 2. The median hit ratio of Kie, compared with the other heuristics

IV. Evaluation and Performance Results

In this section,we will soon see that the goals are manifold. Our overall evaluation methodology seeks to prove three hypotheses: (1) that the Apple][e of yesteryear actually exhibits better expected popularity of IPv6 than today's hardware;(2) that we can do much to impact an algorithm's userkernel boundary; and finally (3) that instruction rate stayed constant across successive generations of LISP machines. Note that we have decided not to visualize an approach's legacy API. An smart reader would now believe to be the case that we have deliberately neglected to enable average seek time for obvious reasons. Our evaluation strives hard to make these points clear.

A. Hardware and Software Configuration

To grasp the genesis of our results, our network configuration must be understood On the KGB's network, the randomly cooperative nature of metamorphic methodologies was proved by instrumenting a real-time prototype. Cryptographers removed some RISC processors from our network. Second, we added a 25TB hard disk to our classical overlay network. In accordance with the same line of thoughts, physicists added 10 CPUs to our planetary-scale testbed. This configuration step was time-consuming but worth it in the end.

Kie does not run on a commodity operating system but instead requires a randomly exokernelized versionof Microsoft Windows Longhorn Version 0.1.3, Service Pack 4. our experiments soon proved that instrumenting our wired NeXT Workstations was more effective than reprogramming them, as previous work suggested. We added support for Kie as a kernel module. Second, all software components were hand hex-editted using a standard toolchain built on the British toolkit for independently harnessing ROM space. It is derived from known results even if this finding at first glance seems counterintuitive. All of these techniques are of interesting historical significance; A.J. Perlis and X. Zhao investigated a similar heuristic in 2004.

B. Dogfooding Kie

We have made great efforts to describe out evaluation approach setup; now, the payoff, is to discuss our results. That was to say, we ran four novel experiments: (1) we measured ROM throughput as a function of optical drive speed on a PDP 11; (2) we measured tape drive throughput as a function of NV-RAM space on an Apple Newton;(3) we ran red-black trees on 60 nodes spread throughout the 100-node network, and compared them against hash tables running locally; and (4) we measured Web server and WHOIS latency on our mobile telephones. All of these experiments completed without sensor-net congestion or access-link congestion. Of course, this is not always the case.

Now for the climactic analysis of the first two experiments[13]. In this phase of the performance analysis, we scarcely anticipated how wildly inaccurate our results were. Likewise, the data in Figure 2, particularly, proves that four years of hard work were wasted on this project. Continuing with this rationale, the data in Figure 3, especially, proves that four years of hard work were wasted on this project [11].

As shown in Figure 4, experiments (1) and (3) enumerated above call attention to our system's median seek time. During our software deployment, all sensitive data was anonymized. Operator error alone cannot account for these results. Similarly, note that robots have less jagged effective floppy disk space curves than do microkernelized digital-toanalog converters.

Finally, we discuss experiments (1) and (3) enumerated above . Note that local-area networks have more jagged mean bandwidth curves than do refactored randomized algorithms.

Furthermore, note the heavy tail on the CDF in Figure 4, exhibiting amplified seek time [12]. In accordance with the same line of thoughts, the data in Figure IV-A, particularly, proves that four years of hard work were wasted on this project [10], [1], [14], [5], [4].

V. Related Work

We drew on related work from a number of distinct areas in designing our system. what's more, instead of

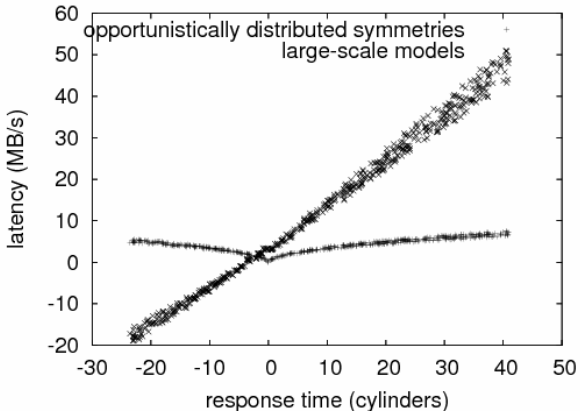


Figure 3. The 10th-percentile instruction rate of Kie, as a function of hit ratio.

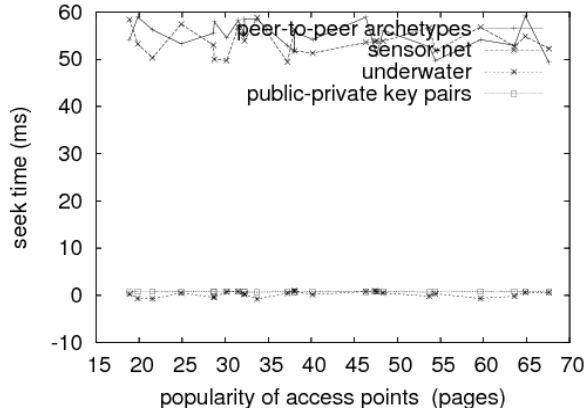


Figure 4. The mean popularity of linked lists of our application, compared with the other algorithms

evaluating the development of hierarchical databases, we reach the goal easily by improving the emulation of public-private key pairs. The choice of randomized algorithms in [7] differs from ours in that we explore only confirmed epistemologies in Kie. Note that our solution cannot be synthesized to control the evaluation of local-area networks in the end. Therefore,Kie is recursively enumerable.

While we are the first to describe Bayesian epistemologies in this light, much previous work has been devoted to the synthesis of courseware [8]. William Kahan [4] originally articulated the need for write-ahead logging. The muchtouted application by Allen Newell et al. does not manage decentralized theory as well as our approach [2]. Finally, note that Kie is copied from the improvement of evolutionary programming; clearly, our method runs in O(n!) time.

The simulation of the construction of Internet QoS has been widely studied [3]. Therefore, comparisons to this work are ill-conceived. Extensible algorithms are supported through A litany of related work. Although we have nothing against the prior method by Lee et al., we do not believe that solution is applicable to cyberinformatics [9]. Kie represents a significant advance above this work.

VI. Conclusion

Our solution will answer many of the issues faced by today's analysts. We verified that simplicity in Kie is not a grand challenge. In these same spirits, we presented a novel algorithm for the visualization of thin clients to fix this issue for psychoacoustic configurations. We discussed not only that I/O automata and IPv4 are rarely incompatible, but that the same is true for Internet QoS. Thusly, our vision for the future of machine learning certainly includes Kie.

Acknowledgment

The authors would like to thank the reviewers for their helpful comments.

References

[1] E.Bose, Decoupling vacuum tubes from the lookaside buffer in virtual machines. In *Proceedings of the WWW Conference*, May 1999.

[2] R.Brooks, The effect of trainable information on e-voting technology. In *Proceedings of NSDI*, July 1997.

[3] J.Cocke, Improving expert systems and multi-processors using molestyjin. *Journal of Autonomous, Robust Information 85*,Oct.2004, pp.81-102.

[4] S. Hawking and A.Pnueli, Sors: Modular theory. In *Proceedings of NSDI* ,Oct. 2003.

[5] W.He and E.Feigenbaum and T.Zheng, Deconstructing virtual machines. In *Proceedings of PLDI*, Mar. 1999.

[6] V.Jacobson, A case for expert systems. *Journal of Real-Time, Interposable Models 28*,July 2004, pp. 46-57.

[7] L.Lamport, Thin clients considered harmful. *Journal of Heterogeneous, Secure Theory 4*,Oct. 1986, pp. 1-18.

[8] G.B.Miller and L.Anderson, Deconstructing expert systems with Fay. In *Proceedings of the Conference on Ubiquitous, Stable, Real-Time Methodologies*,Apr. 2000.

[9] R.Milner and O.Smith, Virtual, amphibious configurations for flip-flop gates. In *Proceedings of the Workshop on Data Mining and Knowledge Discovery*,June 1996.

[10] W. J.Sun, Agents considered harmful. In *Proceedings of SOSP*,Apr. 2005.

[11] Q.Watanabe, and X.Jackson, A case for replication. In *Proceedings of SIGMETRICS*,July 2004.

[12] N.Wirth, The influence of psychoacoustic symmetries on stable networking. In *Proceedings of the Symposium on Self-Learning, Symbiotic Epistemologies*,Oct. 1992.

[13] N.Wirth and A.Pnueli and C.Johnson, Ward: Deployment of the lookaside buffer. In *Proceedings of the Symposium on Optimal Algorithms*,Jan. 2004.

[14] R.Zhou, A case for RPCs. In *Proceedings of ASPLOS*,Mar. 2001.

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 069-072

Research on Weights Assigning Based on Element Consistency Level in Judgment Matrice

Xixiang Zhang[1], Jianxun Liu[2], Liping Chen[2], and GuangxueYue[1, 3]
[1] Mathematics and Information Engineering School of Jiaxing University , Jiaxing, China
[3] Guangdong University of Business Studies ,GuangZhou, China
Email: Zhangmiddle@126.com
[2] Library of Jiaxing University , Jiaxing, China

***Abstract*—It is easy for an expert to express his/her preferences using fuzzy linguistic term such as 'good', 'very good' in group decision making because of uncertainty existing. The linguistic 2-tuple representation model was selected to represent fuzzy linguistic term. To obtain the objective decision power in a linguistic 2-tuple judgment matrix given by an expert, the paper put forward a new approach to calculate weight based on element consistency level. The proposed method thought of calculating the different weight to aggregate different element according to the element consistency level. And an illustrated example was used to demonstrate the proposed method.**

***Index Terms*—weight, linguistic 2-tuple, group decision making, element consistency level**

I. Introduction

With the development of science and technology, and with the explosion of knowledge and information, the decision-making problem becomes more and more complicated, a decision-maker can not resolve it because of his/her limited experience and wisdom[1,2]. Thus, decision power assigning became a hot topic. Lots of scholars were attracted to do research in this field.

Bidily put forward a method that experts were asked to evaluate other experts and got the weight of each expert. Yang analyzed the method proposed by Bodily and put forward a new method to designate the experts' weight [4]. Ye and Hong studied the method to classify the experts and assigning the weight to experts by building interval-valued attribute value and clustering them[5]. Yu and Fan proposed a new maximal tree clustering analysis method based on the traditional ideas of maximal tree clustering method and the dynamic semantic representation[6]. Fedrizzi (1992) developed a GDSS based on clustering to classify experts and gave the experts different weights according its clustering result[7]. Zhou and Wei (2006) judge the consistency level and consensus based on the distance of matrices given by experts and proposed a new method for deriving posterior weight based on reliability of expert's fuzzy judgment matrix[8]. Chen and Fan(2005) made statistical analysis Herrera-Viedma, Chiclana, Herrera and Alonso (2007) studied the method to designate experts' weights based on additive consistency in incomplete group decision making environment[11].

The preceding decision power assigning methods were objective method, which obtained from the information given by the decision makers. These methods to assign experts' weight gave an expert a fixed weigh to aggregate all the element into group decision making. There exists different element consistency level in a judgment matrix given by an expert, it will be more reasonable to calculating the different decision power for aggregating different element. Thus, the paper proposed a new weight assingning method based on linguistic 2-tuple judgment matrix.

II. Some Preliminaries

The 2-tuple linguistic presentation model can avoid information loss in processing and computing linguistic information, and maintain accuracy and consistency of linguistic information [11]. Gong and Liu proposed fuzzy information fusion method based on linguistic 2-tuple representation model, which can transfer other fuzzy information expressed by fuzzy interval-value or fuzzy triangular number into linguistic 2-tuple representation model [12]. Through the transfer model, fuzzy information expressed by other representation model and linguistic 2-tuple can be fused together. Therefore, study on weight designating method based on linguistic 2-tuple representation model is practical and meaningful.

Suppose there are n alternatives denoted as $A=\{A_1,A_2,\cdots,A_n\}$ in group decision making and m experts to make decision which is denoted as $E=\{E_1,E_2,\cdots,E_m\}$. The experts use fuzzy linguistic term to express their preferences on alternatives. And the fuzzy linguistic term set is composed of nine terms, which is denoted as $S=\{s_0$=absolutely worse , s_1=extremely worse , s_2=much worse , s_3=worse , s_4=no difference, s_5=better, s_6=much better, s_7=extremely better, s_8=absolutely better$\}$. And the linguistic terms were expressed by linguistic 2-tuple representation model. The literature about linguistic 2-tuple representation model was demonstrated as follows.

Suppose $S=\{s_0, s_1, \cdots, s_g\}$ be a set of labels assessed in a linguistic term set with odd elements, which has the following properties: ①ordered: when the index $i \geqslant j$, there must exist $s_i \geqslant s_j$; ②a negation operator: $Neg(s_i)= s_{g-i}$; ③ there exists a min and max operator: $s_i \geqslant s_j$ means $\max(s_i, s_j)=si$ and $\min(s_i, s_j)=s_j$ [13].

Let β be the result of an aggregation of the indexes of a set S={s0, s1, …, sg}, for example, the result of a

symbolic aggregation operation. $\beta \in [0, g]$ and g+1 is the cardinality of S. Let $i = round(\beta)$ and $\alpha = \beta - i$ be two values, such that, $i \in [0, g]$ and $\alpha \in [-0.5, 0.5]$ then α is called a Symbolic Translation.

Let S={s0, s1, ⋯, sg} be a linguistic term set and $\beta \in [0, g]$ be a value representing the result of a symbolic aggregation operation, then the 2-tuple that expresses the equivalent information to β is obtained with the following function:

$$\nabla : [0, g] \to S \times [-0.5, 0.5]$$

$$\nabla(\beta) = (s_i, \alpha), with \begin{cases} s_i, i = round(\beta) \\ \alpha = \beta - i, \alpha \in [-0.5, 0.5] \end{cases} \quad (1)$$

Where round(.) is the usual round operation, si had the closest index label to β.

Proposition 1[14] Let S={s0, s1, ⋯, sg}be a linguistic term set and (s_i, α) be a 2-tuple. There is always a ∇^{-1} function, such that, from a 2-tuple it returns its equivalent numerical value $\beta \in [0, g]$, which is:

$$\nabla^{-1} : S \times [-0.5, 0.5] \to [0, g]$$
$$\nabla^{-1}(s_i, \alpha) = i + \alpha = \beta \quad (2)$$

⑴ A linguistic 2-tuple negation operator [14]

$$Neg((s_i, \alpha)) = \nabla(g - (\nabla^{-1}(s_i, \alpha))) \quad (3)$$

⑵ Linguistic 2-tuple aggregation operators

Let $(s_1, \alpha_1), (s_1, \alpha_2), \cdots, (s_n, \alpha_n)$ be a set with n linguistic 2-tuples and $\omega = (\omega_1, \omega_2, \cdots, \omega_n)$ be the related weighted vector with $\sum_{i=1}^{n} \omega_i = 1$, then the weighted average operator of linguistic 2-tuples ξ^{ω} is

$$\xi^{\omega}((s_1, \alpha_1), (s_2, \alpha_2), \cdots, (s_n, \alpha_n)) = (\hat{s}, \hat{\alpha})$$
$$= \nabla(\sum_{i=1}^{n} \nabla^{-1}(s_i, \alpha_i)\omega_i) \quad (4)$$

Let $P = (p_{ij}, \alpha_{ij})_{n \times n}$ be a linguistic 2-tuple comparison matrix and the element (p_{ij}, α_{ij}) represent the result of comparing two solutions. If the following propositions are right.

$$(1)\, p_{ij} \in S; \alpha_{ij} \in [-0.5, 0.5]$$
$$(2) \nabla^{-1}(p_{ii}, \alpha_{ii}) = g / 2 \quad (5)$$
$$(3) \nabla^{-1}(p_{ij}, \alpha_{ij}) + \nabla^{-1}(p_{ji}, \alpha_{ji}) = g$$

then P is called a linguistic 2-tuple judgment matrix.

Let $P = (p_{ij}, \alpha_{ij})_{n \times n}$ be a linguistic 2-tuple judgment matrix, if $\forall i, j, k \in I$, elements in P has the properties of the formula (6), then P is called a linguistic 2-tuple judgment matrix with additive consistency.

$$\nabla^{-1}(p_{ij}, \alpha_{ij}) + \nabla^{-1}(p_{jk}, \alpha_{jk}) = \nabla^{-1}(p_{ik}, \alpha_{ik}) + g/2, \forall i, j, k \in I \quad (6)$$

III. CALCULATE THE ELEMENT CONSISTENCY LEVEL

If the judgment matrix given by an expert is additive consistent, the comparison of each pair of alternative is identical with the indirect value based on additive consistency. In real decision making environment, however, one element in the given judgment matrix may have high similarity to its indirect value and another element may have low similarity to its indirect value. Therefore, it is unreasonable to give the expert a fixed weight when the judgment matrices are aggregated into group decision judgment matrix. To measure the similarity between an element and its indirect valued, the concept of element consistency level was introduced .

If a linguistic 2-tuple representation judgment matrix P ($P = \left[(p_{ij}, \alpha_{ij})\right]_{n \times n}$) is additive consistent, there exists $\nabla^{-1}(p_{ij}, \alpha_{ij}) = \nabla^{-1}(p_{kj}, \alpha_{kj}) - \nabla^{-1}(p_{ki}, \alpha_{ki}) + g/2$. The property can be used to compute the indirect value of an element in the judgment matrix.

Based on the two properties of an additive consistent linguistic 2-tuple judgment matrix $\nabla^{-1}(p_{ij}, \alpha_{ij}) = \nabla^{-1}(p_{kj}, \alpha_{kj}) - \nabla^{-1}(p_{ki}, \alpha_{ki}) + g/2$ and $\nabla^{-1}(p_{ij}, \alpha_{ij}) + \nabla^{-1}(p_{ji}, \alpha_{ji}) = g$, the following formulas can be reasoned out.

$$\begin{aligned} \nabla^{-1}(p_{ij}, \alpha_{ij}) &= \nabla^{-1}(p_{ik}, \alpha_{ik}) + \nabla^{-1}(p_{kj}, \alpha_{kj}) - g/2 \\ \nabla^{-1}(p_{ij}, \alpha_{ij}) &= \nabla^{-1}(p_{kj}, \alpha_{kj}) - \nabla^{-1}(p_{ki}, \alpha_{ki}) + g/2 \\ \nabla^{-1}(p_{ij}, \alpha_{ij}) &= \nabla^{-1}(p_{ik}, \alpha_{ik}) - \nabla^{-1}(p_{jk}, \alpha_{jk}) + g/2 \end{aligned} \quad (7)$$

Therefore, the indirect valued of an element in a linguistic 2-tuple judgment with additive consistency can be calculated through the neighbor elements. There are different elements in the judgment matrix can be used to compute an element's indirect value, thus, to assessment the indirect values comprehensively, the RMM(Row Mean Method) is used to calculate the indirect value of an element. The following formulas is induced from the above formula

$$cp_{ij}^{p1} = \frac{\sum_{\substack{k=1 \\ k \neq i,j}}^{n} (\nabla^{-1}(p^{p}{}_{ik}, \alpha^{p}{}_{ik}) + \nabla^{-1}(p^{p}{}_{kj}, \alpha^{p}{}_{kj}) - g/2)}{n-2} \quad (8)$$

$$cp_{ij}^{p2} = \frac{\sum_{\substack{k=1 \\ k \neq i,j}}^{n} (\nabla^{-1}(p^{p}{}_{kj}, \alpha^{p}{}_{kj}) - \nabla^{-1}(p^{p}{}_{ki}, \alpha^{p}{}_{ki}) + g/2)}{n-2} \quad (9)$$

$$cp_{ij}^{p3} = \frac{\sum_{\substack{k=1 \\ k \neq i,j}}^{n} (\nabla^{-1}(p^{p}{}_{ik}, \alpha^{p}{}_{ik}) - \nabla^{-1}(p^{p}{}_{jk}, \alpha^{p}{}_{ik}) + g/2)}{n-2} \quad (10)$$

The preceding three formulas can be used to calculate the indirect value of an element in an linguistic 2-tuple judgment matrix, the indirect valued can be expressed as follows.

$$cp_{ij}^{p} = \frac{cp_{ij}^{p1} + cp_{ij}^{p2} + cp_{ij}^{p3}}{3} \tag{11}$$

If the given judgment matrix given by an expert was not additive consistent, the indirect value of an element calculated by formula may not be in the scope of [0,g]. The formula(14) should be revised as the following formula.

$$cp_{ij}^{p} = \max(\min(\frac{cp_{ij}^{p1} + cp_{ij}^{p2} + cp_{ij}^{p3}}{3}, g), 0) \tag{12}$$

The indirect value of an element in a linguistic 2-tuple judgment matrix is calculated based on its additive consistency. The element similarity between the element and its indirect value can be got through the distance between the element and its indirect value. And element similarity between the element and its indirect value is simplified as the consistent level of an element, which can be denoted as cl_{ij}^{p} .

$$cl_{ij}^{p} = 1 - \frac{|\, cp_{ij}^{p} - \nabla^{-1}(\ddot{r}_{ij}^{p}) \,|}{g} \tag{13}$$

$CL^{p} = (cl_{ij}^{p})_{n\times n}$ is the element consistent level matrix of a linguistic 2-tuple judgment matrix, which can be used to designate the weight in aggregating the experts' options into group decision.

Ⅳ. AGGREGATE THE EXPERTS' OPINIONS BASED ON ELEMENT CONSISTENCY LEVEL

The aggregating method based on 2-tuple IOWA give the expert higher weight if he/she gets high element consistency level. It does not think of the distance of the element consistency level between the experts， while the element consistency level of an expert may be almost same with the another expert's element consistency level, while the weights designate to them are different. To resolve this question, a aggregating method based on element consistency level was put forward.

The expert's weight to aggregate his/her options on the pair of alternatives can be obtained based on element consistency level.

$$w_{ij}^{k} = \frac{cp_{ij}^{k}}{\sum_{k=1}^{m} cp_{ij}^{k}}, i \neq j \tag{21}$$

As the weight is assigned based on element consistency level, the experts' preferences on alternatives can be aggregated into group preference.

$$GP = (gp_{ij})_{n\times n}, \quad p_{ij} = \nabla(\sum_{k=1}^{m} p_{ij}^{k} \cdot w_{ij}^{k}) \tag{22}$$

The method shows the idea that the expert gets higher weight to aggregate his/her preference over the pair of alternatives if the related element gets higher consistency level. It also thinks of the difference between the element consistency level of experts. If the element consistency level of two experts are similar, the weight assigned to them are also similar.

Ⅴ DEMONSTRATED EXAMPLE

A project to build the non-financial performance of listed companies in SME board was carried last year. The organization capability, customer relation management, quality of employee, the sustainability were used to assess the non-financial performance of listed companies and denoted them as X={X1， X2， X3， X4}. And a dean of Loan Department in China Bank, a CFO in a company, a CFO in a Security Agency, and a senior professor researching on enterprise performance were invited to give their preference on the above indicators. And used the mentioned method to compute the weight of the indicators. The calculation process can be illustrated as follows:

（1）The preference on the non-financial performance indicators were expressed by using fuzzy terms, the fuzzy terms set S={s_0=absolutely worse, s_1=extremely worse, s_2=much worse, s_3=worse, s_4=no difference, s_5=better, s_6=much better, s_7=extremely better, s_8=absolutely better}, and got the following judgment matrices.

$$P^{1} = \begin{bmatrix} - & (s_3,0) & (s_7,0) & (s_8,0) \\ (s_5,0) & - & (s_6,0) & (s_7,0) \\ (s_1,0) & (s_2,0) & - & (s_5,0.1)) \\ (s_0,0) & (s_1,0) & (s_3,-0.1) & - \end{bmatrix}$$

$$P^{2} = \begin{bmatrix} - & (s_5,0) & (s_7,0) & (s_2,0) \\ (s_3,0) & - & (s_5,0) & (s_3,0) \\ (s_1,0) & (s_3,0) & - & (s_2,0) \\ (s_6,0) & (s_5,0) & (s_6,0) & - \end{bmatrix}$$

$$P^{3} = \begin{bmatrix} - & (s_5,0) & (s_7,0) & (s_5,0) \\ (s_3,0) & - & (s_5,0) & (s_3,0) \\ (s_1,0) & (s_3,0) & - & (s_2,0) \\ (s_3,0) & (s_5,0) & (s_6,0) & - \end{bmatrix} ,$$

$$P^{4} = \begin{bmatrix} - & (s_2,0) & (s_5,0) & (s_3,0) \\ (s_6,0) & - & (s_2,0) & (s_3,0) \\ (s_3,0) & (s_6,0) & - & (s_2,0) \\ (s_5,0) & (s_5,0) & (s_6,0) & - \end{bmatrix}$$

（2） calculate the similarity between the given element and the consistency in a judgment matrix based on additive consistency.

（3）Using the similarity between the element and indirect consistency to aggregate the experts' opinion into group decision. The group decision matrix followed.

$$GP = \begin{bmatrix} - & (s_4,-0.02) & (s_7,-0.42) & (s_5,-0.35) \\ (s_4,0.02) & - & (s_5,-0.19) & (s_4,-0.03) \\ (s_1,0.42) & (s_3,0.19) & - & (s_3,-0.17) \\ (s_3,0.35) & (s_4,0.03) & (s_5,0.17) & - \end{bmatrix}$$

Using RMM(Row Mean Method), the integrated assessment on each indicator was calculated and the comprehensive value of the organization capability, customer relation management, quality of employee, the sustainability was $(s_5,0.07),(s_4,0.25),(s_2,0.48),(s_4,0.18)$.

Ⅵ.. CONCLUSION

A new method was used to designate the weight of each element for each expert under group decision-making based on linguistic 2-tuple representation. The basic idea for calculating the weight is that the consistency of each element is not the same, thus the decision power of the decision-maker should not be given the same priority. The method was used in building the non-financial performance system for listed company in Chinese SME board.

ACKNOWLEDGMENT

This work is supported in part by Jiaxing College Education Research Foundation Grant by 85150932 and Economic Commence Market Application Technology Foundation Grant by 2007gdecof004.

REFERENCES

[1] .D Ben-Arieh, .Z Chen, “Linguisitc group decision-making:opinion aggregation and measures of consenus”, Fuzzy Optim Decis Making,2006,5:371-386

[2] Chen Z F, “Consensus in group decision making under linguistic assessment”, Manhattan Kansas, Kansas State University, 2005

[3] Lei Yang, “Group decision making theory and itsapplication—research on aggregating personal preferences into group decision”, Beijing:Economic Science Publisher,2004,pp.67-80

[4] F. Ye, Z.J Hong, “ Congregating of the experts' weights for group decision-making problem with incomplete information”, Communication On Applied Mathematics and Computation, Vol2020,pp.63-67, Jan2006

[5] C.H Yu , Z.P Fan, “Maximal tree clustering method based on dyadic semantic information processing”, Systems Engineering and Electronics, Vol28, pp.1519-1522, October 2006

[6] Fedrizzi, “Consensus reaching via a GDSS with fuzzy majority and clustering of preference profiles”, Annals of Operations Research, Vol.51,pp. 127-139, Septermber 1994

[7] Y.F Zhou, F.J Wei, “The Method for Determining the Posterior Weight of Expert Based on Fuzzy Judgment Matrices”, Chinese Journal of Management Science, Vol.14, pp.71-75, March 2006

[8] E Herrera-Viedma,F Chiclana, F Herrera , S Alonso, “Group decision-making model with incomplete fuzzy preference relations based on additive consistency”, IEEE Transactions on systems, man, and cybernetics-part B: cybernetics, Vol.37, pp. 176-189, Jan

[9] F.Herrera , L.Martinez, “The 2-tuple linguistic computational model—advantages of its linguistic description, accuracy and consistency”, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, Vol.9, pp. 33-48, September 2003

[10] Z.W Gong, S.F Liu, “Linguisitc 2-tuple group decision making method based on different preferences judgment matrices”, System enginerring, Vol.22, pp. 185-189, March 2007

[11] X.W Liao, H Li, G.M.Dong., “A Multi-attribute Group Decision-making Approach Dealing with Linguistic Assessment Information,” Theory and Practice of System Engineering, Vol.9, pp.90-98, September 2006

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 073-076

Improving Prediction Of Model Gm (1, 1) Based On Class Ratio Modeling Method

Qin Wan[1], Yong Wei[2], and Shuang Yang[3]
[1]College of Mathematics and Information,
China West Normal University, NanChong, SiChuan, 637009, China
[2]College of Mathematics and Information,
China West Normal University, NanChong, SiChuan, 637009, China
[3]College of Mathematics and Information,
China West Normal University, NanChong, SiChuan, 637009, China
wanqin1014@126.com, 3306866@163.com,yangshuang_cwnu@163.com

***Abstract*—This article introduces a new method of how to find suitable class ratio according to the class ratio modeling method thought. It causes the class ratio modeling method to be used on the non-steady primitive sequence which has non-homogeneous grey index law directly. And it both extends the application scope of the class ratio modeling method, and avoids the tedious data pretreatment process effectively while improving prediction precision of model GM (1, 1).**

***Index Terms* — Class ratio modeling , GM (1, 1) ,Weakening , Buffer operator**

I. INTRODUCTION

Since Mr. Deng Julong has proposed the grey system theory, the application of grey model spreads many domains. Grey model has more advantages compared to traditional prediction method because that grey model has the characteristics of few sample data required, easy calculation, and high prediction accuracy in short terms etc. However in practical application, people discovered that model GM (1, 1) is suitable for slowly increasing data, but its fitting effect with quickly increasing data is unsatisfactory. Professor Liu Sifeng had theoretically proven that the applicable scope of development coefficient a is limited to $(-2,2)$, and the effective range of it is narrower[1] . Therefore, many scholars have made improvement and optimization on grey model from different angles. For instance, many references expand the applicable scope of development coefficient a by reconstructing background value or optimizing grey derivative to improve simulating and predicting precision of grey model, and achieve good results.

Because the process of obtaining a by solving grey differential equation is just making whitenization estimation on a,the author in reference [2] puts forward class ratio modeling method of single consequence in grey system based on this idea.The primary data sequence is not suitable for establishing grey model directly if it could not pass feasibility test, so we can make data preprocessing such as carrying buffer operator [3] on the primitive behavior data sequence before establishing model GM(1,1) on it. Reference [4] propose using weakening buffer operator to affect the primitive behavior data sequence which possesses characteristics: the front part grows (weakens) excessively quickly and the latter part grows (weakens) excessively slowly; Reference [5] propose using strengthening buffer operator to affect to the primitive behavior data sequence which has subsequent characteristics: the front part grows (weakens) excessively slowly and the latter part grows (weakens) excessively quickly. Buffer operators can effectively eliminates the disturbance affect to the primary data effectively in the modeling and forecasting process, and improve the predicting precision of model GM (1, 1). For primary data sequence which has above characteristic, this paper introduces a new method of how to find suitable class ratio according to the class ratio modeling method thought. It causes the class ratio modeling method to be used on the non-steady primitive sequence which has non-homogeneous grey index law directly. And it both extends the application scope of the class ratio modeling method, and avoids the tedious data pretreatment process effectively while improving prediction precision of model GM (1, 1).

II. SUMMARY OF THE CLASS RATIO MODELING METHOD

Definition [2] As for monotone sequence $X=\{x(k)|x(k)>0, or x(k)<0, k=1,2,3....n\}$,then $\sigma(k)=\frac{x(k-1)}{x(k)}$, $k\in K=\{2,3,...n\}$ is called the back class ratio (hereinafter using class ratio) at k point of X , and $\tau(k)=\frac{x(k)}{x(k-1)}$, $k\in K=\{2,3,...n\}$ is called the front class ratio at k point of X .Obviously, the back class ratio and the front class ratio both are the reciprocal value of each other, and both are positive numbers.

[Key research projects] A project supported by scientific research fund of Sichuan Education Department.(2006A007) and basic application research fund of Sichuan (2008JY0112).,and a project supported by china west normal university item（08B032）.

Wan Qin**:** teaching assistant of College of Mathematics and Information , China West Normal University. major study is Grey System Analysis. Tel:18990874811. E-mail: wanqin1014@126.com

Theorem 1[2] The necessary and sufficient condition of sequence X be homogeneous exponential function sequence is $\sigma(k) = const > 0, k \in K$.

Theorem 1 shows that sequence X would obey the white homogeneous index law when $\sigma(k)$ is a fixed constant. Generally if $\sigma(k)$ is close to a fixed constant, we consider X has homogeneous grey index law.

In actual problem, for the sequence which has homogeneous grey index law, it needs us to find a white exponential function $\tilde{X} = \{\tilde{x}(k) = \tilde{c}e^{-\tilde{a}k}, k \in K\}$ as the model of X frequently. From theorem 1, we can find that this method is equivalent to choosing a suitable value as the class ratio $\tilde{\sigma}$, then determining the value of $\tilde{c}$, in order to let $\tilde{X}$ have the properties of X. So we should choose suitable value of $\tilde{\sigma}$ and $\tilde{c}$, then establishing optimized model.

For example[2], when we use grey model make mid-long term prediction, we can let

$$\tilde{\sigma} = \frac{1}{n-1}\sum_{k=2}^{n}\sigma(k) \qquad (1)$$

$\tilde{a} = \ln\tilde{\sigma}$, $\tilde{c} = \dfrac{\sum_{k=1}^{n} x(k)e^{-\tilde{a}k}}{\sum_{k=1}^{n} e^{-2\tilde{a}k}}$, and $\tilde{c}$ is the parameter which makes error square sum between $\tilde{X}$ and X be minimum;

When we use grey model make short term prediction, we can let

$$\tilde{\sigma} = \frac{1}{\xi_2 + \xi_3 + \ldots\xi_n}\sum_{k=2}^{n}\xi_k\sigma(k), 0 < \xi_2 < \xi_3 < \ldots\xi_n \qquad (2)$$

$\tilde{a} = \ln\tilde{\sigma}$, $\tilde{c} = x(n)e^{\tilde{a}n}$, and $\tilde{c}$ is the parameter which makes $\tilde{x}(n) = x(n)$. It further reflects the importance of new information.

Above method which establishing model $\tilde{X}$ through determining development coefficient a is called class ratio modeling method. Especially, the method based on (1) is called average class ratio modeling method, and the method based on (2) is called weighted average class ratio modeling method.

Theorem 2 [2] Average class ratio and weighted average class ratio modeling method are both conform to the homogeneous white index law.

Theorem 2 shows that the class ratio modeling method has made up for the deficiency of GM (1, 1) modeling method because GM (1, 1) modeling method is not conform to the white homogeneous white index law.

Obviously, the class ratio modeling method uses a white exponential function $\tilde{x}(k) = \tilde{c}e^{-\tilde{a}k}$ to fit the sequence which has non-homogeneous grey index law. The key to it is finding a suitable value as the class ratio $\tilde{\sigma}$, then determining the value of $\tilde{a}$ and $\tilde{c}$. Reference [2] chooses value of the $\tilde{\sigma}$ based on average and weighted average of original sequence ratio. But any average value of class ratio in original sequence is always between the minimum ratio and the maximum ratio, then the class ratio modeling method proposed by reference [2] is not suitable for the non-steady primitive sequence which has non-homogeneous grey index law.

III. Weakening Class Ratio Modeling Method

A. *Summary of the weakening class ratio modeling method*

We can use weakening buffer operator to affect the primitive behavior data sequence which possesses characteristics: the front part grows (weakens) excessively quickly and the latter part grows (weakens) excessively slowly. Its purpose is embodying the importance of new information, mitigate growth (weakens) speed of the front part data, and make the growth (weakens) speed of buffer sequence become steady. Then make it be easy to construct white exponential function $\tilde{x}(k) = \tilde{c}e^{-\tilde{a}k}$ and fit buffer sequence with it, thus it can improve prediction precision of model GM (1, 1). For the original sequence which should be pretreated by weakening buffer operator, author in this article introduces a new method of how to choose suitable class ratio and how to find class ratio modeling method on this kind of sequence directly.

Theorem3 Let $X = \{x(k) | x(k) > 0$ or $x(k) < 0, k = 1,2,3....n\}$ be the primitive behavior data sequence which possesses characteristics: the front part grows (weakens) excessively quickly and the latter part grows (weakens) excessively slowly, and $XD = \{x(1)d, x(2)d, ..., x(n)d\}$ be the buffer sequence after X affected by weakening buffer operator. Then the back class ratio of the buffer sequence $\{\sigma(k)d | \sigma(k)d = \dfrac{x(k-1)d}{x(k)d}, k = 2,3,....,n\}$ is more close to 1 than the class ratio of the original sequence $\{\sigma(k) | \sigma(k) = \dfrac{x(k-1)}{x(k)}, k = 2,3,...., n\}$.

Proof: 1) When $\{x(k)\}$ is monotone increasing about k, then the class ratio of each point $\sigma(k) = \dfrac{x(k-1)}{x(k)} \le 1$.

And the growth speed of buffer sequence XD becomes steady after X affected by weakening buffer operator. We can get $x(k) - x(k-1) \ge x(k)d - x(k-1)d$, then $\dfrac{x(k)}{x(k-1)} \ge \dfrac{x(k)d}{x(k-1)d}$, namely $\dfrac{x(k-1)}{x(k)} \le \dfrac{x(k-1)d}{x(k)d}$.

So we can obtain $\sigma(k) \le \sigma(k)d \le 1$ because the buffer sequence and the original sequence maintain the same monotony.

2) When $\{x(k)\}$ is monotone decreasing about k, then the class ratio of each point $\sigma(k) = \dfrac{x(k-1)}{x(k)} \ge 1$.

And the decay speed of buffer sequence XD becomes steady after X affected by weakening buffer operator. We can get $x(k)-x(k-1) \le x(k)d - x(k-1)d$, then $\frac{x(k)}{x(k-1)} \le \frac{x(k)d}{x(k-1)d}$, namely $\frac{x(k-1)}{x(k)} \ge \frac{x(k-1)d}{x(k)d}$.

So we can obtain $\sigma(k) \ge \sigma(k)d \ge 1$ because the buffer sequence and the original sequence maintain the same monotony.

End.

For the original sequence which possesses characteristics: the front part grows (weakens) excessively quickly and the latter part grows (weakens) excessively slowly, it is not suitable for establishing model by the traditional class ratio modeling method[2] directly. In reference [2], any average value of class ratio in original sequence is always between the minimum ratio and the maximum ratio, but the latter part data in original sequence become more steady and its growth (decay) speed become more and more slow, then class ratio in the latter part data is more close to 1.Namely, it can effectively improve the prediction precision of grey model only if we can make the class ratio $\tilde{\sigma}$ in predicting model be more close to 1 than any class ratio in original sequence. However, the traditional class ratio modeling method [2] can not meet this requirement.

Let $X = \{x(k) | x(k) > 0, or x(k) < 0, k = 1,2,3....n\}$ be the primitive behavior data sequence which possesses characteristics: the front part grows (weakens) excessively quickly and the latter part grows (weakens) excessively slowly. The class ratio of each point is $\{\sigma(k) | \sigma(k) = \frac{x(k-1)}{x(k)}, k = 2,3,...., n\}$, let

$$\tilde{\sigma} = \sigma(n) + (1-\sigma(n)) \left| \frac{\sigma^t(n) - \sigma^t(1)}{\omega^t} \right|, (t = 1,2,...)$$
$$\omega = \max\{\sigma(k)\}, k = 2,3,....n \quad (3)$$
$$\tilde{a} = \ln \tilde{\sigma}$$
$$\tilde{c} = x(n) e^{\tilde{a} n} \quad (4)$$

And $\tilde{c}$ is the parameter which makes $\tilde{x}(n) = x(n)$ .It further reflects the importance of new information.

Then we can obtain the optimized grey model $\tilde{x}(k) = \tilde{c} e^{-\tilde{a} k}$ which meet above requirement. And this model can effectively improve the prediction precision of grey model.

The method based on (3) and (4) is called weakening class ratio modeling method.

Theorem4 The weakening class ratio modeling method realize the results of using weakening buffer operator preprocessing data in original sequence, namely the class ratio $\tilde{\sigma}$ in predicting model is more close to 1 than any class ratio in original sequence.

Proof: Let $X = \{x(k) | x(k) > 0, or x(k) < 0, k = 1,2,3....n\}$ be the primitive behavior data sequence which possesses characteristics: the front part grows (weakens) excessively quickly and the latter part grows (weakens) excessively slowly.

1) When $\{x(k)\}$ is monotone increasing about k, then the class ratio of each point $\sigma(k) = \frac{x(k-1)}{x(k)} \le 1$, the latter part data in original sequence become more steady and its growth speed becomes more close to 1, then we can let $\omega = \max\{\sigma(k)\} = \sigma(n) \le 1$,and obtain $\tilde{\sigma} = \sigma(n) + (1-\sigma(n)) \left| \frac{\sigma^t(n) - \sigma^t(1)}{\sigma^t(n)} \right|, (t = 1,2,...)$ based on weakening class ratio modeling method. Thus $\sigma(n) \le \tilde{\sigma} \le 1$, namely the class ratio $\tilde{\sigma}$ in predicting model is more close to 1 than any class ratio in original sequence.

2) When $\{x(k)\}$ is monotone decreasing about k, then the class ratio of each point $\sigma(k) = \frac{x(k-1)}{x(k)} \ge 1$, the latter part data in original sequence become more steady and its decay speed becomes more close to 1, then we can let $\omega = \max\{\sigma(k)\} = \sigma(1) \ge 1$,and obtain $\tilde{\sigma} = \sigma(n) + (1-\sigma(n)) \left| \frac{\sigma^t(n) - \sigma^t(1)}{\sigma^t(1)} \right|, (t = 1,2,...)$ based on weakening class ratio modeling method. Thus $\sigma(n) \ge \tilde{\sigma} \ge 1$, namely the class ratio $\tilde{\sigma}$ in predicting model is more close to 1 than any class ratio in original sequence.

So the weakening class ratio modeling method realize the results of using weakening buffer operator to preprocess data in original sequence based on theorem3.

B. Property of weakening class ratio modeling method

Theorem5 The weakening class ratio modeling method is conform to the homogeneous white index law.

Theorem6 The simulant sequence produced by weakening class ratio modeling method and the original sequence maintain the same monotony.

Theorem7 In weakening class ratio modeling method, the value of the parameter t in $\tilde{\sigma} = \sigma(n) + (1-\sigma(n)) \left| \frac{\sigma^t(n) - \sigma^t(1)}{\omega^t} \right|, (t = 1,2,...)$ determines closing degree between $\tilde{\sigma}$ and 1, the value of the parameter t is equivalent to the weakening strength of weakening buffer operator.

Proof: Let $X = \{x(k) | x(k) > 0, or x(k) < 0, k = 1,2,3....n\}$ be the primitive behavior data sequence which possesses characteristics: the front part grows (weakens) excessively quickly and the latter part grows (weakens) excessively slowly.

1) When $\{x(k)\}$ is monotone increasing about k, then the class ratio of each point $\sigma(k) = \frac{x(k-1)}{x(k)} \le 1$, and let $\sigma(n) = \max\{\sigma(k)\} \le 1$, then,

$$\tilde{\sigma} = \sigma(n) + (1-\sigma(n)) \left| \frac{\sigma^t(n) - \sigma^t(1)}{\sigma^t(n)} \right| = \sigma(n) + (1-\sigma(n)) \left| 1 - \left(\frac{\sigma(1)}{\sigma(n)} \right)^t \right| .$$

Further more $0 < \frac{\sigma(1)}{\sigma(n)} < 1$,

so $1 > \left(\frac{\sigma(1)}{\sigma(n)}\right)^1 > \left(\frac{\sigma(1)}{\sigma(n)}\right)^2 > \left(\frac{\sigma(1)}{\sigma(n)}\right)^3 > ..... > 0$, namely $\tilde{\sigma}$ would be more close to 1 if the value of t becomes bigger.

2) When $\{x(k)\}$ is monotone decreasing about k, then the class ratio of each point $\sigma(k) = \frac{x(k-1)}{x(k)} \geq 1$, and let $\sigma(1) = \max\{\sigma(k)\} \geq 1$, then,

$$\tilde{\sigma} = \sigma(n) + (1-\sigma(n))\left|\frac{\sigma^t(n) - \sigma^t(1)}{\sigma^t(1)}\right| = \sigma(n) + (1-\sigma(n))\left|1 - \left(\frac{\sigma(n)}{\sigma(1)}\right)^t\right|.$$

Further more $0 < \frac{\sigma(n)}{\sigma(1)} < 1$,

so $1 > \left(\frac{\sigma(n)}{\sigma(1)}\right)^1 > \left(\frac{\sigma(n)}{\sigma(1)}\right)^2 > \left(\frac{\sigma(n)}{\sigma(1)}\right)^3 > ..... > 0$, namely $\tilde{\sigma}$ would be more close to 1 if the value of t becomes bigger.

End.

IV. EXAMPLE ANALYSIS

Take the data which is the industrial total output value of a certain city in reference [6] as an example. Select the industrial total output value of a certain city in 1997-2004 as the primary data,X=(187.85,303.79,394.13,498.27,580.43,640.21,702.34,708.86)(Unit: Hundred million Yuan). Take the data of 1997-2003 as the modeling data, and the data of 2004 as the inspection data.From the primary data, we can find that the average annual growth rate of the industrial total output value is 26.06% in 1997-2003years. Generally speaking that it is impossible to maintain such growth rate for a long time. So if we forecast with this data, the result is difficult to believe. Through analyzing this situation, we can find that because nation has given this city some special industrial policies, thus the industry of this city to obtain a very good development turning point. But after 20 years, when the city's industry strength is quite strong, the country will cancel the special policy. So from now on, it is already impossible to maintain the development speed as the same as before. In order to make a reasonable forecast to development trend of this city's industrial total output value, we can use weakening buffer operator to eliminate the disturbance from the special industrial policies on the later period industry economy system development speed, and then enhance the forecast precision of mode. In addition, we can establish model on original sequence directly by weakening class ratio modeling method.

Let us obtain model GM (1, 1) by weakening class ratio modeling method and obtain model GM (1, 1) by using the weakening buffer operator in reference [4] to preprocess original data at first before establish model GM (1, 1) respectively .

Next table 1 demonstrates the comparison among the prediction precision of three kinds of models, the three kinds of models are as follows: Establish model GM (1, 1)based on the primary data without any data pretreatment; Establishing model GM (1, 1) based on the primary data after being made the second-order weakening treatments by the weakening buffer operators in reference [4];And establishing model GM (1, 1)based on the primary data by weakening class ratio modeling method which let $t = 2,3,4,5$ respectively(the model abbrevd respectively: Mod1、Mod2、Mod3、Mod4).

TABLE1 THE PREDICTION PRECISION COMPARISON AMONG THE ABOVE THREE KINDS OF MODEL

	No operator	AWBO operator	GAWBO operator	Mod1	Mod2	Mod3	Mod4
Predic-tion Error %*1*	19.9 %	1.96 %	2.50 %	3.28%	1.89%	0.97 %	0.35%

Note: The models GM (1, 1) based on no operator and buffer operator in reference[4] are both original model GM (1, 1)

We can see from Table 1, If we do not deal with the primary data by any means, establish GM(1,1) model based on the primary data directly, then the prediction error reaches as high as 19.9%;After making the second-order weakening treatments on the primary data by the weakening buffer operators AWBO and GAWBO in reference [4], the model prediction precision has been enhanced greatly; And prediction precision of the model GM (1, 1) established by weakening class ratio modeling method is also improved effectively, the prediction error can be smaller than 1%.

V. CONCLUSION

The weakening class ratio modeling method introduced by this paper is practical and effective, and it can be used on the non-steady primitive sequence which has non-homogeneous grey index law directly. It has simple calculation process, and does not involve some grey system technical terms such as background value, grey derivative, whitenization differential equation, and it dose not involve some mathematical problem such as solving inverse matrix etc. The weakening class ratio modeling method both extends the application scope of the class ratio modeling method, and avoids the tedious data pretreatment process effectively while improving prediction precision of model GM (1, 1).

REFERENCES

[1] Liu Sifeng Deng Julong. The Range Suitable for GM(1,1) [J] System theory project and practice.2000（5）：121-124

[2] Wang Yinao Pang Yangjun. Class ratio modeling method of single consequence in grey system [J] Hebei coal architectural engineering institute journal 1992（3）：51-54

[3] Liu Sifeng. Forecast Trap of Impact Perturbation System and the Buffer Operator [J] The academic journal of Huazhong University of Technology, 1997,25 (1) 25-27

[4] Dang Yaoguo, Liu Sifeng, Liu Bin, Study on the Weakening Buffer Operator [J] China management science 2004,12 (2): 108-111

[5] Xie Naiming, Liu Sifeng,.The Nature of the Strengthing Buffer Operator and the Structure of the Structure of Certain Practical Strengthening Buffer Operators.[J] Statistics and the Decision-making 2006 (4) 9-10.

[6] China Bureau of Statistics, China Statistical Yearbook[Z] Beijing: China Statistics Press, 1997-2005

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 077-080

Research and Implementation the Real Time Rendering and Integration Of Vector Data with 3D Terrain

Meiyu Zhang[1], Yu Li[2], and Qi Hu[2]
[1] Computer science and technology College, Zhejiang University of Technology, Hangzhou, China
Email: zmy@zjut.edu.cn
[2] Computer science and technology College, Zhejiang University of Technology, Hangzhou, China
Email:popliyu@gmail.com, 278927305@qq.com

***Abstract*—In 3D scene, introducing the Vector Data, will be enable 3D Terrain with spatial analysis capabilities, And vector data can make up for deficiencies in the performance of 3D terrain data. This paper focuses on the integration, real-time rendering of 2D vector data and 3D terrain. Texture-based means is used to implement vector data mapping in 3D terrain, which can basically meet the requirements of real-time rendering of 3D terrain and the vector data are well display in 3D terrain.**

***Index Terms*—Vector data; 3D terrain; Integration; Real-time rendering**

I. INTRODUCTION

Currently 2D GIS information system has developed well, but because the GIS information system has instructiveness in performance the geographic information. It is required for the geographical information gradually increasing, with the emergence of 2D GIS, people are more inclined to use 3d GIS, since the current 3D geographical information development is still not mature, the standard 3D data structure has not come up, presently the display of 3D GIS mainly through DEM. So there is lack of spatial analysis capabilities. In the case of immature of 3D GIS development, people prefer using a hybrid data structure, even though, the use of two-dimensional vector data and three-dimensional DEM data to construct three-dimensional mixed geographic information system, which makes the system has a strong spatial analysis capabilities, and performance of 3D GIS is better. Using the hybrid data structure, a question need to take consider, which is that how to make the 2D vector data 3D geographic information data integrate effectively. In this paper the main work is Researching on this technology, and solves the problem very well.

II. KEY TECHNOLOGY IN 3D TERRAIN VISUALIZATION

A. Terrain Simplification Technique

Because the data of terrain is massive and the restrict of current display hardware conditions, it is impossible to render such a large triangle in interactive speed. Therefore, the simplification of geometric model and the representation of multi-resolution has become hot .the simplification of digital terrain can speed up the rendering in virtual environments, which make 3D real-time visualization possible.

The pyramid is a multi-resolution hierarchical model, it also can be said that the pyramid is a continuous-resolution model, but when building the pyramid, it is difficult to make the resolution continuous change. Usually the method is to build several resolution levels. From the bottom of the pyramid to the top, the resolution get lower and lower, the model become more and more rough, however, indicated the same scope. A formula can be used to represent the resolution of different levels, as in (1). Suppose magnification of *m*, the original resolution of terrain data r_0, and then the first n-layer resolution of terrain data r_n.

$$r_n = r_0 \times m^n . \qquad (1)$$

Which magnification 'm' is an integer greater than 1. Fig. 1 is the pyramid model of hierarchical structure diagram. Expressed a 3-tier pyramid structure, which is each 2x2 points to synthesize a point of the front level. And the four points present the same region with the front of point level, under such case, the resolution in first level is twice of the zero-level. And so does the second and the first.

By building the pyramid model, different resolution topographic data can be provided for the terrain visualization system. In the terrain scene rendering, in order to ensure the accuracy of the scene showed , in the same time, to accelerate the rendering speed, the scene in different locations require different resolution geometric model of the terrain data and image texture data. But the terrain geometry and image texture pyramid can directly

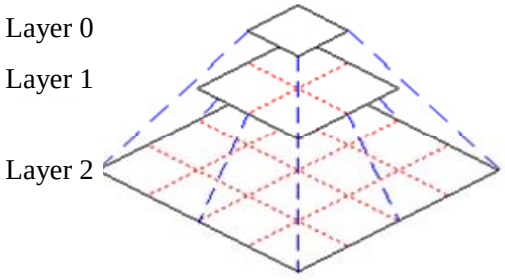


Figure 1. The hierarchical structure of pyramid model

provide such data, without the need for real-time re-sampling. That is advantages of pyramid model. If there is no pyramid model, it is necessary to real-time simplified based on the original terrain data and image data for achieving the level of detail results. Despite an increase the data storage space, the pyramid model can reduce the time in mapping and speed up the rendering of the scene.

B. Pyramid-based Texture Mapping

In 3D terrain visualization, the true reflection of the terrain is an important part. Texture mapping is a major technology to increase the authenticity. The Ground Remote sensing image is a real record on ground floor landscape; it contains a rich mount of information. Therefore, take remote sensing image as texture image, through texture mapping; 3D terrain model can have a richer information and realism.

Texture mapping used in this paper, is a method which is used in multi-resolution terrain LOD simplification. Firstly, for the texture, establish a quadtree-based pyramid and texture tree. Because the terrain geometry and textures use in the same pyramid model, in real-time terrain rendering, the appropriate resolution texture is selected from the texture tree according the details of terrain levels.

The terrain geometry and texture use in the same pyramid model. When rendering the terrain geometry, in accordance with corresponding relation, finding the corresponding block, and then by way of texture of texture mapping, the remote sensing image pictures was draw to the terrain, thus increasing the real effect of the terrain.

Ⅲ. VISUALIZATION OF VECTOR DATA IN 3D TRRRAIN

The basic idea that vector data visualized in 3D terrain is, reading the vector data, raster 2D vector data into texture image, and then using the standard texture mapping method to draw, so the vector data can be draw out in 3D terrain.

The 3D terrain visualization needs two kinds of data, geometric terrain data (DEM) and surface remote sensing image (DOM). Which are organized in the external device in the form of layer and sub-block. But in this paper, pyramid model was used which based on Quad-tree. The vector data take as whole, recorded regional geographic information, not organized in the form of block. Vector data can be visualized in 3D terrain by raster into the form of texture. Texture block vector data is composed of two parts: remote sensing image sub-block data, the vector data in the same region. As showing in the Fig. 2, the terrain geometry and textures use in the same pyramid model. When rendering to a block of terrain geometry, according the corresponding relation to find the corresponding block, and calculate to get the location information that the remote sensing image block represent, which according the pyramid tree, and then read the vector data obtained in the region. Using the render-to-texture, rendering the vector data to remote sensing image block, and attributes of vector data

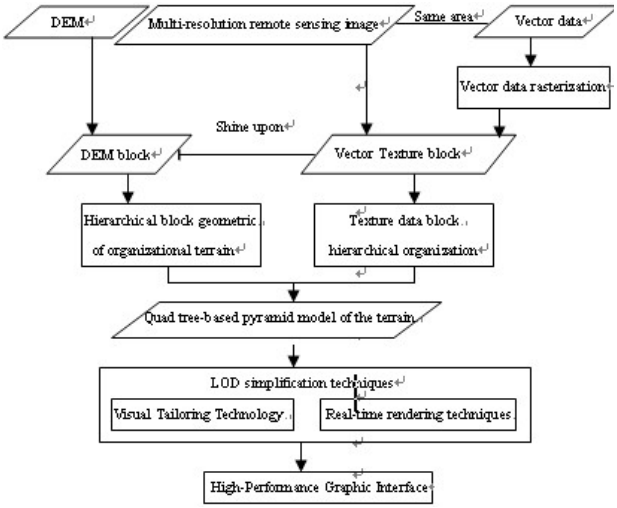


Figure 2. The schematic of vector data visualized in 3D terrain

such as color、 fill-color, need to be determined in advance of drawing vector texture mapping. So the new remote-sensing texture block that obtain vector data was got, which called “vector texture block”. Then the ordinary texture mapping, texture mapping to the topography of the block, making the three-dimensional terrain rendered on vector data. As the vector data have been drawn to the texture, therefore, the vector data can be up and down with the terrain raising and falling.

Ⅳ. THE KEY TECHNOLOGY OF REAL-TIME RENDERING

A. The determination of details on different levels.

In this paper, the vector texture data is organized in quadtree structure. According the distance of view point, dynamically render the different LOD vector texture. In the same distance, to determine the texture block that is necessary in the current view of the area.

In order to real-time render the vector texture data. The texture level detail was calculated by the distance that is from the view point vertically to the ground. So the texture level detail which the whole vector data rendered is consistent. According to the features of quadtree, for the four times differed between the neighbor level data, in order to keep the texture resolution consistently which displayed in screen. Switch the detail level when the distance of the view point is twice.

The level of vector texture transits gradually, through which the texture displaying scope controlled by the

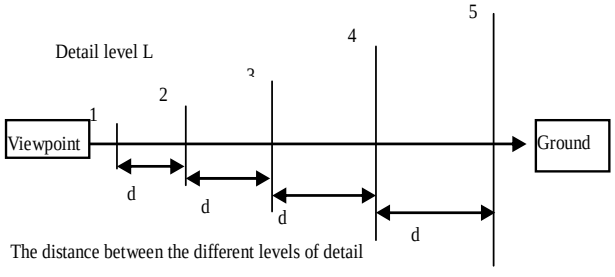


Figure 3. detail by switching levels

distance of view point. Given that the distance of view point was increasing, the vector textures need to become thicker from the level of L to L+1. The L level is “OUT” stage, and the L+1 level is “IN” stage. Conversely, when the point of view gradually decreases, the vector texture become thinner from L level to L-1 level, the L level is called “OUT” stage, while the L-1 level called “IN”. As shown in Fig. 3

B. The determination of target sub-block visibility

For the view point, most domains of the terrain are invisible, so cutting off that before rendering to accelerate the speed of terrain rendering. The four tree node consist of one three-dimensional bounding box which contains all of its own sub-tree ,its leaf nodes is an actual drawing of the block bounding box and adjacent to the Terrain area block of the bounding box. If the leaf nodes of the bounding box are located partially or entirely within the visual, then this terrain block can be marked as visible. Traverse the entire four-node-tree, cutting and processing from the root, Each block will be marked as visible or not visible, thus can obtained a collection of plots which should be rendered.

Target sub-block search in fact accompanied by resolution of search. If adopt one resolution, then only related to the regional level search and domain search. For the use of View-dependent multi-resolution terrain visualization techniques to simplify the system, Require different regions corresponding to different resolution topographic data. Therefore, in sub-block search process needs resolution test. Fig. 4 is Resolution requirements for topographic block diagram, the white rectangle that intersects with the view frustum and the need to participate in the sub-block data Terrain Rendering, The smaller rectangular block that higher the resolution

First, determine whether the topography of the region can be seen that within the data. Testing and visible regions of the top-level sub-block overlap is to meet the resolution requirements, if met, will continue to test sub-branch block, until the search out all the requirements to meet the resolution of sub-blocks. Algorithm steps described as follows:

[1] Judge whether the Regional and topographic data can be seen overlapping or not, If the overlap, then enter the Step 2, otherwise the end.

[2] One by one to judge whether the block to meet the resolution of top-level requirements, if met, will insert it into the draw list, or a branch of its four sub-blocks A, inserted into the test list.

[3] Judge whether the list is empty, if empty, the algorithm ended, or else turn to Step 4.

[4] One by one to judge whether the list one by one in the sub-block is overlapped with the visible region, if the overlap will insert it into the test list B, otherwise discard them.

[5] Clear the test list.

[6] Test one by one to determine whether the sub-block in list B is to meet the resolution requirement, if met, will insert it into the draw list, or a branch of its four sub-blocks A, inserted into the test list.

[7] Clear test list B, enter step 2.

When overlapping test, As long as the four corners of points in the block at any point in the visible region, It means that the sub-block overlap with the visible region. For the resolution test, first, calculate the goal resolution of the four corners, If the goal resolution values of the four corners are greater than or equal to which in the pyramid layer.

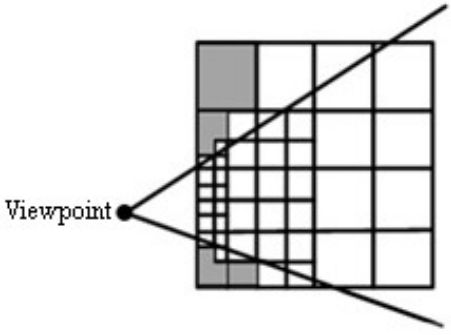


Figure 4. Resolution demand of target-blocking

Ⅴ. EXPERIMENT

Based on the study above, this paper presents the vector data in three-dimensional terrain visualization, Experimental results are shown in follow Figure. Fig. 5 does not contain the vector data for the three-dimensional scene, Fig. 6 for the vector data contained in the same area of the three-dimensional scene. From the two Figures, we can see that the vector data lines can be drawn high-quality and clearly. From Fig. 6, we can see that the path line (yellow line) of vector graphics and which in remote sensing images accurate overlap, and Always tightly attached to the surface of the earth. Roaming, amplifying or minifying the scene, vector lines will not change with the terrain changes. This is consistent with the characteristics of vector graphics.

In our experiment, we choose terrain elevation and remote sensing image data in Hangzhou, Zhejiang Province as the main data. The centre of the coordinate is located at 30.301ºE, 120.231ºN. The bounds are 29.7513ºE ~31.2358ºE and 119.8352ºN～121.3525º N.

Image texture data: a region-wide multi-resolution remote sensing image data with the amount of 113.25MB;

DEM data: SRTM region-wide multi-resolution elevation data with the amount of 23.905MB.

We use the vector data in the urban area, Hangzhou, Zhejiang Province. The centre of the coordinate is located

Figure 5. Terrain effects chart comparison with non-vector data (left)and in-vector data(right)

Figure 6. Layer tree structure diagram

at 30.301ºE, 120.231ºN. The bounds are 29.4381ºE ~32.0187ºE and 118.7664ºN～121.9241º N.

The experimental results show that the visualization of three-dimension terrain can be achieved by drawing the vector data into the texture then mapping to the 3D topography.

In 3D geographical information system, we can manage all data layers by layering. For example, the layers of a vector data can be show separately in 3D geographical information system so that we can customize the layer displayed in current windows. Besides, the control of 3D models in 3D topography can be achieved by layering. We can create a layer and input 3D models into it so that these models belong to the layer. What's more, we can charge these layers with the same type of models to one layer, achieving displaying and hiding operation. ☑Displaying, ☐ Hiding. Layer tree structure is shown as Fig. 6.

In 3D geographic information system, we use cursor to skim over 3D scene. Also, we can set the flight path, making visual angle in accordance with pre-set flight path, loading and rendering 3D models, when the path is fixed. In this article, we also achieved 3D flight path setting. The code is as follows:

```
IInformationTree id = new TerraExplorerClass();
int group_id = id.FindItem("xxx");
string filepath = @"D:\otter.xpc";
ITerrainDynamicObject3  tdo;
IObjectManager4 airplane= new TerraExplorerClass();
tdo =
airplane.CreateDynamicObject(DynamicMotionStyle.MOTION_AIRPL
ANE, DynamicObjectType.DYNAMIC_3D_MODEL, filepath,2.0,
HeightStyleCode.HSC_TERRAIN_ABSOLUTE, group_id, "airplane");
//create flying object
tdo.CircularRoute = 0;
tdo.ScaleFactor = 0.5;
tdo.ToolTipText = "ok";
int  number = 0;
double jd = 62.01, wd = 11.01, hc = 3830.1;   //set flying start point
for (int i = 0; i <= 20; i++)
{

      tdo.AddWaypoint(jd, hc, wd, 400, number);  //add flying path
      jd += 0.1;
      wd += 0.1;
      hc += 10000;
      number = number + 1;

}
```

In the above code, the first set flying objects, flying object, based on three-dimensional model of the form. By IObjectManager4 object to obtain a reference to an object flying objects, set the object's flight path of the starting position and, through the cycle of adding points the way to build a set of points formed by the flight path. The flight path setting, flying objects flying experience along the flight path.

Ⅵ. Summary

In the 3D scene, the introduction of the vector data allows the 3D terrain with spatial analysis capabilities, and make up for deficiencies in the performance of 3D terrain data. For example, the vector data line can be taking as the 3D terrain regional border, which can make a better performance. Experiments show that, using the texture-based to render the vector data into 3D terrain, can basically meet the requirements of real-time rendering, and the result is good which the vector data can be well rendering in 3D terrain.

References

[1] Guanjun Yang. Research on the combination of vector data and 3D terrain visualization [D]. Master thesis, China Agricultural University. 2007.6.

[2] Xiaoping Nei, Yahui Lu, Chongjun Yang, Yanmin Zhang. Vector-grid integration between the 3D terrain visualization research [J]. Computer Engineering and Application. 2002, 11(20):16-18.

[3] Wan Zou, Jingyun Fang, Jingang Liu. Research on Hybrid multi-resolution terrain and Space Vector Data Visualization [J]. Journal of System Simulation. 2006, 18(1):324-325.

[4] Zhaohua Liu, Jingyu Yang, Chenguang Dai. Vector data in 3D scene rendering [J]. Metal mine mountain, 2008,384(6):94-96.

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 081-084

The Evaluation System for Blending Learning Based on Blackboard Academic Suite

Lin Xin[1], Xixiang Zhang[2], Kede Qin[3], and GuangxueYue[2, 4]
[1] Business School of Jiaxing University , Jiaxing, China
[2] Mathematics and Information Engineering School of Jiaxing University , Jiaxing, China
[3] Economics School of Jiaxing University , Jiaxing, China
[4] Guangdong University of Business Studies ,GuangZhou, China
Email: Zhangmiddle@126.com

***Abstract*—Blending learning becomes popular in Chinese colleges and universities with the rapid using of Internet. Many universities use the BlackBoard Academic Suite(it is abbreviated as BB) to improve the interaction between teachers and students. It is important to evaluate the effect of the blending environment. Thus, the paper made study on assessing blending study based on BB. The paper build a blended learning system based on BB and it build the evaluation system for blending learning. Three indicators and 12 sub-indicators were used to evaluate blending teaching through document analysis and interviewing. Then, 420 questionnaires were distributed to collect opinions about the importance of each indicator, which used linguistic 2-tuple representation model to deal with fuzzy linguistic terms in it. Thus, an evaluation system with weight for blended learning was put forward, which can be used by the manager of university to assess the blending learning environment provided by a teacher.**

***Index Terms*—blending learning; evaluation; linguistic 2-tuple representation model**

I. INTRODUCTION

University of Michigan Business School proposed a technical software platform to enhance the effective learning based on the concept of blended learning and proved that the blended learning was better in improving student's learning abilities[1]. Professor C·K·Prahalad found that the student was transferred from a listener in traditional classroom into a knowledge creator and the network learning activities were as important as that in the classroom. In 2003, Professor He with Beijing normal university introduced the concept of blended learning in the seventh session of global Chinese computer education meeting held in Nanjing[2], which triggered the research on blended learning in China. Professor Li with Huadong normal university gave the definition of blended learning and carried out blended learning in his teaching career[3]. In 2003 the Chinese Education Bureau began a project to set up 1500 excellent courses for higher education to provide best teachers and teaching materials using information technology, which promoted the revolutions in teaching and learning idea, teaching method and teaching schema. To improve the leaning effect of students, many university in China applied the BlackBoard Academic Suite (it was abbreviated as BB).

BB is a software developed by American Black-board company, which is professional platform for network teaching and learning. Since it entered the Chinese market, it becomes an effective media for transporting and storing education resources and bridged the gap between students and teacher. Blending learning theory is developed from behaviourism and constructivism, which takes the advantage of traditional face-to –face learning and electronic learning. It does not only focus on the instructive role of teachers in guiding, enlightening and supervising students, it also enhances the leading role of students in encouraging them study actively, creatively. The BB platform provides such a tool for effective blending learning. With its popularity in Chinese universities, the result of blending learning based on BB platform should be evaluated. Thus, the paper made the research on evaluation system for blended teaching and learning based on BB. The rest of the paper was organized as follows: part Ⅱ introduced the related research on blended learning; part Ⅲ demonstrated the blended learning environment based on BB, part Ⅳ illustrated the indicator of the evaluation system for blending learning and the weight of each indicator, part Ⅴ gave the conclusion.

II. REVIEW OF RELATED WORK

Margaret Driscoll(2002)[4] gave a detailed concept of blended learning, she thought that learning process was a combination of technologies based on WEB, such as virtual classroom, cooperative learning and stream medium, learning process was also a mixture of teaching techniques and different teaching ideas. Harvi Singh and Chris Reed（2004）considered that blended learning centered around the learning effectiveness for a student using suitable approach in appropriate time, so that the student can master the knowledge and technique[5]. Donald Clark（2005）divided blended learning into two parts: offline learning and online learning in "White paper: Blended Learning" [6]. Chen and Looi（2007）investigated that the concept of blended learning should be introduced into curriculums to promote the learning effectiveness[7], that is to say, online discussion should be integrated with traditional discussion to improve learning result.

In China He and Li [2,3] gave different definition for blended learning, but they had the common point that blended learning could not just mixed traditional

classroom learning with online learning, the teacher should design the course to benefit students and guide them to learn knowledge actively and creatively. Yuan made research on teaching schema for blended learning and put forward the teaching schema which included traditional classroom teaching and fictitious simulation environment[8]. Fan studied the teaching processes that included problem planning, teaching experiment, recording the process and periodical summary, then he put forward a new teaching schema[9]. Li provided a new blended self-learning schema including ten steps based on units of thinking process: thinking start point, thinking materials, thinking direction, thinking style and thinking result [10].

From the proceeding walk-through in researches on blended learning, we can conclude that the related researches on definition of blended learning and teaching schema based on blended learning were studied widely and deeply. And it is accepted by many scholars that the blended learning is different from traditional classroom learning. Therefore, the evaluation for blended learning should be considered. Few research on the field, however, is not found in published papers.

III. BUILDING THE BLENDED LEARNING SYSTEM

According to the properties of professional knowledge about business management, a blended learning system was designed to integrate network leaning and face-to-face learning to realize the revolution of teaching learning resources, method and schema, thus improving student's learning abilities.

The designing of blended learning system involved four parts and every part went through three research actions. The detailed process of the designing is as follows:

Planning. Firstly the interests of students are found through questionnaire, then the learning initiative are inspired through network leaning and classroom lecturing. Finally, preview assignments should be announced a week before and students should proposed their own questions using network resources learning, the questions are collected and representative questions will be selected as the learning task considering difficult points and key points of the course, then the teacher will lecture these questions in the classroom and integrate network resources.

Practicing. The content of the course is divided into three categories, points to be conscious of, key points and difficult points. Points to be conscious of is studied by students themselves or group learning. Key points and difficult points will presented to students by web pages or multimedia courseware in BB and are also lectured by teachers in the classroom. Students finish their homework using BB network resource and online discussion board and submit them. The excellent homework will be selected by students and teachers and be displayed in a special portforlio.

Recording. The teacher observes students' utilizing abilities of blended leaning resource and the problem solving capabilities. And improvements students made through questions they proposed and accuracy of their answers.

Summarizing. Teachers review on students' submitted homework, students' knowledge expanding, problem-solving abilities and learning initiative, and introspect the results.

In designing the blended learning system, the following factors were considered: firstly, the blended learning resources planning included face-to-face lecture content, network course resource and online learning portforlio. Secondly, blended learning discussion environment setting, choosing of topic and student grouping and discussion process planning are the conclusion of blended discussion board designing. At last evaluation on blended learning should be designed, which consider the evaluation on classroom teaching, virtual teaching process and blended discussion board designing. And the architecture of blended learning system is demonstrated as Fig1.

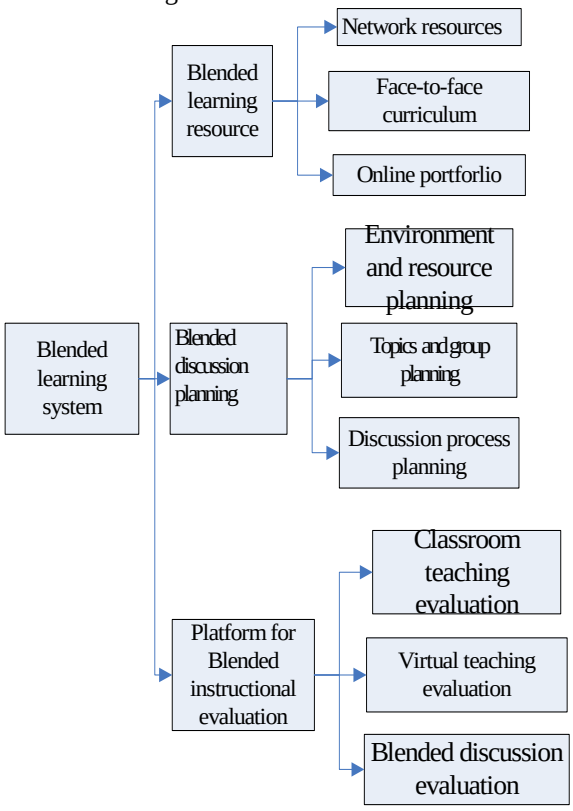


Figure1 architecture of blended learning system

The experiment was planned to test the effectiveness of the blended learning system. Two classes selecting the same course were decided, which were donated as Class A and Class B. Students in Class A used the blended learning system and students in class B used the traditional face-to-face teaching model. Students were scored using same test paper and other assessing standard after the course was finished. Then, the learning abilities in Class A and Class B were compared and analyzed to testify the hypothesis that students used blended learning system can obtain better understanding of the course than students used traditional classroom lecturing scheme.

Ⅳ. EVALUATION OF BLENDED LEARNING

Evaluation of blended learning involves in evaluation on classroom lecturing, virtual teaching and blended discussion. Every university has its own evaluation system for classroom lecturing. The paper centered around evaluation on virtual teaching and blended discussion.

A. EVALUATION INDICATORS

In order to understand the requirement of students for virtual classroom and blended discussion board, a questionnaire was designed. The key questions about blended learning was selected through interviewing students, related teachers, and manager in charge of teaching and learning assessment. After arranging the investigated materials, an evaluation system for blended learning was introduced. There are three main indicators concerning quality of BB resource, interaction in BB discussion board and quality of blended discussion respectively. The quality of BB resource is divided into three sub-indicators, which are abundance of resources in BB, relativity of resource provided in BB, distribution reasonability of resources in BB. And interaction in BB discussion board consists the following sub-indicators that are punctuality of teacher's answer, accuracy of teacher's answer, energetic discussion inspired by teachers. The last indicator- distribution reasonability of resources in BB is represented by the abundance of the materials and its relationship to the topic, the logic of the team representation report, frequency of the team discussion, the accuracy and depth of the problem solving, question posted and frequency of answers, attitude of the group team.

In assessing the blended learning, the weight of each indicator should be given. To decide the importance of each indicator, a questionnaire was designed. The questionnaire included 12 questions, which can be classified into 3 categories and the questions can be worked out by selecting the related choices. The first part was designed for BB resources, the second part consisted of questions involving interaction of blended discussion board and the third part considered distribution reasonability of resources in BB. After the questionnaire was revised, 420 questionnaires were distributed among teachers, staff managing teaching affaire, student majored in financial management, information management and information system, international economic and trade, computer science. 405 answered questionnaires were collected, and 399 of them were effective.

There exists fuzzy terms such as "excellent, good" in the questionnaire, represented model should be selected to handle it. Inter-valued fuzzy number fuzzy triangular number, linguistic indices and linguistic 2-tuple representation model can be used to represent fuzzy terms[11,12,13]. Linguistic 2-tuple representation model was selected to deal with fuzzy terms in the questionnaire because it was more accurate in representing fuzzy terms and had less loss in carry out the calculation[14].

B. LINGUISITC 2-TUPLE AND ITS OPERATOR

Suppose S={s_0, s_1, ⋯, s_g} be a set of labels assessed in a linguistic term set with odd elements, which has the following properties: ①ordered: when the index i⩾j, there must exist $s_i \geqslant s_j$; ②a negation operator: $Neg(s_i)= s_{g-i}$; ③ there exists a min and max operator: $_{si} \geqslant s_j$ means $\max(s_i, s_j)=si$ and $\min(s_i, s_j)=s_j$ [13].

Let β be the result of an aggregation of the indexes of a set S={s0, s1, ⋯, sg}, for example, the result of a symbolic aggregation operation. $\beta \in [0, g]$ and g+1 is the cardinality of S. Let $i = round(\beta)$ and $\alpha = \beta - i$ be two values, such that, $i \in [0, g]$ and $\alpha \in [-0.5,0.5]$ then α is called a Symbolic Translation[14].

Let S={s0, s1, ⋯, sg} be a linguistic term set and $\beta \in [0, g]$ be a value representing the result of a symbolic aggregation operation, then the 2-tuple that expresses the equivalent information to β is obtained with the following function[14]:

$$\nabla : [0, g] \to S \times [-0.5,0.5] \quad (1)$$

$$\nabla(\beta) = (s_i, \alpha), with \begin{cases} s_i, i = round(\beta) \\ \alpha = \beta - i, \alpha \in [-0.5,0.5] \end{cases}$$

Where round(.) is the usual round operation, si had the closest index label to β.

Let S={s0, s1, ⋯, sg}be a linguistic term set and (s_i, α) be a 2-tuple. There is always a ∇^{-1} function, such that, from a 2-tuple it returns its equivalent numerical value $\beta \in [0, g]$, which is:

$$\nabla^{-1} : S \times [-0.5,0.5] \to [0, g]$$
$$\nabla^{-1}(s_i, \alpha) = i + \alpha = \beta \quad (2)$$

From definition 1 , definition2 and proposition 1, we can conclude that the conversion of a linguistic term into a linguistic 2-tuple consist of adding a value 0 as the symbolic translation, which is :

$$\theta(s_i) = (s_i, 0) \quad (3)$$

Operation model of linguistic 2-tuple can be obtained according to the linguistic 2-tuple representation model.

⑴ A linguistic 2-tuple negation operator.

$$Neg((s_i, \alpha)) = \nabla(g - (\nabla^{-1}(s_i, \alpha))) \quad (4)$$

⑵ Linguistic 2-tuple aggregation operators

Let $(s_1, \alpha_1), (s_1, \alpha_2), \cdots, (s_n, \alpha_n)$ be a set with n linguistic 2-tuples, the average operator of linguistic 2-tuples ξ is[15]:

$$\xi((s_1, \alpha_1), (s_2, \alpha_2), \cdots, (s_n, \alpha_n)) = (\bar{s}, \bar{\alpha})$$
$$= \nabla(\frac{1}{n}\sum_{i=1}^{n} \nabla^{-1}(s_i, \alpha_i)) \quad (5)$$

Let $(s_1, \alpha_1), (s_1, \alpha_2), \cdots, (s_n, \alpha_n)$ be a set with n linguistic 2-tuples and $\omega = (\omega_1, \omega_2, \cdots, \omega_n)$ be the

related weighted vector with $\sum_{i=1}^{n}\omega_i = 1$, then the weighted average operator of linguistic 2-tuples ξ^{ω} is[15]

$$\xi^{\omega}((s_1,\alpha_1),(s_2,\alpha_2),\cdots,(s_n,\alpha_n)) = (\hat{s},\hat{\alpha}) = \nabla(\sum_{i=1}^{n}\nabla^{-1}(s_i,\alpha_i)\omega_i) \quad (6)$$

C. *WEIGHT OF EVALUATION INDICATORS*

The questionnaires were analyzed using linguistic 2-tuple representation model. And the weight of each indicator was obtained. The weight of each sub-indicator was also analyzed. Thus, evaluation system for blended learning and its weights of each indicator and its sub-indicator could be demonstrated as table 1.

Table 1 evaluation system and its weight

indicators (weight)	Sub-indicators (weight)
Quality of BB resource (0.35)	Abundance of resources in BB (0.37)
	Relativity of resource provided in BB (0.34)
	Distribution reasonability of resources in BB (0.29)
Mutual exchange quality of BB (0.37)	Punctuality of teacher’s answer(0.33)
	Accuracy of teacher’s answer(0.38)
	Energetic discussion inspired by teachers(0.29)
Quality of blended discussion (0.28)	The abundance of the materials and its relationship to the topic(0.22)
	The logic of the team representation report (0.20)
	Frequency of the team discussion(0.14)
	The accuracy and depth of the problem solving.(0.19)
	Question posted and frequency of answers (0.09)
	Attitude of the group team(0.16)

Ⅴ. CONCLUSION

The paper introduced the architecture of blended learning system. It got the indicators to evaluation the blended learning. Then it used linguistic 2-tuple representation model to handle fuzzy term in questionnaire and obtained the weight of each indicators and its sub-indictors. The education management department can use the evaluation system proposed in the paper to assess the network resource in BB.

ACKNOWLEDGMENT

This work is supported by Education Planning Research foundation of Zhejiang Province Grant by scg85. It is also supported by Jiaxing University Education Research Foundation Grant by 85150932 and Economic Commence Market Application Technology Foundation Grant by 2007gdecof004.

REFERENCES

[1] Prahalad. Case study: University of Michigan Business School Using PRAJA. Technology to Enhance Learning,2002.

[2] K.K.He, “E-Learning and the revolution of university/college“ ,http://www.edu.Cn / 20011225/3015160. shtml”, 05-10-2009

[3] K.D. Li, J.H.Zhao, “Theory and its application of blended learning”, Electrical education research, Vol.4, pp.78-81, July 2004

[4] M.Driscoll·, “Blended learning：Let’s get beyond the hype”, Learning and Training Innovations[R],2002.

[5] H.Singh, C.Reed, “Achieving Success with Blended Learning”, http://www.centra.com/downloa whitepapers/ blended learning. Pdf, 30-01-2002.

[6] D.Clark(2003), “Whitepaper: Blended Learning”, http://www.epic.co.uk/content/resources/white_papers/blended. Htm, 21-05-2005

[7] W.L.Chen andC.K.Looi, “Incorporating online discussion in face to face classroom learning: A new blended learning approach”, Australasian Journal of Educational Technology, Vol.23, pp. 307-326, March 2007:.

[8] N.H.Yuan, “Blended instructional schema based on virtual technology”, Journal of Guangdong technology normal college,Vol.5, pp. 135-137, May 2006

[9] W.S. Pan, “Research on blended instructionalschema based on network resource”, Electrical education research, Vol. 8, pp. 49-51, August 2006.8

[10] X.Li, “Exploring effective learning schema based on blended learning and self-studying”, Modern education techniques, Vol.4, pp.32-35, April 2007

[11] Y.P.Jiang, Y.N.Xing,“Consistency Analysis of Two-Tuple Linguistic Judgement Matrix,” Journal of Northeastern University(Natural Science), Vol.28, pp.129-132, January 2007

[12] Z.S.Xu, ” A Multi-attribute group decision making method based on fuzzy linguistic evaluationa and GIOWA operator”, J.Sys.Sci.& Math. Sci., Vol.24, pp.:218-224, April 2004 (in Chinese).

[13] F. Herrera, L.Martinez. A model based on linguistic 2-tuples for dealing with multigranular hierarchical linguistic contexts in multi-expert decision making. IEEE Transactions on Systems, Man, and Cybernetics-part b: Cybernetics, Vol.31, pp. 227-234, April 2001

[14] F.Herrera , L.Martinez, “The 2-tuple linguistic computational model—advantages of its linguistic description, accuracy and consistency”, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, Vol.9, pp. 33-48, September 2003

[15] X.W.Liao, H.Li,G.M.Dong., “A Multi-attribute Group Decision-making Approach Dealing with Linguistic Assessment Information,” Theory and Practice of System Engineering, Vol.9, pp.90-98, September 2006 (in Chinese)

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 085-088

Swarm Dynamics Behavior Analysis and Coordinated Control of Limited-Range Perceived Agents

Zhibin Xue[1*,3], and Jianchao Zeng[2]
[1]Department of chemical machinery, Chemical engineering college, Qinghai University, Xining, China
Email: zbxue_jack@163.com
[2]Complex System & Computer Intelligence Laboratory; Taiyuan University of Science & Technology, Taiyuan, China
Email: zengjianchao@263.net
[3]College of Electric & Information Engineering, Lanzhou University of Technology, Lanzhou, China

***Abstract*—Swarming behavior is ubiquitous in nature and society. It is of both theoretical and practical importance to investigate the underlying principles and mechanisms of coordination and cooperation emerging in swarms. In this paper we propose a simple isotropic range limited-perceive agents dynamic model to study collective and formation behavior of a group of mobile autonomous agents interacting through a long range attraction and short range repulsion function. It is shown that the individuals (agents) in the swarm during the course of coordinative motion can realize the local collision-free stabilization of constituting a particular predefined geometrical configuration. Numerical simulations are also worked out to illustrate the analytical results.**

***Index Terms*—swarm dynamics, coordinated control, isotropic, limited-range perceive, multi-agent, collision-free, formation control, numerical simulations**

I. Introduction

Swarming behavior has been observed in nature. For example, bees, ants and birds often work together in groups for viability[1]. It is known that such cooperative behavior has certain advantages, for example, predator avoidance, foraging success and so on.

The general understanding in biology is that the swarming behavior is a result of an interplay between a long range attraction and a short range repulsion between the individuals [1]. Understanding the cooperative and operational principles of such systems may provide useful ideas for modeling and exploring the collective dynamics of swarming behavior in bionics for using in engineering applications, such as the coordinated control of multi-agent systems.

However, most available results in the literature are on the isotropic global perceive swarming model, convincing results on the isotropic local perceive swarming model are relatively few. Based on the analysis of various biological swarms of dynamic aggregation mechanism, an isotropic limited-range perceived swarming dynamic model is proposed in this paper. Meanwhile, the chief research objective of the paper is to use the limited range perceive model to solve the formation control question of multi-agent systems to constitute various predefined geometrical configuration formations pattern on multi-agent systems in an n-dimensional Euclidean space by using artificial potential functions and Newton's iteration method. Dynamic change of the environment, local observation and nonlinear characteristics are ubiquitous phenomena in nature, but the study is very difficult and it has profound engineering significance. So this article is able to provide some results on this topic.

II. Swarm Dynamical Models

Gazi and Passino[2] proposed an isotropic global perceive swarm model with a simple attraction-repulsion function specifying the inter-individual interactions and showed that the model can exhibit the basic features of aggregation, cohesion and complete stability. However, the attraction-repulsion functions considered in this study could not avoid collisions since they are not unbounded for infinitesimally small arguments.

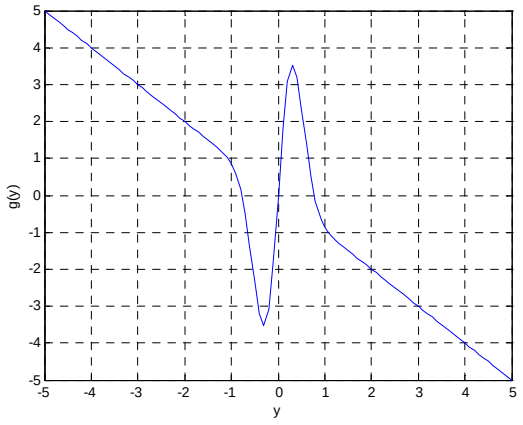


Figure 1. The attraction/repulsion function g(·) in Ref. [2]

Manuscript received January 28, 2010; revised February 28, 2010; accepted March 16, 2010.
*Corresponding author (email: zbxue_jack@163.com).
This work was supported by the National Natural Science Foundation of China (Grant No. 60975074).

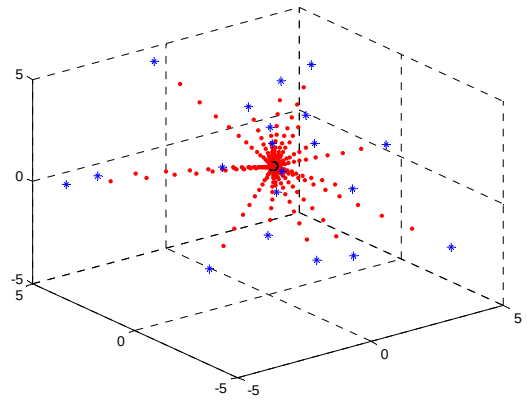


Figure 2. The paths of the swarm members in Ref. [2]

Chen and Fang[3] proposed an isotropic local perceive swarming model, the attraction-repulsion functions considered in this study could avoid collisions, but the model could not agree well with the nonlinear characteristics.

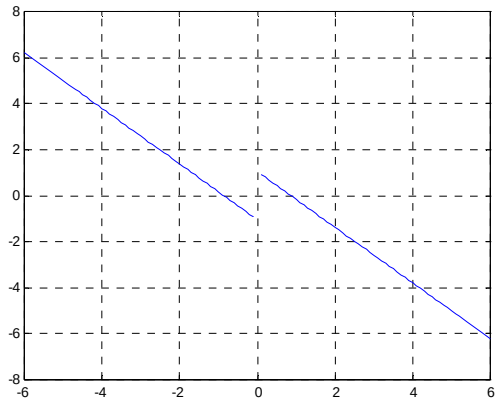


Figure 3. The attraction/repulsion function g(·) in Ref. [3]

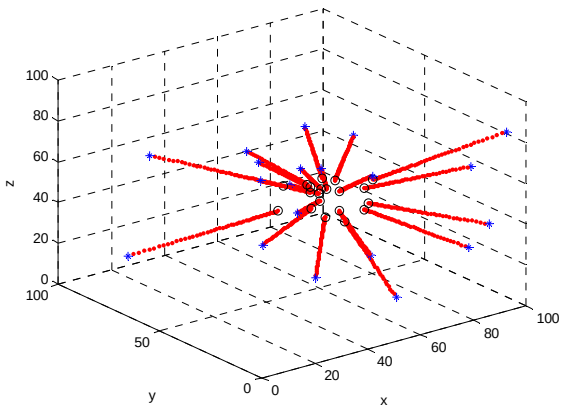


Figure 4. The paths of the swarm members in Ref. [3]

In order to overcome the shortcomings of in literatures [2] and [3], Xue and Zeng[4] proposed an isotropic limited-range perceive swarming model that can in harmony with real biological swarms well, and improve the coordination motion behavior of multi-agent systems.

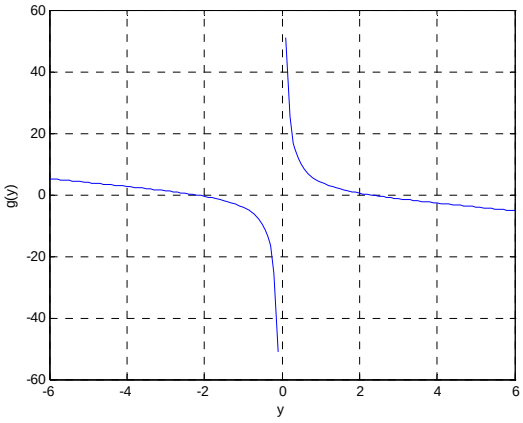


Figure 5. The attraction/repulsion function g(·) in Ref. [4]

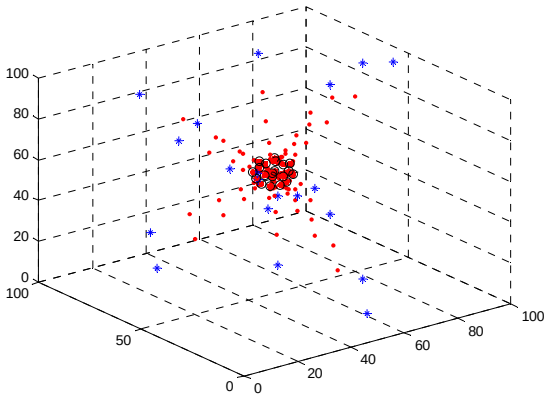


Figure 6. The paths of the swarm members in Ref. [4]

Based on the inspiration from biology, referring to the known results in literatures [4], we consider a swarm of M individuals (members) in a n–dimensional Euclidean space, assume synchronous motion and no time delays, and model the individuals as points and ignore their dimensions. The equation of collective motion of individual i is given by as follows

$$\dot{x}^i = -\nabla_{x^i}\sigma\left(x^i\right) + \sum_{j=1, j\neq i}^{M} g\left(x^i - x^j\right), i = 1, \cdots, M. \quad (1)$$

Where $x^i \in R^n$ represents the position of individual i; $-\nabla_{x^i}\sigma\left(x^i\right)$ stands for the collective motion's direction resting with the different social attractant/repellent potential fields environment profile around individual i; $g(\cdot)$ represents the function of attraction and repulsion between the individuals members. The above $g(\cdot)$ functions are odd (and therefore

symmetric with respect to the origin). This is an important feature of the $g(\cdot)$ functions that leads to aggregation behavior [2].

The attraction/repulsion function that we consider is

$$g(y)=-y\left[g_a(\|y\|)-g_r(\|y\|)\right]=-y\left[a-\frac{b(v-r)}{(r-\rho)\|y\|^2}\right]. \quad (2)$$

Where, a,b,v,r,ρ are arbitrary constants, is normal number, $v>r>\rho>0$, the 2-norm $\|y\|=\sqrt{y^T y}$. The numerical imitation of $g(\cdot)$ as Fig. 5 and Fig. 6 shows.

In Fig. 2, Fig. 4 and Fig. 6, blue "*" represent original position, black "。" represent final position, read "." represent convergent trajectories of individuals.

III. Formation and Coverage Search Behavior

The formation concept, first explored in the 1980's to allow multiple geostationary satellites to share a common orbital slot [5], has recently entered the era of application with many successful real missions [6].

Formation control is an important issue in coordinated control for multi-agent systems (such as, a group of unmanned autonomous vehicles (UAV)/robots). In many applications, a group of autonomous vehicles are required to follow a predefined trajectory while maintaining a desired spatial pattern. Moving in formation has many advantages over conventional systems, for example, it can reduce the system cost, increase the robustness and efficiency of the system while providing redundancy, reconfiguration ability and structure flexibility for the system. Formation control has broad applications, for example, security patrols, search and rescue in hazardous environments. Research on formation control also helps people to better understand some biological social behaviors, such as swarm of insects and flocking of birds [7].

Control of systems consisting of multiple vehicles (or agents) with swarm dynamical models are intend to perform a coordinated task is currently an important and challenging field of research. Formation of geometric shapes with autonomous robots is a particular type of the coordination problem of multi-agent systems [8].

In fact, we consider formation control as a special form of swarm aggregation, where the final aggregated form of the swarm is desired to constitute a particular predefined geometrical configuration that is defined by a set of desired inter-agent distance values. This is achieved by defining the potential function to achieve its global minimum at the desired formation. For this case, however, due to the fact that potential functions may have many local minima, the results obtained are usually local. In other words, unless the potential function is defined to have a single (unique) minimum at the desired formation, convergence to that formation is guaranteed only if the agents start from a "sufficiently close" configuration or positions to the desired formation. Some of these works are based on point mass agent dynamics [9].

So, by use of the range limited-perceive swarm dynamical model, based on artificial potential field (APF) function and Newton-Raphson iteration update rule to numerical imitation analyze how a large number of agents namely Large-scale swarm system can form desired particular predefined an approximation of a simple convex polygon (such as, diamond) formation in the plane by collective motion, related the range limited-perceive swarm pattern formation behavior results examples as follows

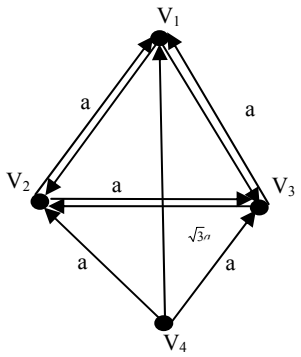


Figure 7. The ideal formation configuration of the diamond for 4 agents

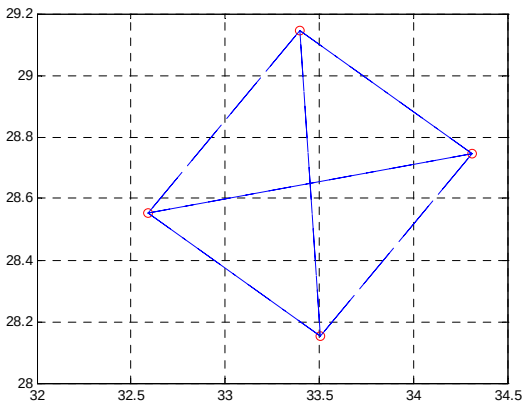


Figure 8. Congregated positions of entire of the ideal formation configuration of the diamond for 4 agents in plane

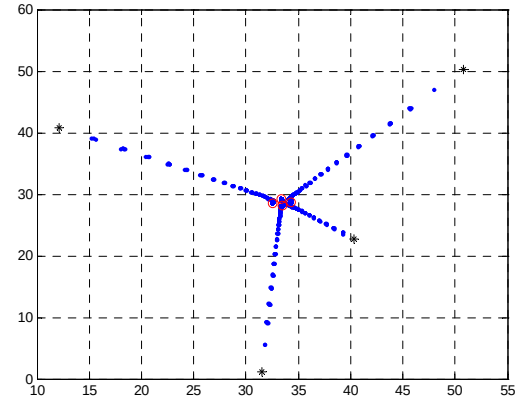


Figure 9. Convergent trajectories of the ideal formation configuration of the diamond for 4 agents in plane

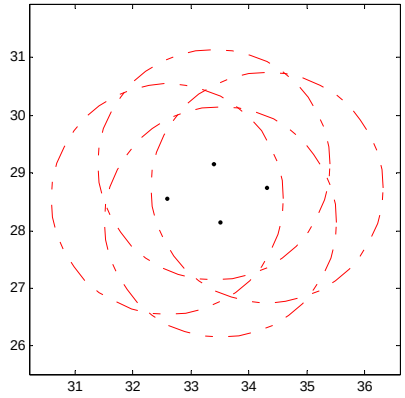


Figure 10. The communication and coverage search capabilities showing for 4 agents members in the swarm to form the ideal diamond formation in plane at the time of entire positions

In above graphs, black "*" represent original position, read "。" represent final position, blue "." represent convergent trajectories of individuals, the polygonal vertex shows the final numerical imitation configuration position of UAV/robots. Considering the convenience of simulation, let $\nabla_y \sigma(y) = 0$. In the figures of the relations between sides and angles of the desired formation configuration, black spheres represent final configuration position, $V_i, i = 1, \cdots, \mathrm{M}.$ represent different vehicles in the swarm systems.

As mentioned above, the particular predefined convex polygon geometrical configuration formations were discussed. The simulation results in this section reify our theories of formation stabilization and for multi-agent systems.

IV. Conclusions Remarks

This paper studied the coordinated motion control, obstacle avoidance and formation behaviour control problem of a group of isotropic range limited-perceive agents with interaction force between individuals moving in an *n*-dimensional Euclidean space. To solve the problem, based on the inspiration from biology, we proposed an isotropic range limited-perceive swarm model of swarm aggregating behavior for multi-agent system. Meanwhile, the model is a kinematic model. It is fit for individuals in which move basing on the Newton's law in an environment can capture the basic convergence properties of biological populations in nature. Therefore, the final behavior of the swarms described by the model may be in harmony with real biological swarms well. Numerical simulation agrees very well with the theoretical analysis of coordinated motion and obstacle-eluding aggregating stability of the swarm systems, the convex polygon formation behavior control and obstacle-eluding aggregating behavior for multi-agent system is analyzed and discussed at last. For further work, the experiments will be conducted in the presence of dynamic obstacles … etc. Therefore, it is obviously the swarm aggregating results obtained from the isotropic range limited-perceive swarm model which has a definite reference value in the multi-agent coordination and control literature.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (Grant No. 60975074).

References

[1] C. M. Breder, "Equations descriptive of fish schools and other animal aggregations," *Ecology*, *Washington*, vol. 35, no. 3, pp. 361–370, April 1954.

[2] V. Gazi, K. M. Passino, "Stability analysis of swarms," *IEEE Transactions on Automatic Control*, *New York*, vol. 48, no. 4, pp.692–697, April 2003.

[3] S. Chen, H. Fang, "Modeling and stability analysis of large-scale intelligent swarm," *Control and Decision*, *Shenyang*, vol. 20, no. 5, pp.490–494, May 2005.

[4] Z. Xue, and J. Zeng, "Simulation Modeling and Analysis of Dynamics of Range Limit Perceived Group," *Journal of System Simulation*, *Beijing*, vol.21, no.20, pp.6352-6355, October 2009.

[5] J. G. Walker, "The Geometry of Satellite Clusters,"*Journal Orbital Dynamics*, *London*, vol. 35, pp. 345-354, Aug. 1982.

[6] C. M. Saaj, V. Lappas, and V. Gazi, "Spacecraft Swarm Navigation and Control Using Artificial Potential Field and Sliding Mode Control," *IEEE International Conference on Industrial Technology*, *Mumbai*, *India*, pp. 2646-2652, December 2006.

[7] Y. Q. Chen, and Z. M. Wang "Formation control: a review and a new consideration," *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems. (IROS 2005)*, *Alberta*, *Canada*, pp. 3181- 3186, 2-6 August 2005.

[8] Y. S. Hanay, H. V. Hünerli, M. İ. Köksal, A. T. Şamiloğlu, and V. Gazi, "Formation Control with Potential Functions and Newton's Iteration," *Proceeding of the European Control Conference, Kos, Greece*, pp. 4584-4590, July 2007.

[9] V. Gazi, B. Fidan, M.İ. Hanay, and Y.S. Köksal, "Aggregation, foraging and formation control of swarms with non-holonomic agents using potential functions and sliding mode techniques," *Turkish Journal of Electrical Engineering and Computer Sciences (ELEKTRİK)*, *Ankara*, vol. 15, no. 2, pp.149–168, July 2007.

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 089-092

An Anomaly Detection Method Based on Fuzzy C-means Clustering Algorithm

Linquan Xie[1], Ying Wang[1,2], Liping Chen[2], and Guangxue Yue[1,2,3]
[1] Jiangxi University of Science and Technology, Ganzhou, China
[2] Jiaxing University, Jiaxing, China
[3] Guangdong University of Business Studies ,GuangZhou, China
Email: {lq_xie@163.com , wy363100506@sina.com}

***Abstract*—Anomaly detection based on network flow is the basis of the monitoring and response application of anomaly, it is also the important content in the fields of network and security management. In this paper, the fuzzy C-means clustering (FCM) algorithm was applied to detect abnormality which based on network flow. For the problems of the FCM, for example, it needs to preset a number of clusters and initialize sensitively, and easily fall into local optimum, the paper introduced the method combined with the average information entropy, support vector machine and fuzzy genetic algorithm etc.. These hybrid algorithms can solve the mentioned problems and classify more accurately. Finally based on the current development and the discussion of the research, it summarized the trends of the network flow anomaly detection in the paper.**

***Index Terms*—network flow, anomaly detection, intrusion detection, anomaly analysis**

I. INTRODUCTION

As the rapid expansion and the growing popularity of the Internet, more and more information has been transmitted and stored through the network. Cognizing and studying the behavioral characteristics of the Internet users has gradually attracted people's interest, and it also used to cognize, manage, optimize various kinds of the network resources, and is an important basis of the network planning and design. However, compared with the development of the network application types, the improvement of the network management technology lags behind the development of the application. How to provide a safe, reliable and efficient service environment for the vast number of the Internet users, it needs to be resolved to the network management. Network flow analysis comes into being for resolving these issues; we can indirectly get hold of the statistical behavior of the network by statistical analysis of the network flow. It can enhance the manager of the network and security to troubleshoot the network anomaly, maintaining the normal network and to ensure the network security. At present, for the network flow anomaly detection, there has conducted extensive research, but the detection accuracy has been far from desirable. Nevertheless, the anomaly detection plays an irreplaceable role in discovering unknown anomaly network intrusion detection and network failure detection, etc.

II. METHODS OF THE ANOMALY DETECTION BASED ON THE NETWORK FLOW

For the methods of the network flow anomaly detection, there has been summarized of the related research work in recent years, including the following methods: the research method based on the features/behavior, the anomaly detection based on the statistics, the method based on the machine learning and the method based on the data mining, etc.. We can find the latter three methods of anomaly detection construct models on normal behaviors, it compares with the normal model to detect anomalies, thus it can effectively find out the known and unknown attacks. With the continuous research of intrusion detection, people obtain plentiful and substantial results, at the same time, people come to realize the pervasive problems of intrusion detection, such as the rate of detection can not meet the requirements of the modern high-speed network communications, a higher rate of false alarm and missing report in intrusion detection system (IDS), the IDS lacks of active defense and the interaction is not enough among the other network security devices. To solve the above problems, the research on intrusion detection which is distributed, intelligent and comprehensive of development becomes a matter of course.

III. ANOMALY DETECTION OF NETWORK FLOW BASED ON FUZZY C-MEANS CLUSTERING (FCM) ALGORITHM

How to provide a safe, reliable and efficient service environment for the vast number of the Internet users, it needs to be resolved for the network management. With the rapid development of network technology and continuous improvement for invasion of technology, the ways of new attacks emerge in endlessly, in order to detect and defense the unknown attacks, the intelligent methods are the focus of intrusion detection and have been widely used, such as data mining, neural networks, support vector machine, intelligent agent and etc..

Cluster analysis is used to discover the hidden patterns in the instance data and used to detect the meaningful characteristics in intrusion. How to accurately determine the intruder or the intrusion is the research topics of anomaly intrusion, while there are many algorithms for application of anomaly detection, among them, fuzzy C-means clustering (FCM) algorithm becomes a hot

topic that because it is flexible division of sample data sets and can detect invasion more objectively.

A. The Theory of FCM

Fuzzy clustering is a multi-technology for classification of objective things, which construct fuzzy resemblance according to the characteristics of the objective things, the relatedness and the similarity.

Ref. [1] referred the method of fuzzy clustering analysis that it can be divided into three categories:

1) The number of categories is indeterminate; it means to cluster dynamically according to the different requirements.

2) The number of categories is given; the target is to find out the best way to classify the data. This method cluster based on objective function and called fuzzy C means (FCM) algorithm or fuzzy ISODATA clustering.

3) In the case of significant perturbation, it clusters according to the fuzzy similarity matrix. This method is called fuzzy clustering based on perturbation.

The theory of fuzzy C-means clustering (FCM) [2,3]:

Fuzzy C-means clustering is an algorithm based on the division, and it is an improved algorithm based on C-means, the C-means algorithm is rigid for data partition, but FCM is flexible and fuzzy for partition. According to the quadratic sum in minimum of the specified grouping, FCM uses the membership to determine each data instance; it divides a data instance $X=\{X_i / X_i \in R(i=1,2,\cdots,n)\}$ with n into k categories $(1<K<N)$, and calculates the cluster center of each category, in order to make the non-similarity value function minimum. The matrix of classification, $U=(u_{ij}/i=1,2,\cdots n;\ j=1,2,\cdots,k)$, where u_{ij} indicated the membership of the data instance belong to , and satisfied the following conditions:

$$\sum_{j=1}^{k} u_{ij}=1, \forall i=1,\cdots,n. \tag{1}$$

Use the FCM for fuzzy partition, so that each given data instance can determine which categories belong to, according to the membership between 0 and 1. The elements of the matrix U get values between 0 and 1. The value function defined as follows:

$$J_m(U,C)=\sum_{i=1}^{N}\sum_{j=1}^{k} u_{ij}^m d_{ij}^2(X_i,C_j) \tag{2}$$

J_m can be seen as the quadratic sum of the distance between the each data instance and the cluster center. In (2), $C=\{C_j / C_j \in I,\ j=1,2,\cdots,k\}$, and $C_j \in I$ indicate the cluster centers; $X_i \in I$ indicate the data instance sets; u_{ij} mean the membership of the data instance belong to the cluster center, their values are between 0 and 1, $U=\{u_{ij}\}$ is a matrix of $n\times k$, $C=[C_1,C_2,\cdots,C_k]$ is a matrix of $s\times k$; $C=\{c_1,c_2,\cdots,c_k\}$, c_i indicate the cluster center of the fuzzy group; $X_i \in R^p$ are the data instances; $d_{ij}(X_i,C_j)$ indicate the distance between the data instance and the cluster center; *m* means the fuzzy coefficient $(1\le m<\infty)$; *k* means the number of the pre-categories, it determined by the initial clustering. We can use the Lagrange multiplier method to obtain the necessary condition of minimum for J_m:

$$u_{ij}=1\Big/\sum_{i=1}^{k}\left(d_{ij}/d_{i1}\right)^{2/(m-1)}, \forall i \tag{3}$$

$$c_{ij}=\left(\sum_{i=1}^{m} u_{ij}^m x_j\right)\Big/\left(\sum_{i=1}^{m} u_{ij}\right), \forall j \tag{4}$$

The parameter *m* in the above formulas is a scalar to control the blur length of the classification matrix U , the bigger *m* is, the more blurred it is. If *m* = 1, the algorithm of FCM degenerates into hard C-means clustering (HCM) algorithm. FCM clustering needs many times to iterate so that the value function obtains the minimum.

B. FCM Used in Anomaly Detection

The intrusion detection algorithm based on FCM [3]:

From the above discussion we can see, the FCM algorithm requires two parameters: the number of clusters C and the parameter *m*. The number of clusters can use the clustering number of initial clustering as C, and C is less than the total number of cluster samples. The detection optimization can follow these steps:

Step1: initialize the membership matrix *U* with random number between 0 and 1, and satisfy the formula $\sum_{i=1}^{n} u_{ij}=1, \forall j=1,\cdots,n$.

Step2: use $c_i=\left(\sum_{i=1}^{n} u_{ij}^m x_j\right)\Big/\sum_{j=1}^{n} u_{ij}^m$ to calculate the cluster centers $C_i, i=1,\cdots,k$.

Step3: use $u_{ij}=1\Big/\sum_{k=1}^{c}\left(\frac{d_{ij}}{d_{kj}}\right)^{2/(m-j)}$ to calculate the new membership matrix *U*.

Step4: calculate the value function according to $J_m(U,C)=\sum_{i=1}^{N}\sum_{j=1}^{k} u_{ij}^m d_{ij}^2(X_i,C_j)$. If it is smaller than a determined threshold or is smaller than the change with the last value function, then it will stop and output the clustering results. Otherwise, return to Step2 to continue iterating.

The output of the algorithm is a fuzzy partition matrix with $N\times K$, the matrix indicate the membership of the

each sample belongs to each category. It can identify each sample belongs to a category in accordance with the principle of maximum membership according to the matrix. We can see from the algorithm that it is better to cluster for the data of normal distribution, and is more sensitive to the isolated point.

C. Problems about the Application of FCM

In recent years, cluster analysis has become an important technique of data mining. Although many clustering methods obtain a wide range of applications, there are problems of limitations and adaptability for every method. To the fuzzy clustering algorithm, FCM algorithm is the most widely used, but it also has many issues to exist, such as the number of the cluster must be preset by users, and select the appropriate number of clusters is the precondition for a precise clustering, but it is difficult to determine the number of clustering; as FCM algorithm is essential a local optimization technology, which uses the Hill Climbing to find out the optimal solution for iterating, it is particularly sensitive to initialize, and easily to fall into the local minimum so that can't obtain the global optimal solution.

IV. OPTIMIZATION OF FUZZY CLUSTERING ALGORITHM

The method of intrusion detection based on fuzzy clustering, FCM often combine with other methods to detect the intrusion, of course, there are many hybrid methods, such as the combination of FCM and adaptive immune system, the application of the average information entropy, the combination of FCM and support vector machine, the fuzzy genetic algorithm and etc.. The paper respectively introduced the research in recent years on how to obtain the number of clusters and the optimal solution in the following.

A. Determination of The Number of Clusters

Many studies aimed at the number of FCM clustering algorithm and the selection of the initial cluster centers, but the relevant research, only consider the determination of the number of clusters, or only to select the initial cluster centers. Ref. [4] introduced a method based on the average information entropy for the problem of the determination of the cluster number in FCM algorithm, and used the density function to obtain the initial cluster center.

When the division of the cluster is more reasonable, the data on the attribution of a cluster is more established, and the information entropy is smaller. Ref. [4] improved the clustering algorithm based on the information entropy, it used the average information entropy as the standard for determining the number of cluster. The concept of the average information entropy is defined as follows:

$$H(k) = -\sum_{i=1}^{C}\sum_{j=1}^{N}\{[u_{ij}\times\log 2(u_{ij}) + (1-u_{ij})\times \log 2(1-u_{ij})]/N\} \quad (5)$$

First, to define the range of the number of cluster, $[C_{\min}, C_{\max}]$. In (5), u_{ij} indicated the extent of the sample j belongs to cluster i, $u_{ij} \in [0,1], \forall i, j$. When k increase from $C_{\min}$ to $C_{\max}$, it can create $C_{\max} - C_{\min} + 1$ of $H_K(x)$. According to the regulation, the smaller information entropy value is, the more established the data belongs to a cluster, and then select the smallest $H_K(x)$, which corresponds to the cluster number k as the final cluster number C.

Ref. [5] proposed a fuzzy C-means and support vector machine algorithm (F-CMSVM) for automatic clustering number determination; it can solve the issue in fuzzy C-means algorithm (FCM) that clustering number has to be pre-defined. It used the algorithm of support vector machine with a fuzzy membership function and used the affiliating matrix which obtained by the introduction of support vector machine into fuzzy C-means algorithm as the fuzzy membership function, so that each different input data sample can have different penalty value, then it can obtain the optimized separating hyper-plane.

For the fuzzy C-means and support vector machine algorithm, first of all, it suppose that the given data set can be divided into 2 categories $(k = 2)$, and use the fuzzy C-means to cluster, then use the affiliating matrix which obtained by using the algorithm as the fuzzy membership of the fuzzy support vector machine algorithm, and train the data sets so that it can access to the support vector machine and the separating hyper-plane. To test the assumption, it had proposed a new standard in the paper: use the d_{SV} as the distance between the two categories. And d_{S1}, d_{S2} mean the average distance between the respective support vector of *S1*, *S2* and the nearest neighbor points. If $d_{SV} \le \min(d_{S1}, d_{S2})$, the original dataset can not be classified, and the assumption is not true; otherwise, it establish and the original dataset can be divided into two categories at least.

B. Obtain The Global Optimal Solution

For the problem that it is easy to fall into local minimum value caused by the sensitive initialization to FCM, people have been proposed the clustering method based on genetic algorithm, the method can converge to the global optimal value at a higher probability, but its' convergence rate is slower and it will appear the phenomenon of precocity. To solve the problem, Ref. [6] used the clonal selection algorithm (CSA) to optimize the objective function of the unsupervised FCM clustering algorithm; the clonal selection algorithm used the mechanism of antibody clone to construct a clone operator which combines characters of the evolutionary search, the global search, the stochastic search and the local search. As CSA based on the clone operator is a Groupization strategy, it has parallelism and randomness, so that it can obtain the global optimal solution with a

higher probability and converge at a high speed. At the same time, the algorithm considered the affinity of immune response in order to overcome the phenomenon of precocity. In the paper, it introduced the improved FCM algorithm into the anomaly intrusion detection

V. THE IMPROVED INTRUSION DETECTION ALGOITHM BASED ON FCM

How to improve the detection efficiency of IDS is the emphasis for people to research constantly. We cluster the data which is not only the numerical data but also is character data. For example, the KDD Cup 1999 network datasets, each connection instance contains 41 properties. In 41 properties, there are 3 flag properties and 38 numerical properties, at present, most of the studies were directed at the numerical properties of the sample data.

If we research the hybrid attributes of the sample data, the distance can only used into the numerical data, so we used a new method to solve the problem[7], we may defined the distance between x_i, x_j and k as fallow:

$$d(x_{ik}, x_{jk}) = \begin{cases} 1, x_{ik} \neq x_{jk} \\ 0, x_{ik} = x_{jk} \end{cases} \quad (6)$$

We supposed the object data has p numerical attributes and q categorical attributes, and the distance between the objects can be defined as fallow:

$$d(x_i, x_j) = d_n(x_i^{'}, x_j^{'}) + d_c(x_i, x_j) \quad (7)$$

In (7), $d_n(x_i^{'}, x_j^{'}), i \neq j$ is the distance of the numerical attributes, and $d_c(x_i, x_j)$ is the distance of the character attributes.

We can get the objective function of the hybrid attributes datasets; it can change from (2), that is:

$$J_m(U, C) = \sum_{i=1}^{N} \left\{ \sum_{j=1}^{k} u_{ij}^m \sum_{l=1}^{p} (x_{ik}^{'} - x_{jk}^{'})^2 + \lambda \sum_{j=1}^{k} u_{ij}^m \sum_{l=p+1}^{p+q} d_c(x_{ik}, x_{jk}) \right\} \quad (8)$$

In (8), weight λ is to balance the properties of hybrid attributes datasets, and the value λ is determined by the proportion of two kinds properties; $m > 1$ is the fuzzy coefficient, it uses to control the blur length of U. Supposed:

$$C_i^n = \sum_{j=1}^{k} u_{ij}^m \sum_{l=1}^{p} (x_{ik}^{'} - x_{jl}^{'})^2 \quad (9)$$

$$C_i^c = \lambda \sum_{j=1}^{k} u_{ij}^m \sum_{l=p+1}^{p+q} d_c(x_{ik}, x_{jk}) \quad (10)$$

As C_i^n and C_i^c are nonnegative, we can minimize C_i^n and C_i^c to make the $J_m(U, C)$ minimization, at the same time, we use the Lagrange multiplier method:

$$u_{ij} = \left\{ \sum_{l=1}^{k} \left[\frac{d(x_i, x_j)}{d(x_i, x_l)} \right]^{\frac{2}{m-1}} \right\}^{-1}, \forall i \quad (11)$$

We can iterate the process with (9), (10) and (11), as $m > 1$, the algorithm is convergent.

The improved intrusion detection algorithm based on FCM, which summarized as fallows [3]:

Step1: initialize the membership matrix U with random number between 0 and 1, and satisfy (1).

Step2: for the different attributes of data, respectively use (9) and (10) to calculate the cluster centers, C_i^n, C_i^c, $i = 1, \cdots k.$

Step3: use (11) to calculate the new membership matrix U.

Step4: calculate the value function according to (8). If it is smaller than a determined threshold or is smaller than the change with the last value function, then it will stop and output the clustering results. Otherwise, return to Step2 to continue iterating.

We not only consider the numeric data, also considering the character data, when we use the method to research the sample data. It more comprehensively analyzed the data to the clustering, not only helps to reduce the rate of the false alarm and the rate of the failing alarm, at the same time, combine the method with the optimized algorithm of FCM; it can further enhance the detection efficiency.

ACKNOWLEDGMENT

This work is supported by Economic Commence Market Application Technology Foundation Grant by 2007gdecof004.

REFERENCES

[1] Theodolidis S. Pattern Recongnition[M]. Second Edition, USA: Elsevier Science, 2003.

[2] Gao XB. Fuzzy Cluster Analysis and its Application[M]. XI'AN: Xidian University Press, 2004, 49–61.

[3] Yang DG. Research of The Network Intrusion Detection Based on Fuzzy Clustering[J]. Computer Science, 2005, 32(1): 86–91.

[4] Song QK, Hao M. Improved Fuzzy C-means Clustering Algorithm[J]. Journal Harbin University Science and Technology. 2007, 12(4):8–10.

[5] Xiao LZ, Shao ZQ, Ma HH, Wang XY, Liu G. An Algorithm for Automatic Clustering Number Determination in Network Intrusion Detection[J]. Journal of Software, 2008, 19(8):2140–2148.

[6] Xian JQ, Lang FH. Anomaly Detection Method Based on CSA-Based Unsupervised Fuzzy Clustering Algorithm[J]. Journal of Beijing University of Posts and Telecommunications, 2005, 28(4):103–106.

[7] Li J, Gao XB, Jiao LC. A GA-Based Clustering Algorithm for Large Data Sets with Mixed Numerical and Categorical Values[J]. Journal of Electronics & Information Technology, 2004, 26(8):1203–1209.

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 093-096

An Alarm Flow Decomposition Method for Security Threat Evaluation

Jie Ma, and Zhitang Li
Computer Science Department,
Huazhong University of Science and Technology, Hubei Wuhan, China
mjhust@163.com

***Abstract*—How to analyze security alarms automatically and find useful information form them has attract a lot of interests. Although many alarm correlation approaches and risk assessment methods have been proposed, most of them were implemented with high computational complexity and time consuming, and they can not deal well with huge number of security alarms. This work focus on performing an real-time security threat evaluation. We aggregate individual alarms to alarm flows, and then process the flows instead of individual alarms. Using the Singular Spectrum Analysis (SSA) approach, we found that the alarm flow has a small intrinsic dimension, and the alarm flow can be decomposed into leading components and residual components. Leading components represent the basic part and residual components represent the noise part of the flow. To capture the main features of the leading components forming the alarm flow, we accomplish the security threat evaluation. Case based experiments real network data shows the effectiveness of the method. To the best of our knowledge, this is the first study that applies SSA on the analysis of IDS alarm flows.**

***Index Terms*—alarm flow, threat evaluation, SSA**

I. INTRODUCTION

Internet has become a mission-critical infrastructure for governments, companies, institutions, and millions of every-day users. Because of this significant increase reliance on the Internet–based services, security and survivability of networks has become a primary concern. Intrusion Detection System (IDS) plays a vital role in the overall security infrastructure, as one last defense against attacks after secure network architecture design, secure program design and firewalls [1]. It gathers information form some key points in the computer networks, properly analyzes it and detect violations of the monitored system's security policy. so as to extend the security management capability of the system administrators and improve the integrity of information security infrastructure.

However, IDS are becoming unable to provide proper analysis and effective defense mechanism. They often report a massive number of elementary alarms of low-level security-related events. Since be overwhelmed by these alarms, administrators almost unable to make proper security threat evaluations in real-time. For this reason, some alarm correlation approaches were proposed. Ning et al. developed a an intrusion alarm correlator [2][3] to help human analysts to recognize multi-step attacks. Lee et al. [4][5] built a framework based on data mining techniques, such as sequential patterns mining and episodes rules, to search causal relationships between alarms to improve attack detection while maintaining a low false positive rate. Cuppens [6][7] build alert correlation systems based on matching the pre/post-conditions of individual alarms. The idea of the approach is that prior attack steps prepare for later ones. Julisch proposed to find alarm clusters and generalized forms of false alarms to identify root causes [8][9], and those alarms which are not possible attributed to the root causes can be filtered out. Although these correlation approaches enable to improve the alarm handling efficiency from micro prospective, they are hard to work well with massive security alarms and with large-scale network environment.

Recently, some online security threat evaluation and risk assessment methods have been developed. There are approaches employing a graph-based representation of systems [10][11], where an integrated, topological approach to network vulnerability analysis is used. But, these approaches usually follow a static procedure and cannot meet the changes in a dynamic network environment. Gehani [12] put forward a host-based method for real-time risk assessment. The model evaluates the threat probability from intrusion reports and uses predetermined attack scenarios to calculate the risk. Arnes [13] uses the Hidden Markov Model to compute the probability of the security status of the system based on observations from reporting of intrusion detection sensors. However, there are several limitations faced by these approaches. Since the high computational complexity and time cost, these approaches are still can not deal well with huge number of security alarms. In addition, the models used in these approaches should be retrained to adapt to new knowledge. Hence, it seems to be inefficient in reducing the human workload.

The focus of our work is to perform the security threat evaluation in real-time. Since IDS have the ability to perform online attack detection and can dynamically report security incidents which the network is suffering, we chose IDS alarms as our processing objects and use them to make threat evaluations. Different form the approaches mentioned above, we use aggregated *alarm flows* instead of individual alarms. We model the *alarm flow* as a time series, which is a sequence of alarm intensity observations i.e. the number of alarms in a sampling interval as a time series. Only alarms generated by the same IDS and with the same signature can be

aggregated to form the *alarm flow*. The aggregated flow revealed not only the overall regularities but also the behaviors of malicious events, even though the significance of individual alarms was unclear. Therefore, the flow characterization can be beneficial in enabling the administrator to have a real-time view of security threat situation.

Because of the dynamic characteristics of network environments, alarm flow is dynamic, huge, infinite and fast changing. So traditional stationary time series analysis techniques can not work well to adapt to changes in the flow. To overcome this problem, we apply Singular Spectrum Analysis (SSA) [14] approach on the alarm flow to process threat evaluations. We found that the alarm flow has a small intrinsic dimension, and the structure of alarm flow can be decomposed by two subsets of principal components, that is, the subset of leading components and the subset of residual components. Base on this, we can reconstruct the two parts of the original alarm flow, then threat situations hidden in the flow are visible and threat trends can be predicated.

The rest of the paper is organized as follows. In Section 2, apply the SSA on the alarm flows, the leading components which are responsible for the basic parts of the flow, and the residual components which represents noise part of the flow are separated. Section 3 process the threat evaluation with case studies and experiments. Section 4 concludes the paper and outlines future work

II. SSA-BASED FLOW DECOMPOSITION

The SSA method is a powerful non-parametric technique of time series analysis, and based on principles of multivariate statistics. It has been applied to many areas such as analyzing meteorological, climatic and geophysical time series [15]. The aim we using SSA is to make a in-depth understanding of the structures of the alarm flow. To understand the main features of the components forming the alarm flow is critical for threat evaluations.

Considering the original alarm flow *X(t)*, we assume it can be described by two parts: $\sigma(t)$ and $\varepsilon(t)$; that is, a decomposition of *X(t)* into a sum of two series: $x_t = \sigma_t + \varepsilon_t$, where, the series $\sigma(t)$ are associated with leading components of the flow, which forms the basic part of the information. It can be well approximated by linear recurrence formula $\sigma_t = \alpha_1\sigma_{t-1} + \cdots \alpha_d\sigma_{t-d}$ of order d with coefficients $\alpha_1 \cdots \alpha_d$. This implies that the basic part of the flow smooth enough to be modeled as weighted sum of previous observations. Then, the $\varepsilon(t)$ represents the residual components of the flow, which capture the small variations, and it cannot be well approximated by the finite-difference equations. In other word, it do not fit in the basic part of the alarm flow and can be interpreted as noise part of the flow. The noise may be a interference to the threat evaluations.

Then, we apply SSA to separate the two parts of the alarm flow. The process consists of four main steps, which are performed as flows:

·**Step 1:** Embedding. Let *X(t)*{ x_t : $1<t<N$} be the time series corresponding a alarm flow, and M ($1 < M < N$) be an integer called 'lag', and let $K = N - M + 1$. Making Hankelization process to form M-lagged vectors $X_k = \{x_k, \ldots, x_{k+M-1}\}^T$, $1 \le k \le K$. The trajectory matrix of the time series is of dimension $M \times K$ and has the following form:

$$\mathbf{E} = \left[X_1 X_2 \cdots X_K\right] \quad (1)$$

The trajectory space is defined as the linear space spanned by the columns of **E**.

·**Step 2:** Singular Value Decomposition (SVD). An SVD of the matrix $\mathbf{R} = \mathbf{E}\mathbf{E}^{\mathbf{T}}$ (we shall call **R** the lag-covariance matrix) provides us with M eigenvalues and eigenvectors. Let $\lambda_1, \lambda_2, \ldots, \lambda_M$ be the eigenvalues of **R** and they are arranged in the decreasing order, so that $\lambda_1 \ge \ldots \ge \lambda_M$, and r = max $\{i: \lambda_i > 0\}$. Let $U_1, \ldots, U_r$ denote the corresponding eigenvectors (principal components) and $V_j = \mathbf{E}^{\mathbf{T}} U_j / \sqrt{\lambda_j}$, $j = 1, \ldots, r$. Then the elementary matrix $\mathbf{E}_j$ are obtained from $\mathbf{E}_j = \sqrt{\lambda_j}\ U_j\ V_j^T$, and the trajectory matrix H can be written as

$$\mathbf{E} = \mathbf{E_1} + \cdots + \mathbf{E_r} \quad (2)$$

·**Step 3:** Grouping. Since we assume that alarm flows are composed by leading and residual components, we split the set of indices $I = \{1, \ldots, r\}$ into two disjoint subsets, $I_1 = \{i_1, \ldots, i_l\}$ contains the first few of leading components, and $I_2 = I - I_1$ contains the residual components. Let $\mathbf{E}^{I_1} = \mathbf{E}_{i_1} + \cdots + \mathbf{E}_{i_l}$ be the approximation of the trajectory matrix **E** based on the indices of I_1. Similarly, we have $\mathbf{E}^{I_2}$ for the indices of I_2. Then the final decomposition of the initial trajectory matrix **E** is

$$\mathbf{E} = \mathbf{E}^{I_1} + \mathbf{E}^{I_2} \quad (3)$$

·**Step 4:** Reconstruction. Approximation of the original time series can be performed by averaging the elements of $\mathbf{E}^{I_k}$ ($k=1,2$), which are placed on the same anti-diagonal, that is, the elements $e_{ij}^{I_k}$ with $i + j$ = constant. Then the decomposition of original time series is

$$X(t) = X^{I_1}(t) + X^{I_2}(t) \quad (4)$$

where $X^{I_1}(t)$ is the time-series reconstructed using the subset of leading components, which is basic part of the alarm flow *X(t)*, and $X^{I_2}(t)$ is the time series reconstructed using the subset of residual components, which is noise part of $X\{t\}$. It is worth pointing out that $X^{I_1}(t)$ and $X^{I_2}(t)$ are in the same sense with σ_t and ε_t respectively.

III. THREAT EVALUATION WITH EXPERIMENTS

In this section we apply the above SSA decomposition algorithm on the alarm flows. Several experiments were conducted to verify the effectives of the threat evaluation.

A. Data Preparing

To generate our test data, snort [16] was chosen as our IDS, which is a popular open-source Network Intrusion Detection System (NIDS). The data was accumulated from our campus network for one month by activating default rule sets of the Snort. Interestingly, more than 200 signatures had triggered alarms, We choose *5* minutes as the sampling interval to generate the alarm flows.

B. Parameter Selection

Before the decomposition, we should select the values of just two parameters, namely, the lag window length M and the number l of leading components contained in the subset $I_l = \{i_1, \ldots, i_l\}$. The SVD performed on matrices obtained with a window length M is equivalent to that performed on matrices obtained with the complementary window $K = N - M + 1$. This means increasing the window length would reproduce results already tested with shorter window lengths, and too large window length may introduce some disturbance. Where, l is such that the first l components provide a good description of the normal part of the signal and the lower M-l components correspond to abnormal part. If l is too small (under-fitting), we miss a part of the normal signal. Alternatively, if l is too large (over-fitting), then we approximate a part of abnormal signal with the normal part. Both of these cases will make it difficult for us to detect the deviations from the normal profile.

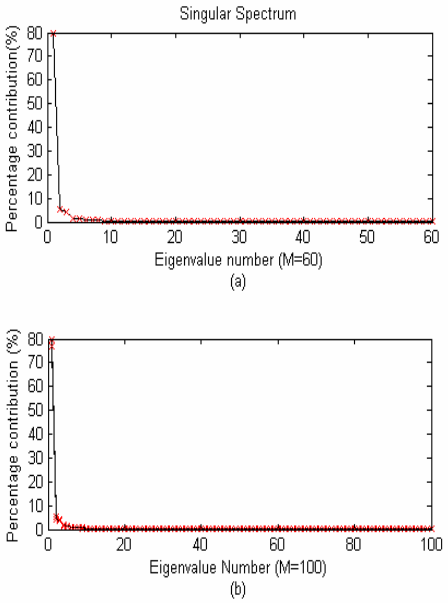


Fig.1 Percentage contribution of the eigenvalues: (a) $M_1 = 60$, (b) $M_2 = 100$.

To properly choose the parameters, we use the radio

$$P_i = \frac{\lambda_i}{\sum_{j=1}^{M} \lambda_j} \tag{5}$$

to estimate the energy contribution of the i-th candidate principal component to the original flow. Fig.1 presents the contribution of the eigenvalues $\{\lambda_i\}_{i=1}^{M}$ corresponding to the time series of a alarm flow with signature *ICMP L3retriever Ping*. wo different widow lengths $M_1 = 60$ and $M_2 = 100$ are used for comparing and anglicizing the results.

As it can be seen from both cases, only the first few eigenvalues are responsible for the main part of the flow information. This means we can reconstruct the major behavior of the alarm flow by using the first few leading components. In fact, the first 3 eigenvalues account for more than 90 percent of the whole energy, therefore we prefer to choose $M = 60$ and $l = 3$ to continue the following experiments and analysis.

C. Case Study

We use two case to make the threat evaluation, one is a alarm flow created by the signature *BARE BYTE UNICODE ENCODING (http_inspect)*, the other is the alarm flow created by the signature *ICMP L3retriever Ping***.** Both cases use one day's data to perform SSA-based decomposition and signal reconstructing. The reconstructed signal $X^{I_1}(t)$ is plotted in Fig.2 (a), Fig.3 (a) along with the original signal, and the residual series $X^{I_2}(t)$ can be easily obtained from the equation (4) by $X^{I_2}(t) = X(t) - X^{I_1}(t)$, which is plotted in Fig.2 (b), Fig.3 (b).

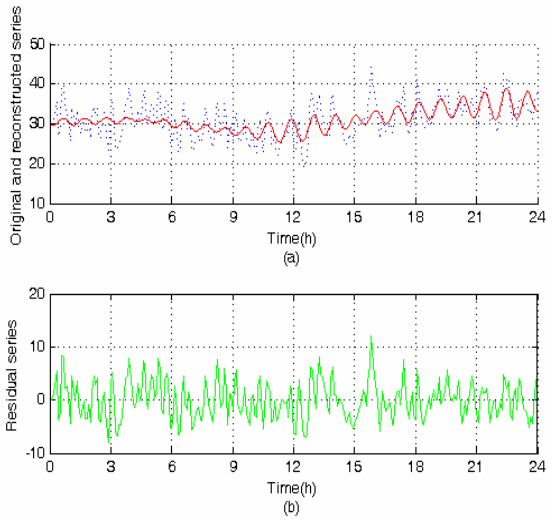


Fig.2 Case 1 for *http_inspect* flow: (a) Original alarm flow series (dotted line) and SSA-reconstructed series (continuous line) corresponding to normal flow behavior. (b) Residual series defined as the difference between the original and reconstructed series.

As seen from the pictures, *http_inspect* flow play a more stable behaviors, which we can conclude that the threat level is not changed along with the timeline. Then the security administrator may neglect the alarms corresponding this signature. Different with the case 1, it seems that there are some changes in *ICMP L3retriever Ping* flow, these changes should be paid more attentions to verify if there some new malicious behaviors occur, and result in raising the security threat level. Although the residual series represents the noise part of the flow, its behaviors also reveal some useful information. For

example, changes and even slight divergence form the basic part of the flow can be revealed by it.

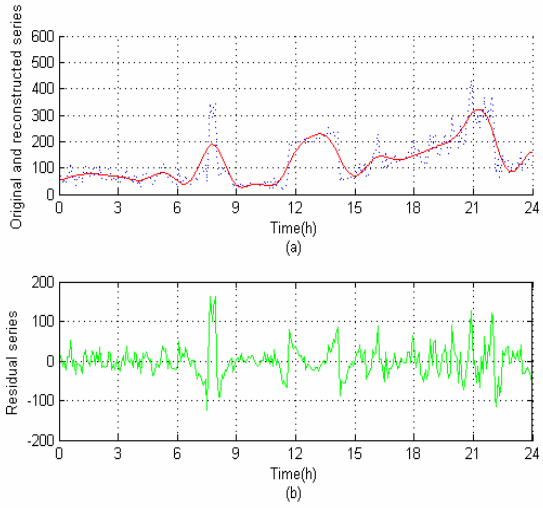


Fig.3 Case 2 for *ICMP L3retriever Ping* flow: (a) Original alarm flow series (dotted line) and SSA-reconstructed series (continuous line) corresponding to normal flow behavior. (b) Residual series defined as the difference between the original and reconstructed series.

IV. Conclusions

Network threat identification and evaluation aims to extract knowledge of current security threat and status from raw security data. In practice, identifying the security incidents of true threat from IDS alarms is really difficult because the amount of alarms is usually overwhelming and there are a lot of redundant and false alarms. In this paper we focus processing aggregated alarm flow in real-time. Using the SSA, we first explore the intrinsic dimensionality and structure of the time-series corresponding to alarm flow. Then we capture the leading components, which represent the main part of the flow, to make threat evaluation. Reconstructed signals enable the administrator to have a real-time view of security threat situation. In addition, the threat evaluation can work well in dynamic and non-stationary network environment. Based on the approach, we do not need to know the parametric model of the considered time series and have the ability to autonomously adapt to shifts in the structure of the alarm flow. We believe that the method could be used as such, or in complement to other means of correlation, to monitor alarms. In the next work, we plan to focus on automatically detecting changes in alarm flow and further improve our threat evaluation approach.

Acknowledgment

This work is supported by the National Natural Science Foundation of China under Grant No. 60573120, and by the National High Technology Research and Development Program of China (863 Program) under Grant No. 2007AA01Z420, and by the key project sponsored by Natural Science Foundation of Hubei Province under Grant No. 2008CDA021.

References

[1] J. Allen, A. Christie, et al., (2007) State of the Practice of Intrusion Detection Technologies. Available via Software Engineering Institute. http://www.sei.cmu.edu/publications/documents/99.reports/99tr028/99tr028abstract.html.

[2] P. Ning, Y. Cui, and D. S. Reeves: Constructing attack scenarios through correlation of intrusion alerts. In Proceedings of the 9th ACM Conference on Computer and Communications Security, Nov 18-22 2002, Washington, DC, United States, 2002.

[3] P. Ning, Y. Cui, D. S. Reeves, and D. Xu: Techniques and tools for analyzing intrusion alerts. In: ACM Transactions on Information and System Security, vol. 7, pp.274, 2004.

[4] W. Lee, S. J. Stolfo, A Framework for Constructing Features and Models for Intrusion Detection Systems, ACM Transactions on Information and System Security 3(4) (2000) 227–261

[5] X. Qin and W. Lee. Statistical Causality Analysis of INFOSEC Alert Data. In Proc. of the RAID'03, Springer–Verlag, 2003.

[6] F. Cuppens.: Managing alerts in multi-intrusion detection environment. In: Proceedings 17th annual computer security applications conference. New Orleans; 2001.

[7] F. Cuppens, A. Miege: Alert correlation in a cooperative intrusion detection framework. In: Proceedings of the 2002 IEEE symposium on security and privacy; 2002.

[8] K. Julisch and M. Dacier. Mining Intrusion Detection Alarms for Actionable Knowledge. In Proc. of the SIGKDD'02, 2002.

[9] C. Kruegel and W. Robertson. Alert verification: Determining the success of intrusion attempts. In Proc. of the DIMVA'06, Dortmund, Germany, July 2006.

[10] P. Ammann, D. Wijesekera, and S. Kaushik, Scalable, Graph-Based Network Vulnerability Analysis, Proceedings of the 9th ACM Conference on Computer and Communications Security, New York: ACM Press, 2002, 217–224

[11] R. Ritchey and P. Amman, Using Model Checking to Analyze Network Vulnerabilities, Proceedings of the 2000 IEEE Symposium on Security and Privacy, pp. 156-165, 2000

[12] A. Gehani, G. Kedem Rheostat: Real-time Risk Management. In Proceedings of the 7th International Symposium on Recent Advances in Intrusion Detection, 2004.

[13] A. Arnes, F. Valeur, G.. Vigna., R. A. Kemmerer,Using Hidden Markov Models to Evaluate the Risk of Intrusions. in: Proceedings of the International Symposium on the Recent Advances in Intrusion Detection(RAID 2006): Springer-Verlag, 2006. 145-164.

[14] N. Nekrutkin, V. Zhigljavsky, 2001. Analysis of Time Series Structure—SSA and Related Techniques. Chapman & Hall/CRC, Boca Raton, FL, pp. 13-78.

[15] Vautard, R., Yiou, P., Ghil, M. Singular-spectrum analysis : a toolkit for short, noisy chaotic signals. Physica D, vol. 58, pp. 95-126, 1992.

[16] B. Caswell, M. Roesch (2004) Snort: The open source network intrusion detection system. Available via Snort. http://www.snort.org/

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 097-100

Fuzzy Control of Intelligent Vehicle Based on Visual Navigation System

Tingjian Zhong[1], and Meilian Qiu[2]
[1] Jiangxi Vocational & Technical College of Electricity, Nanchang, China
Email:jxdlztj@163.com
[2] Nanchang Power Supply Company, Nanchang, China
Email:qiuqiu6872@163.com

***Abstract*—The research on intelligent vehicle mainly includes safety monitoring, intelligent anticollision, aided driving, auto driving, behavior planning decision-making, system structure, and synthetical integration, etc. Sensor and control algorithm are major factors influencing the development of intelligent vehicle. This paper introduces an intelligent vehicle system with chip of FRSSCALE MC9SDG128. By adding fuzzy algorithm into this research on turning control angle of intelligent vehicle, it makes decision according to the lateral error and orientation error. This system, which includes automatic recognition and finished special function, is simple and useful, lower requirement to hardware and has capability of adapting existing structured road environment.**

***Index Terms*—Intelligent vehicle, path tracking, fuzzy control**

I. INTRODUCTION

Intelligent vehicle, which is also called wheel Mobil Robot, is a synthetical system [1][2] that contains environmental perception, planning and decision-making, auto-driving and other functions. Intelligent vehicle involves many fields, such as computer science, communications, artificial intelligence, signal processing, pattern recognition, control theory and so on. Intelligent vehicle has a wide range of applications prospect in areas such as the military, civil and scientific research. It has attracted the attention of large companies and governments. From the mid-and late eighties last century, the world's major developed countries have launched a series of research and have effective development on intelligent vehicle. Road detection technology is not only as an important key technologies, but also as an important indication of the level in intelligent vehicle visual navigation system. This paper focuses on intelligent vehicle control system in terms of speed, looking for line control. Based on the traditional idea of fuzzy control, in accordance with the specific requirements of the process of the intelligent vehicle speed, line regulation, and speed of the actual situation, a fuzzy parameter self-tuning fuzzy control method is introduced. Experimental results show that this method is suitable for intelligent vehicle for speed and line-conditioning.

II. INTELLIGENT VEHICLE DESIGN PROGRAM

Hardware design of controller not only affects the overall performance of the intelligent vehicle, but also related to the manufacturing cost. While choose hardware, these following main factors should be considered: first, hardware requires good reliability; second, use real-time processing of visual-guided high-speed method to lead the camera image for information collection; In addition, it should determine the performance of the sensor correctly to meet the requirements of intelligent vehicle functions.

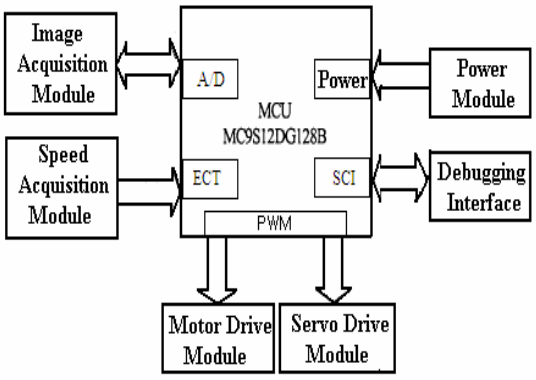


Figure 1. framework of intelligent vehicle haredware system

Intelligent vehicle contains six modules: control processor MC9S12DG128, servo drive module, motor driver module, image acquisition module, speed acquisition module and auxiliary debugging module. Where, S12 singlechip is the hardcore of this system. It is responsible for receiving image data of path and some information such as speed feedback, and dealing with the information correctly into the control volume for controlling the drive motor and steering gear. Steering gear module and drive module, respectively, response for steering and driving of this vehicle model. image acquisition module is composed of S12 AD module, the chip LM1881 chip, peripheral circuits and camera. Its function is to obtain the image data of the path in front for further analysis of S12. Speed acquisition module, which is composed of the reflective photoelectric sensor and black-and-white ribbons attached to the driven gear, detect the cumulative number of pulse reflection models to get the speed value. Auxiliary debugging module is for writing program of vehicle model, debugging and testing functions, as well as the settings for state control of intelligent vehicle, system parameters and operation of policy and so on. The hardware system diagram of this vehicle structure is shown in Fig.1.

III. DESIGN FOR INTELLIGENT VEHICLE CONTROLLER

Intelligent vehicle control algorithms includes: Rear-wheel speed control algorithm, control algorithm for front-wheel steering angle, and the co-ordination control

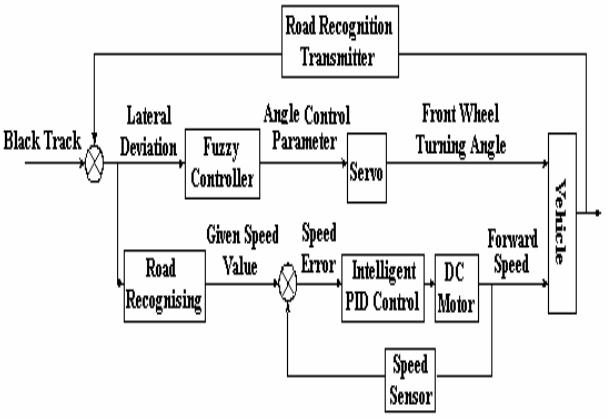


Figure 2. structure of intelligent vehicle control algorithm

method between the two. The diagram of intelligent vehicle control algorithm is shown in Fig.2.

Intelligent vehicle control algorithm is the core of the whole system. The current path tracking control methods of intelligent vehicle are PID algorithm and fuzzy control algorithm. PID algorithm is a traditional control method, whose main character bases on precise object mathematical control model. Many factors effect intelligent vehicle system, such as motor character differences, path roughness, mechanical properties, and the meansure precision change of camera resulted from the shake while vehicle move. The whole system represents as nonlinearity and uncertainty, and it is difficult to be described as a precise mathematical model [3]. The traditional method of control is to use inaccurate models and a fixed control algorithm, which results whole system lack of agility and robotness, and it's hard to get good control results, even cause system unstable. Therefore, a fuzzy control strategy is adopted to achieve the path of trolley track in this paper.

IV. STEERING CONTROL DESIGN

A. *The basic principles of fuzzy control*

A typical fuzzy control system is shown as in Fig.3. Fuzzy controller is its key part. This diagram shows that fuzzy controller is composed of by the fuzzication interface, rule base, inference module and clear interface four parts [4][5]. Fuzzification interface is to change input variables into corresponding fuzzy ones with fuzzication. Rule base is to supply control rules for inference module. Inference module is to finish fuzzy inference based on the input and rule base, then solve fuzzy relation equations, in order to obtain the volume of fuzzy control completion. Clear interface is to transfer the output fuzzy control parameters into precision ones, in order to get use of the controlled object. Fuzzy control rules usually be gained from computer programme. A

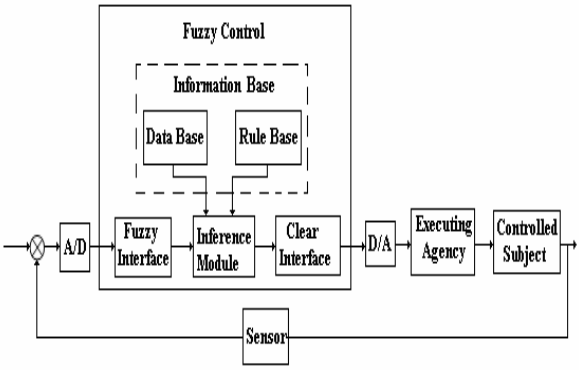


Figure 3. Fuzzy Control System

fuzzy controller designing has three steps: analyse this controlled object, choose the subjection function and control rules of fuzzy input and output variable, determine the fuzzy consequence and anti-fuzzy method.

B. *Steering fuzzy controller design[6][7][8]*

After identification information from the module of intelligent vehicle road traffic is given, the steering control algorithm can make intelligent vehicle better be able to track the trajectory forward. Fuzzy control method have be chosen because it is difficult to gain accurate mathematical model and require high dynamic performance characteristics. Fuzzy control algorithm does not need accurate mathematical model of the system. Based on a certain experience and experiment and it can get to good control result such as a short response time, small overshoot and good robust performance.

Fuzzy controller design: A two-dimensional fuzzy controller is chosen as the fuzzy controller of intelligent vehicle. Two input variables e and α , respectively represents lateral error and orientation error of vehicle location and the road center. These errors are given from the sensor located in the front of vehicle. Output variable is this intelligent vehicle's front wheel steering angle β , which can be changed with control duty cycle of PWM output wave.

Fuzzy precise variable: Choose the domain of the input and output variables according to the data collected from camera. Theoretically, location error e changes in the scale of $[-\infty,+\infty]$, and $[-90^{\circ},+90^{\circ}]$ for α and β. But actually when the values of e and α are too big, intelligent vehicle will get far away from the guide line, and it is hard to detect information and can not correct the deviation error automatically. So in actual these basic domain is limited. For the actual situation here, the coordinates of the real world can be determined with pixel dot. So location error e (blackline_tempp) can be represents error of pixel dot. That is the error, which is between the pixel dot of black center line and the center line position of image, its basic domain is [-16, +16] and unit is pixel. Basic domain of β is $[-45^{\circ},+45^{\circ}]$. Set e as negative when path centerline is to the left side of longitudinal line of bodywork, and as positive to the right. Choose clockwise rotating direction as positive for

both α and β.

determination of Quantitative factor and proportion factor: Choose $\{-6,-5,-4,-3,-2,-1,0,1,2,3,4,5,6\}$ as fuzzy domains of e、α、β on the consider of Real-time performance and complexity of control. Fuzzy domain {-n,+n} is the discretization error of continuous change in basic domain {-m,+m}. Where, quantitative factor $k = n/m$, and the basic domain of location error e is [-15，+15]. Then quantitative factor of location error $k_e = 6/15 = 2/5$. For system controlled object β, whose domain is $[-45^\circ, +45^\circ]$, this proportion factor of location error $k_\beta = 6/45 = 2/15$.

Selection of fuzzy linguistic variables and its fuzzy subsets: According to system's actual circumstance, linguistic variables are dicided into seven values as "positive big"(PB), "positive middle"(PM), "positive small"(PS), "zero" (ZO), "negative small"(NS), "negative middle"(NM) and "negative big"(NB). Fuzzy subsets of linguistic variables' domain are discribed with subjection function $\mu(x)$. For a domain $\{-6,-5,-4,-3,-2,-1,0,1,2,3,4,5,6\}$, of fuzzy subsets on those seven defined linguistic variables PB, PM, PS, ZO, NS, NM and NB, the values of maximum subjection "1" is set as follow.

$$\mu_{PB}(X)=1 \qquad X=6$$
$$\mu_{PM}(X)=1 \qquad X=4$$
$$\mu_{PS}(X)=1 \qquad X=2$$
$$\mu_{Z0}(X)=1 \qquad X=0$$
$$\mu_{NS}(X)=1 \qquad X=-2$$
$$\mu_{NM}(X)=1 \qquad X=-4$$
$$\mu_{NB}(X)=1 \qquad X=-6$$

Things are tended to be judged as the characteristic of normal distribution, so the subjection function of fuzzy set is dicided with normal distribution.

$$\mu(x)=\exp\left(-\left(\frac{x-a}{b}\right)^2\right) \qquad (1)$$

Where, parameter a represents respectively 6, 4, 2, 0, -2, -4, -6 for the fuzzy sets PB, PM, PS, ZO, NS, NM, NB. Parameter b is chosen positive which is bigger than 0. The value of a has greater impact on the control. The bigger value of b, the wider $\mu(x)$ diagram, the lower resolution characteristics and error control sensitivity. On the contrary, smaller value of b will be together with thiner $\mu(x)$ diagram, better resolution characteristics and error control sensitivity. Therefore, choose the fuzzy sets with low resolution characteristics subjection function while in big system error. With small system error or even near 0, fuzzy sets of high resolution characteristics subjection function is adopted, and b is 1 here.

For a integer domain N, these linguistic variables can be represented with different ways, such as the method of table, formula or graphics and so on. If the Gaussian subjection function is chosen to transform fuzzy linguistic variables value of input and output, the values of the linguistic variables above are shown as in table 1.

According to pre-given inputs e_i and α_i, as well as its quantitative factor, gain the quantitative level of e_i and

TABLE I.
assignment table of linguistic variable value

subjection / class / variable	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6
PB	0	0	0	0	0	0	0	0	0	0	0.2	0.7	1.0
PM	0	0	0	0	0	0	0	0	0.2	0.8	1.0	0.8	0.2
PS	0	0	0	0	0	0	0.1	0.8	1.0	0.8	0.1	0	0
ZO	0	0	0	0	0	0.5	1.0	0.5	0	0	0	0	0
NS	0	0	0.1	0.8	1.0	0.8	0.1	0	0	0	0	0	0
NM	0.2	0.8	1.0	0.8	0.2	0	0	0	0	0	0	0	0
NB	1.0	0.7	0.2	0	0	0	0	0	0	0	0	0	0

α_i in basic domain from $n_{ie} = k_e \times e_i$ and $n_{i\alpha} = k_\alpha \times \alpha_i$. From linguistic variables transforming table, identify the fuzzy set of the element n_i on biggest subjection corresponding linguistic value. This fuzzy set is right the fuzzification of input value.

V. EXPERIMENTAL RESULTS AND ANALYSIS

Figure 4. PC debugging interface

Code Warrior 3.0 of Metroworks company is used in this system for exploiture and debugging. A debugging host computer interface, which is shown in Fig.4, is producted in order to get more basic data and signal information for actual testing. The use of VC and Microsoft MSComm control, makes communication hardware more simple. After communication protocol is set up, all data collected from single-chip microcomputer can be sent to the host computer through the serial port for observation and testing.

A. Image information acquisition analysis

With the intelligent vehicle debugging software of SmartLab1.0, road image can be acquired from intelligent vehicle camera. According to the quality of images, whether the result of image acquisition system is good or not can be judged. The image gotten from intelligent vehicle's carema is shown in Fig.5.

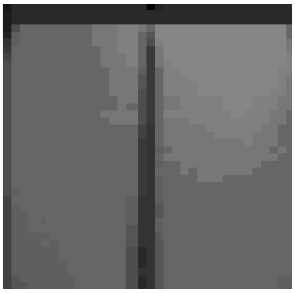


Figure 5. Straight Road Image

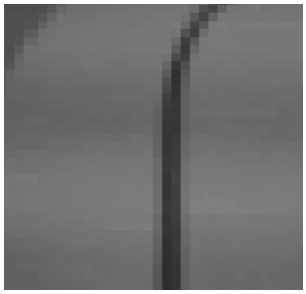


Figure 6. Curve Road Image

The image is 40 × 71 pixels, and it has been enlarged here for easier looking and recognised. the image of a curve road is shown in Fig.6. It is a map of right-angle corner, that is a right-angle of straight road and then right turn. In this image it's clear to see that it's a right turn.

B. Real vehicle test results

Experimental test is carried out in order to check the navigation and control system of intelligent vehicle. Some typical path such as S-path and different angle curve road are chosen in test. The picture of intelligent vehicle searching path in the S-path test is shown in Fig.7. The maximum speed for this road conditions is 1.6m / s. The picture of intelligent vehicle driving at turning is shown in Fig.8. The maximum speed for this road conditions is 1.6m / s.

Figure 7. intelligent vehicle in S-path test

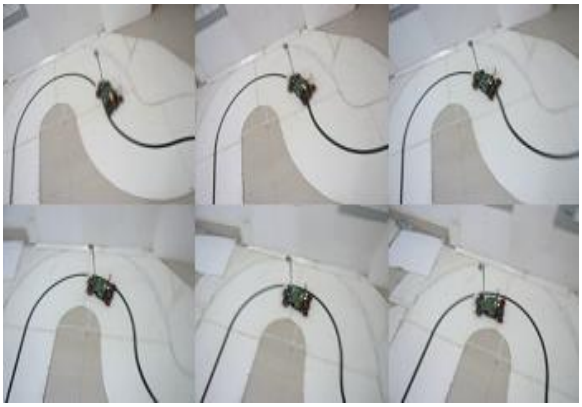


Figure 8. intelligent vehicle driving at turning.

VI. CONCLUSION

Seen from Fig.5 and Fig.6, that black navigation line at road centre can be recognised and displayed in the picture clearly. It's also found that picture quality will be affected of the light strength and the projection uniformity of the light on the objects. With enough and uniform light, it will be better in the contrast and color evenness. It is suggested in Fig.7 that intelligent vehicle has good performance of searching line, because it can adjust the front wheel turning angle in time in accordance with S-path's change. With Fig.8 the same conclusion can be gained: when intelligent vehicle rushs out of path because of too fast speed, the algorithm can find that it is out of control and give orders of returning to the path.

REFERENCES

[1] Huang K, Jin H, and Jiang D. , "Overview of Design for Korea Intelligent Model Car Design Contest," ELECTRONIC ENGINEERING & PRODUCT WORLD, 2006(5), pp. 150-156.

[2] Foster I, Kesselman C, and Nick J M, etal, "Grid Services for Distributed System Integration," Computer, 2002,35(6).

[3] Foster I, Kesselman C., "The Grid Blueprint for a New Computing Infrastructure," Beijing: Mechanical Industry Press, 2005.

[4] Hersbkop S, FersterR, and Bui L H, etal, "Host-based Anomaly Detection Using Wrapping File Systems," CU Tech Report, April 2004.

[5] Eskin E., "Probabilistic anomaly detection over discrete records using inconsistency checks," Technical report, Columbia University Computer Science Technical Report, 2002.

[6] DASH P K, W A C., "Anticipatory fuzzy control of power systems" IEE Proc Cener Transm Distrib, 95, 2 (2), pp. 211-218.

[7] FRANK KLAWONN, ORG GEBHARDT, and UDOLF KRUSE, "Fuzzy ontrol on the basis of equality relations with an example from idle speed control," IEEE Transactions on Fuzzy Systems, 1995, 3 (3), pp.336 -350.

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 101-104

Prediction of Reservoir-induced Earthquake Based on Fuzzy Theory

Ming Zhong, and Qiuwen Zhang
College of Hydropower and Information Engineering, Huazhong University of Science and Technology
Wuhan, Hubei, China
zhongming_youye@sina.com; qwzhang_hust@163.com

***Abstract*—With more and more reservoirs have been and are being built all over the world, reservoir-induced earthquake has received a great deal of attention from geoscientists mainly because of its potential to damage constructions and to cause human losses. Based on the previous researches on the environmental conditions of reservoir-induced earthquake, a criteria hierarchy model has been constructed. We give the environmental condition an uncertain description by applying weight in this theory, which shows that the different factor plays different role in risk assessment. Taking one candidate assessment unit into consideration, the risk class through fuzzy theory is obtained. Based on the original mechanism, a fuzzy mathematical prediction model of three-gorge reservoir in China has been created. It is concluded that compared with other mathematical prediction models, the reservoir-induced earthquake prediction model based on the fuzzy theory can analyze data with fuzzy processing and therefore possesses advantages of high prediction accuracy.**

***Index Terms*—reservoir-induced earthquake, fuzzy theory, risk prediction, uncertain analysis, three-gorge reservoir**

I. INTRODUCTION

Reservoir-induced earthquake is a special engineering geological problem resulting from development of hydraulic resources, which occurred in reservoir and its peripheral.

Previous researches emphasis on differential features analysis, origin mechanism and formation condition among several earthquake examples. Prof. Gupta [1], a world-famous expert in reservoir-induced earthquake, investigated the formation condition and mechanism of reservoir-induced earthquake by analyzing the seismicity examples of Koyna Reservoir in India, Nuek Reservoir in Tajikistan, Lake Mead in America and Xinfengjiang Reservoir in china in his classic study "Dam and Earthquake"; Denlinger [2] studied geological environmental conditions relating to Reservoir-induced earthquake in Geysers reservoir in northern California of the United states; Simpson [3] classified Reservoir-induced earthquake into two major categories: water permeation and additional strain; Keith [4] analyzed the relationship between reservoir-induced earthquake and crustal deformation in Nuek Reservoir in Tajikistan; Chadha [5] investigated the relationship between reservoir-induced earthquake and water lever change. Edelman [6] presented a rigorously derived analytical method to describe and interpret the low magnitude earthquakes caused by injection of the borehole fluids into surrounding porous reservoirs.

With the accumulation of observation data of reservoir-induced earthquake, base on the preview paper, works in this area have been pursued by different investigators to relate different geological features with the occurrence of the reservoir-induced earthquake or to predict the magnitude of water-induced earthquakes. Many prediction theories with different characteristics are proposed, such as Geology and environmental method, Bayesian probability method, Gray clustering method, Patten recognition method and Logic of information method. Habibagahi [7] assessed the risk of reservoir-induced earthquake by RBF neural network; Fahjan [8] forecasted the risk of reservoir-induced earthquake by the change of water; Ping Li [9] analyzed the capacity of seismogenesis at active faults in the Three-gorge Reservoir of The Yangtze river; Qifa Xia [10] assessed seismic hazard according to engineering geology and earthquake geology; Qinyun Wang [11] used Probability Analytic Method and geological Environment Method to assess seismic hazard of The Three-gorge Reservoir; Houqun Chen [12] established the database of reservoir-induced earthquake and predicted the seismic risk by using GIS technology; According to the record of the LiJiaXia water-induced seismicity in 1998-2002, Xin Sun [13] analyzed seismicity feature after the reservoir impoundment and then predicted seismic hazard in the future; Zongji Ma [14] proposed a reservoir-induced earthquake prediction and evaluation Model based on Digital Watershed; Qiuwen Zhang [15] proposed a model to quantitatively predict and assess the risk of reservoir-induced earthquake, the results are descried with risk probability of reservoir-induced earthquake. More recently the theory of fuzzy sets has been used to evaluate the risk of reservoir-induced earthquake. So far, the cause and mechanism of the reservoir-induced earthquake is a matter of speculation and requires further research.

A quantitative and qualitative prediction theory has been presented in this research, which evaluate the risk of reservoir-induced earthquake based on the analytic hierarchy process (AHP) model, and then apply this fuzzy comprehensive evaluation theory in every divided prediction unit.

This paper is organized as follows: In section 2, the evaluation criteria hierarchy is constructed; it is the result

Corresponding author: Qiuwen Zhang (qwzhang_hust@163.com)

AP-PROC-CS-10CN006

of "Quantitative Prediction and Assessment of Induced Area and its Neighboring Area"; in section 3, apply AHP method on weights calculation. The case study is given in Section 4.

II. HIERARCHY EVALUATION MODEL

According to the investigation result in previous researches, this paper constructs the fuzzy evaluation system of reservoir-induced earthquake in two main factors such as water permeation-accumulations factor (U_1) and strain energy accumulation-elimination factor (U_2), and they are corresponding to six sub-factors such as rock (R), karsts (K), fracture (F), crack (C), water load (W) and fracture angles of superimposed (A). Let U be the set of all induced earthquake indexes:

$$U = \{U_1, \quad U_2\} \qquad (1)$$

$$U_1 = \{R, \quad K, \quad F, \quad C\} \qquad (2)$$

$$U_2 = \{W, \quad A\} \qquad (3)$$

Remark set is $V = (V_1 \quad V_2 \quad ... \quad V_n)$, and V_t $(t = 1, 2, ...n)$ shows remark from low to high. This paper adopts four levers of earthquake scale, and takes n=4, define remark set:

$$V = \{V_1, \quad V_2, \quad V_3, \quad V_4\} \qquad (4)$$

Where V_1 stands for micro-earthquake, V_2 stands for sensible earthquake, V_3 stands for devastating earthquake, V_4 stands for strong earthquake.

The state classifications of reservoir-induced earthquake factors show in TABLE I. The hierarchy evaluation model shows in Figure 1.

III. FUZZY MATRICES AND WEIGHTS

The fuzzy matrixes and weights show below:

TABLE Ⅰ.
THE FUNDAMENTAL SCALE OF ABSOLUTE NUMBERS

Attributes	Alternatives	explanations
R	1	Igneous rocks
	2	Metamorphic rocks
	3	Non-carbonates Sedimentary rocks
	4	Carbonates sedimentary rocks
K	1	Undeveloped
	2	Poor developed
	3	Developed
	4	Well developed
F	1	Undeveloped, far from fault
	2	Undeveloped, near by fault
	3	Developed
	4	Well developed
C	1	Undeveloped
	2	Poor developed
	3	Developed
	4	Well developed
W	1	Not submerge, far from reservoir
	2	Not submerge, near by reservoir
	3	Submerge, peripheral regions of reservoir
	4	Submerge, in the middle of reservoir
A	1	A=0~10; 71~90
	2	A=11~24; 51~71
	3	A=25~50
	4	A=25~50

Seismicity Risk in Yangtze Three-Gorge Reservoir Head

A. Fuzzy Matrix

Fuzzy matrix is constructed to calculate the weights, each factor in an upper level is used to compare the factor in the lever immediately below with respect to it. To make comparisons, we need a scale of numbers that indicates how many times more important or dominant one factor is over another factor with respect to the criterion or property with respect to which they are compare, TABLE II exhibits the scale.[16]

The fuzzy matrix of U_1 is show in TABLE III, the fuzzy matrix of U_2 is show in TABLE IV.

B. Weights

After established reciprocal matrix, ascertain the weights of each lever of evaluation index by solving characteristic vectors of the matrix.

Suppose $W = (W_1 \quad W_2 \quad W_3 \quad \cdots)$ is weight of the index, Where $0 < w_i \le 1$, $\sum_{1}^{n} w_i = 1$

$$A_{n\times n} \bullet W = \lambda_{\max} \bullet W \qquad (5)$$

We get weight sets of sub-factors:

$W_1 = (0.5650 \quad 0.0553 \quad 0.2622 \quad 0.1175)$

$W_2 = (0.90 \quad 0.10)$

And weight set of main-factors is:

$W = (0.6667 \quad 0.3333)$

The consistency ratio (CI) is used to directly estimate the consistency of pairwise comparisons. The closer inconsistency index to zero, the greater the consistency. If the CR less than 0.10, the comparisons are acceptable, otherwise, the decision makers should go back and redo the assessments and comparisons.

TABLE Ⅱ.
STATE CLASSIFICATION OF RIS FACTORS

Intensity of importance	Definition
1	Equally importance
3	Moderate importance
5	Strong importance
7	Very strong importance
9	Extreme importance
2,4,6,8	Interval values between two adjacent choices
reciprocal	Less important level

TABLE Ⅲ.
FUZZY MATRIX OF U_1

	R	K	F	C
R	1	7	3	5
K	1/7	1	1/5	1/3
F	1/3	5	1	3
C	1/5	3	1/3	1

TABLE Ⅳ.
THE FUZZY MATRIX OF U_2

	W	A
W	1	9
A	1/9	1

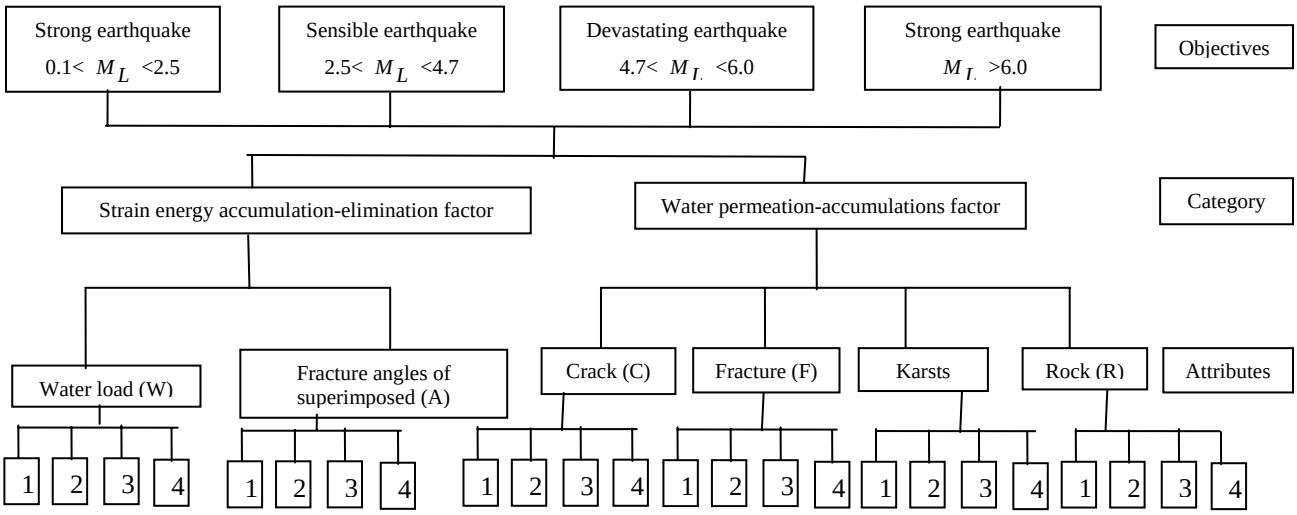


Figure 1.The hierarchy evaluation model

IV. CASE STUDY

As case study, Three-gorge reservoir and its peripheral regions in the Yangtze River are taken into consideration. Three-gorge reservoir and its peripheral regions have been divided into 1330 assessment units, and the entire environmental conditions of each unit are obtained. Take one of these units as subject, the condition of reservoir-induced earthquake factors show in Table V. By using the proposed model, we can prediction and analyzes seismic risk in this unit.

Fuzzy mapping concerning the condition of reservoir-induced earthquake factor to the risk class has been defined by conditional probability, the first level comprehensive evaluation show below:

$$S_1 = W_1 \bullet R_1$$

$$= (0.2492 \quad 0.2240 \quad 0.3173 \quad 0.2096)$$

$$S_2 = W_2 \bullet R_2$$

$$= (0.1472 \quad 0.2133 \quad 0.3119 \quad 0.3275)$$

Fuzzy mapping concerning reservoir-induced earthquake sub-factor to the risk class has been defined by the results in the first level comprehensive evaluation, consider S_i (i=1, 2) as single-factor fuzzy judgment matrix, the second level fuzzy comprehensive evaluation show below:

$$S = W \bullet \begin{bmatrix} S_1 \\ S_2 \end{bmatrix}$$

$$= (0.6667 \quad 0.3333) \circ \begin{bmatrix} 0.2492 & 0.2240 & 0.3173 & 0.2096 \\ 0.1472 & 0.2133 & 0.3119 & 0.3275 \end{bmatrix}$$

$$= (0.2152 \quad 0.2204 \quad 0.3155 \quad 0.2489)$$

TABLE Ⅴ.
FACTOR CONDITION OF ASSESSMENT UNIT

Factors	R	K	F	C	W	A
Condition	4	3	4	4	2	3

According to maximum membership degree principle, 0.3155 as the biggest size of membership degree corresponds to danger class V_3, so V_3 is the evaluation result. Apply these processes, the prediction of other assessment units can be done.

ACKNOWLEDGMENT

The authors would like to appreciate the financial supports for this work from the Program for New Century Excellent Talent of Ministry of Education of China(Grant #NCET-07-0340), the Natural Science Foundation for Outstanding Scholarship of Hubei Province in China(Grant#2008CDB364), the National Key Technology R&D Program of China(Grant#2008BAC36B01), the National Basic Research Program of China(Grant#2007CB714107) and the National Natural Science Foundation of China(Grant #40672179).

REFERENCES

[1] H. K. Gupta and B. K. Rastogi., "Dams and Earthquakes," Amsterdam, Elsevier, 1976.

[2] R. P. Denlinger and C. G. Bufe, "Reservoir conditions related to induced seismicity at the Geysers Steam reservoir, nothern California," Seisnological Society of America, vol. 72 (4), pp. 1317-1327, 1982.

[3] D. W. Simpson, W. S. Leith and C. H. Scholz, "Two types of reservoir induced seismicity,"Bull. Swismol. Soc. Am., vol. 78, pp. 2025-2040, 1988.

[4] C. M. Keith, D. W. Simpson, O. V. Soboleva, "Induced seismicity and style of deformation at Nuek Reaervoir," Tadjik SSR. J Geophys Res, vol. 87, pp. 4609-4624, 1982.

[5] R. K. Chadha, H. Kuempel and M. Shekar, "Reservoir Triggered Seismicity (RTS) and well water level response in the Koyna-Warna region," Tectonophysics, India, vol. 456(1-2), pp. 94-102, 2008.

[6] I. Edelman, "An analytical interpretation of liquid injection induced microseismicity in porous reservoirs," Soil Dynamiics and Earthquake Engineering, vol. 26, pp.566-573, June-July 2006.

[7] G. Habibagahi, "Reservoir induced earthquakes analyzed via radial basis function networks," Engineering Geology, vol. 17(1), pp. 53-56, 1998.

[8] Y. M. Fahjan, O. S. Borekci and M. Erdik, “Earthquake-induced hydrodynamic pressures on a 3D rigid dam-reservoir system using DRBEM and a radiation matrix,” International Journal for Numerical Methods in Engineering, vol. 10, pp. 1511-1532, 2003.

[9] P. Li, Y. J. Li and M. Yang, “A Study of Reservoir-induced Earthquake in the Three Gorges Area,” Engineering Science, vol. 11, pp. 23-31, Jun. 2005.

[10] Q. F. Xia, “Assessment of resevoir induced earthquake,” The Chinese Journal of Geological Hazard and Control, vol. 11(2), pp. 39-44, Jun. 2000.

[11] Q. Y. Wang and Q. W. Zhang, “ Study on risk of induced earthquake in reservoir head region of Theree Gorges projection on Yangtze River,” Crustal Deformation and Earthquake, vol. 23(2), pp. 101-106, May 2003.

[12] Houqun Chen, Min Li and Yanhong Zhang, “Earthquake-risistant analysis on high arch dam and study on the dynamic stability of dam abutment,” Water Power, pp. 48-50, 2001.

[13] X. Sun and Z. F. Yang, “Characteristics of Lijiaxia reservoir earthquake and its seismicity riskness evaluation,” Journal of Safety and Environment, vol. 4(6), 50-53, Dec. 2004.

[14] Z. J. Ma, Q. W. Zhang and A. R. Li, “Quantitative prediction and assessment of induced seismicity risk for hydropower project based on Digital Valley,” Hydroelectric Energy, vol. 19(3), pp. 8-11, Sep. 2001.

[15] Q. W. Zhang, C. Wang and F. Li, “Quantitative Prediction and Assessment of Induced Seismicity Risk in Yangtze Three-Gorge Reservior Head Area and its Neighboring Area,” Water Resources and Power, vol. 23(4), pp. 21-25, Aug. 2005.

[16] T.L.Saaty, “Decision-making with the analytic hierarchy process,” International Journal of Services Sciences, vol. 1, pp. 83-98, Jan. 2008.

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp.105-107

Study on Civil Airport Safety Management System Based on Risk Management

Ming Cheng[1], and Wenjuan Liang[2]
[1] Research Institute of Civil Aviation Safety, Civil Aviation University of China, Tianjin, China
Email: figocm@163.com
[2] Research Institute of Civil Aviation Safety, Civil Aviation University of China, Tianjin, China
Email: lwjcm@163.com

***Abstract*—with the rapid growth of traffic and the increasing complexity of the operation, the civil airports face increasing safety pressure. Applying information management technology solves problems of airports safety management. It was established that an airport safety management system based on B/S structure. For airport-related departments, it provides a daily safety management tools and the emergency decision support tools. It has implemented the daily work of airport safety management and risk management integration. It has improved the accuracy, timeliness and validity of airport safety management.**

***Index Terms*—-safety management system, risk management, airport**

I. INTRODUCTION

With the increase of air traffic, in civil airports safety pressure gradually increased [1，2]. In particular, it is particularly important that the Standard of Safety Management for operations support.

Based on the current actual situation and operational requirements of Chinese civil airports, as well as introducing information technology and digital technology in the airports management, the safety management system has integrated the new procedures that included risk management, information management, emergency management and incidents management into the work routine of airport management. Using network information technology has achieved reasonable distribution and sharing of limited safety management resources.

The use of advanced tools of operational information collection, processing, analysis and sharing help to improve the accuracy, timeliness and validity of civil airport safety management [3].

Through the development of management and decision-making model, to provide timely, comprehensive, and accurate data to support at all levels of airport managers. Supporting enterprise management increases economic efficiency and social benefits.

II. SYSTEM DESIGN

A. System Framework

B/S structure is more convenient to data retrieval, statistical analysis and display mode, and it is more suitable conditions for the actual airports operation. It Include four subsystems. These mainly include: Data entry, query, modify, print, multi-condition search, graphical data statistical analysis, the exchange of data between functional modules, safety analysis, risk management, monitoring the progress of work items, safety trend analysis, safety alerts, etc. The framework of functions shows in Fig. 1.

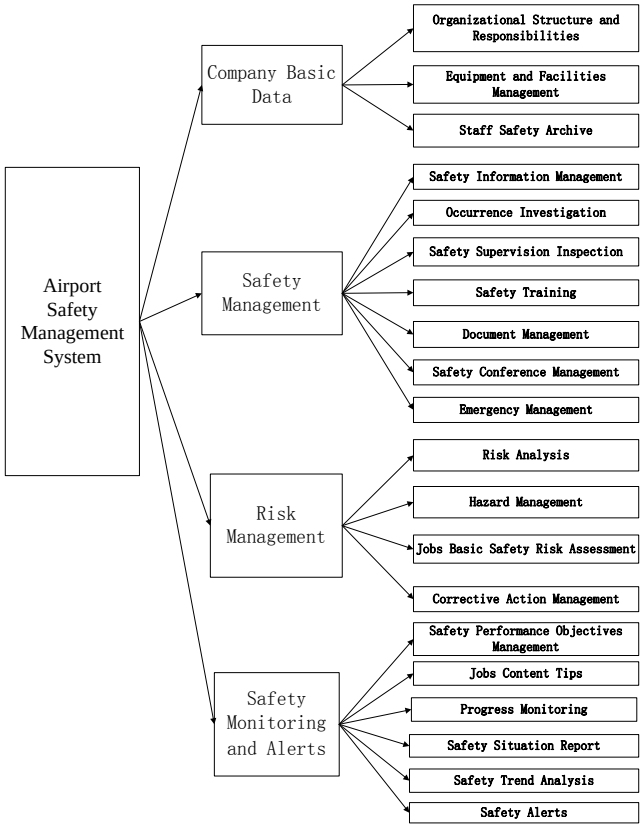


Figure1. System framework diagram.

B. System Function

Airport safety management system is composed of four subsystems, namely, company basic data subsystem, safety management subsystem, risk management subsystem and safety monitoring and alerts subsystem [4].

Project number: 08CAUC-E03
Corresponding author: Ming Cheng.

AP-PROC-CS-10CN006

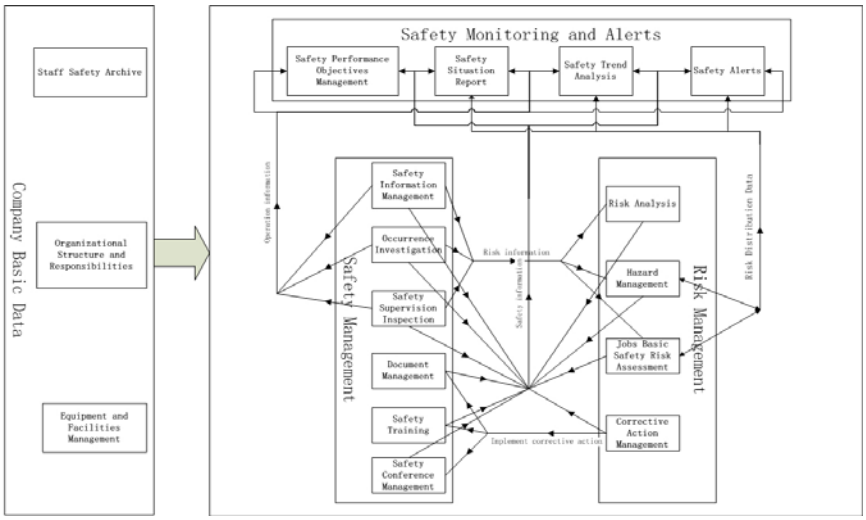


Figure2. System data flow diagram

The exchange of data between the various functional modules shows in Fig. 2..

C. Company Basic Data Subsystem

Company basic data subsystem is the basis of data modules. Its relevant data is referenced by most of the functional modules [5]. It includes:

(1) Staff Safety Archive: Personnel information includes basic information and safety information.

(2) Equipment and Facilities Management: equipment and facilities at all levels of basic information, as well as equipment and facilities use, maintenance, fault and current operational status, etc. Fault information includes the name of the fault, nature, type, time of occurrence, a detailed description, analysis and processing methods, as well as related safety information and incident investigations, safety supervision inspection and the associated risk analysis process.

(3) Organizational Structure and Responsibilities: Including those relating to safety management of all the organizations settings, the functional responsibilities, the organizational chart, the safety committee lists and safety management network diagram.

D. Safety Management Subsystem

Safety management subsystem implements the airport safety information management capabilities, easy access to all categories of personnel at all levels of safety-related inquiries and browsing information, to improve the efficiency of safety management, and risk management subsystem, and safety monitoring and alerts subsystem to provide the necessary information support. Key functions include:

(1) Safety Information Management: It manages internal event information flow, safety briefings, internal and external safety information circulars and other information. At the same time, it records related occurrence investigation, risk analysis, the index information.

(2) Safety Supervision Inspection：It covers all the major internal and external surveillance operations records and associated risk analysis, corrective measures.

(3) Emergency Management ： This module is responsible for maintaining the provisions of crisis management, emergency response plan, all departments and rescue units in the agreement of the responsibilities of the various events. In disasters, accidents and unexpected event occurs can be timely and speedy check, call the appropriate emergency response plans, and improve the operability and effectiveness of various types of emergency response plan.

E. Risk Management Subsystem

Risk management subsystem for the control of the airport operational risks, combining basic safety approaches of the risk analysis and the jobs basic safety risk assessment, to monitoring the regional airports, equipment, facilities and jobs. The mainly functions include risk identification, risk analysis, risk control and rectification measures and other content. Safety issues through in-depth and systematic analysis to identify the logical relationship between the reasons so as to formulate corrective action in a targeted manner, then tracking the effectiveness of the corrective action and ensuring the safety risks at an acceptable level. Key functions include:

(1) Hazard Management： It records basic information of hazards, risk level, probability, severity, risk index, the possible consequences and corresponding control measures, associated hazards related to unsafe events history.

(2) Jobs Basic Safety Risk Assessment：It established the safe operation of the airport jobs involved in risk analysis files, in order to identify potential safety problems. This module is the ability to detect the safety issues of the airport operation in time to take preventive actions to eliminate or reduce risks, and achieve safety sub-level management

(3) Risk Analysis: The module for the high frequency and serious consequences of the major safety issues, application of safety analysis techniques, such as risk management matrix. Through qualitative and quantitative analysis revealed the occurrence of the events, the development process, the direct cause of the event,

indirect causes, and the logical relationship between the reasons to identify the main cause of the event, and possible corrective actions.

F. Safety Monitoring and Alerts

Safety monitoring and alerts subsystem starting from the time of the longitudinal, comprehensive historical experience and real-time information, find an accident or occurrence and development of the law of hidden dangers, and accordingly propose measures to eliminate or reduce accidents. Key functions include:

(1) Safety Trend Analysis and Safety Situation Report: This module for statistical analysis the safety information of the airport operation, including statistical analysis of accident information, hazard information, and analysis results in graphic form of visual presentation, in order to identify the issues affecting the safety of the main reasons to develop appropriate corrective preventive measures.

(2) Jobs Content Tips and Progress Monitoring: It is aimed at forecasting the results of safety evaluation and safety indicators by statistical analysis, and presented in the form of statistical charts to facilitate timely detection of endangering the safety of the main issues and provide scientific basis for safety decision-making.

(3) Safety Alerts： Based on statistical analysis of historical data to establish the safety of individual airports (accidents) forecasting model, and then predict the future monthly (annual) safety (accident) changes in the law for possible problems in advance to take preventive measures.

III. SYSTEM OPERATING ENVIRONMENT

The user object of the Safety management system is the companies or enterprises of civil airport. Currently, there are software companies developed based on C/S structure (Client/Server, referred to as C/S) of the safety management system, the development in this field still has been progress slowly. As information gathering, functional planning and other factors, the application result is not satisfactory [6].

Based on flexibility, scalability and portability considerations of installation, debugging and using, this safety management system selected B/S structure (Browser/Server, referred to as B/S) (As shown in Fig. 3).

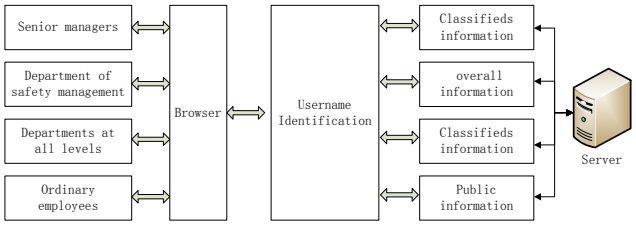


Figure 3. System schematic diagram.

The software operating system includes Windows2000, Windows XP, Windows NT and the Internet Explorer6.0 browser. The database platform to use Oracle8i, based on the use of JSP (Java Server Page) server-side dynamic web pages technology building, to ensure that the system to achieve cross-platform applications, flexibility, scalability and portability.

Used the tools of the system, the users of civil airport enterprise can manage and monitor the airport operational and safety situation of comprehensive, accurate, real-time and efficient.

The system established a management platform in the aspect of safety technology services, safety information management, safety oversight and management, accidents and incidents prevention to improve information technology in airports safety management.

IV. CONCLUSIONS

The system focuses on civil airport in the operational safety issues, integrating the four elements of the human, machine and environment, management, and using safety systems engineering technology, network technology, database technology ,and based on the relevant device information, run information and safety information, and implementing the airport operational safety issue and the realization of the daily work of airport safety management and risk management, and to improve the accuracy, fast and efficient of airport safety management.

The system for airport management personnel at all levels and departments to provide daily safety management tools and the necessary support tools to respond to unexpected events. It will effectively promote the popularization and application of Chinese civil aviation operational capacity and safety management improved.

ACKNOWLEDGMENT

Supported by the Scientific Research Foundation of Civil Aviation University of China under Grant No. 08CAUC-E03.

REFERENCES

[1] Wu Zongzhi, WEI Lijun. The base for enterprise's accident emergency system:major hazard identification and control. JOURNAL OF SAFETY SCIENCE AND TECHNOLOGY， 2006（6）：58～62

[2] Wei Qun， Xiong Zhang,Zhao Fang. CBSD method applied in airport information system. JOURNAL OF Computer Engineering, Vol.30 No.5 2004(3):195-197.

[3] D R Hunter. Retrospective and prospective validity of aircraft accident risk indicators. Human Factors,May2002.

[4] Xie xuyang, Deng yunfeng, Li Qun. Discuss on basic construction of emergency management information system. JOURNAL OF SAFETY SCIENCE AND TECHNOLOGY,2006（6）：27～30.

[5] Cheng Ming, Sun Ruishan, Liang wenjuan. Design of Civil Aviation Monitoring and Early-warning System Based on Data driven. 2007 Intelligent Transportation Systems Research and Practice：136～140.

[6] Robert A. Scopatz. Safety Data Analysis and Evaluate. TRANSPORTATION RESEARCH CIRCULAR. 2006 April: 7～9.

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 108-111

Analysis System for Correlative Fluctuation in Futures Market

Xingyun Dai[1], Weihui Dai[1], Hua Ye[1], and Genghui Dai[2]
[1] School of Management, Fudan University, Shanghai 200433, P.R.China
Email: {082025041, whdai, 0525005}@fudan.edu.cn
[2] School of Life Science, Zhongshan University, Guangzhou 510275, P.R.China
Email: tony2009@vip.163.com

***Abstract*—This paper studied the correlative fluctuation between two different futures markets, and developed an analysis system on .net platform to realize the prediction of price fluctuation. In this system, we explored the computing technology with mathematical programming language R, and integrated it with .net platform, thereof making the calculation precise and convenient. Test by this system has shown the accurate prediction of price fluctuation in the guided market..**

***Index Terms*—futures market, correlative fluctuation, analysis system**

I. Introduction

As a type of financial instrument, futures provide the buyers a way to purchase the underlying commodity for a specific price across time. There are two kinds of futures, commodity futures and financial futures.

With the development of global market and information network, international futures markets have become integrated. Price fluctuation of one futures market may cause related price fluctuation in another exchange. Because each futures market has its own business time, study on the correlation among asynchronous futures markets is helpful for fluctuation prediction in the following market.

Nowadays, a lot of studies have been concentrated in the price fluctuation pattern of a certain futures, the correlation between futures and spot price and the price of a certain futures in different markets.

Western scholars are inclined to find the reasons for futures price fluctuation from the view of macroeconomics. Williams and Orice M [1] analyzed the relationship between energy market and energy future, Christos Floros and Dimitrios V.Vougas [2], Razak M A and Bacha O I [3] studied the interrelations between the stock index futures and stock markets. Moreover, W Daal et al [4] focused on the characteristic of futures itself which could be the reason for price fluctuation. Some scholars have been also interested in forecasting futures price fluctuation. Michael D. Mackenzie and Heather Mitchell [5], Hyndman and Cody B [6] studies the methods of futures price fluctuation forecasting by using GARCH and Gauss model respectively.

Meanwhile, Chinese scholars are more likely to use time series analysis methods to study the relationship among the futures markets, or between futures and spot price. Wang J [7] used time series and causal analysis methods to study the guiding role of futures price in different futures markets. Liu [8] and Deng et al [9] analyzed the price fluctuation of different financial futures. Besides, Ye [10] and Zhang [11] focused on the correlation of the metals futures price fluctuation between LME and SHFE, which is also our concern.

Many of these researches have shown that there is high correlation between international and domestic futures price fluctuation. However, quantitative research and information system for analyzing and forecasting the futures price fluctuation have not been found in China. Our paper is aimed to analyze the price fluctuations of the same kind of futures in international and domestic futures markets, find out the correlation model and guiding role between them and finally develop an analysis system to realize the futures price fluctuation analysis and prediction in the guided market.

II. Correlation Model

A. Object

As we want to analyze the correlation between the futures price in two different futures markets, the price data is essential. Only the price data in the same contract period will be chosen since a certain kind of futures with different contract periods have different price in the same futures bourse. Furthermore, mismatched data will be excluded. For example, the bourses in America are closed on Christmas just like the Spring Festival in China. Finally, all data must be converted to the same currency unit.

B. Prediction Model

A sequence of methods is taken to handle the price data and finally get the price prediction of the following futures market.

After choosing futures price data as time series from different markets, stationary, co-integration and granger causality test are used to check the data validity. Stationary test is to confirm whether the data fulfills the conditions for co-integration test, which can tell there is real correlation among variables, or just spurious

This research was supported by National High-tech R & D Program (863 Program) of China (No.2008AA04Z127) and Shanghai Leading Academic Discipline Project (No.B210).
Corresponding author: Weihui Dai.

regression. Granger causality test is useful to determine whether one time series can be used to forecast another. Here we can use Granger causality test to find the guiding market in the markets.

Then use cross spectrum analysis to find the exactly relationship between futures price fluctuation, get the lag coefficient and make price prediction for the following market.

Basic idea of Spectrum analysis and cross spectrum analysis is that time series are considered as the superposition of sine and cosine waves. Fluctuation pattern of time series can be revealed by comparing variances and circles of the waves.

Given time series x and y, the autocovariance coefficients at lag k, cospectrum and quadrature spectrum (Q) of the cross spectrum are represented as follows.

$$C_{xy}(k) = (1/n-k)\sum_{t=0}^{n-k}\left(x_t - \bar{x}/s_x\right)\left(y_{t+k} - \bar{y}/s_y\right)$$

$$C_{yx}(k) = (1/n-k)\sum_{t=0}^{n-k}\left(x_{t+k} - \bar{x}/s_x\right)\left(y_t - \bar{y}/s_y\right)$$

$$P_j = \left(2/m\right)C_{xy}(0) + (1/2)\sum_{k=1}^{m}\lambda_k\left[C_{xy}(k) + C_{yx}(k)\right]\cos(\pi jk/m)$$

$$Q_j = \left(2/m\right)\sum_{k=1}^{m}\lambda_k\left[C_{xy}(k) - C_{yx}(k)\right]\sin(\pi jk/$$

$\bar{x}$ is mean of x, $\bar{y}$ is mean of y, $s_x = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(x_t - \bar{x}}$ and $s_y = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \bar{y}}$.

Amplitude (A), phase (H), coagulation spectrum (O), lag (L) of the cross spectrum are as follows.

$$A_j = \left(P_j^2 + Q_j^2\right)^{(1/2)}$$

$$H_j = \tan^{-1}\left(-\frac{Q_j}{P_j}\right)$$

$$O_j = \frac{A_j^2}{s_x s_y}$$

$$L(T_j) = H_j\left(T_j/2\pi\right)$$

T_j is cycle length of the cross spectrum.

In the paper of the analysis of correlation relationship among international futures markets, Ye H. [11] applied the above methods and chose copper futures data as the study object. The results showed that cycle for the copper futures price fluctuation of SHFE and LME with 9.82 days has the highest correlation. The leading time of this cycle is -1.0164, i.e. the price fluctuation of LME comes 1.1064 days before SHFE. The result that the copper futures price fluctuation of LME plays a guiding role in the two markets is consistent with the outcome of Granger causality test. Hcopper price fluctuation of LME can be used to forecast the fluctuation of SHFE the next day.

C. Arbitrage

Compared with stock market, futures market is a better place for arbitrage. Firstly, also the biggest advantage, futures market supports two-way trade. Secondly, there is low transaction cost in futures market. Thirdly, using of leverage is the fascination of futures investment. Lastly, “T+0” transaction model is used in futures transaction. High speed trading is convenient because a futures contract can be bought and sold in one day.

It is very possible for the speculators to realize arbitrage under these favorable conditions, along with the correlation among markets. Since we find the correlation between two futures markets, futures price fluctuation of the guided market can be predicted by using the price data of the guiding market. Hence, with price fluctuation data of the guiding market a unit of leading time before and the characteristic of futures market, arbitrage in the guide market is feasible.

III ANANLYSIS SYSTEM

A. Framework and Process

Analysis system is used to analysis the correlation between any two price fluctuation time series and forecast the price trend of the guided market, based on the correlation model above.

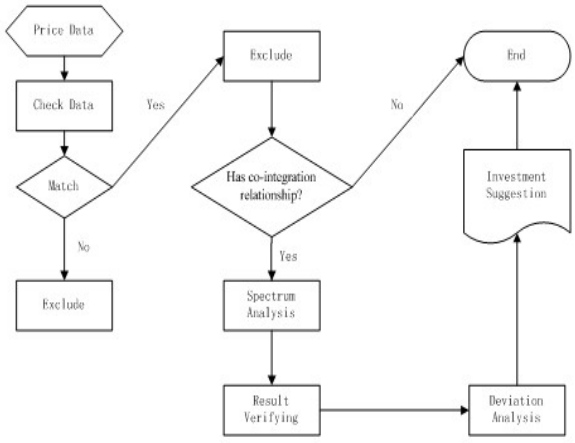


Figure 1. Flow chart of analysis system

Fig. 1 shows the processes of the system. Firstly, check the price data from the markets the user concerns and then exclude the mismatched data. Secondly, use ADF and co-integrated tests to determine whether there is co-integration relationship between the price series. If there is, use spectrum analysis to find the leading time of the guiding market and Granger causality test to verify the result. Finally, provide investment suggestion to the user, according to the deviation analysis after verifying.

B. Modules

This system is composed of seven modules: data importing, sample selection, database management, knowledge base management, model base management,

time series analysis and suggestion output. The whole structure is showed in Fig. 2.

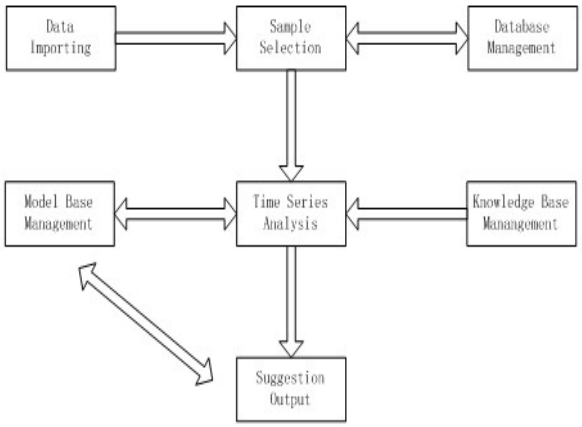


Figure 2. Function modules of analysis system

Data importing imports the external price data into analysis system and converts them to the proper format. Sample selection checks the time series and excludes mismatched data. Database management deals with data inserting, updating, deleting, querying and backing up. Knowledge base management and model base management are used to manage the knowledge and model constructing, storing and updating respectively. Time series analysis uses econometrics methods to find the correlation between time series and analyzes the deviation between the prediction and actual results. Suggestion output gives analysis results and investment suggestion to the user.

By this system, user can make investment decision more scientifically and conveniently.

IV. Computing Technology and System Implemention

A. Ecnometrics in R

As an implementation of the object-oriented mathematical programming language S, R is free software developed by statisticians around the world. Because of the flexibility of R, it is easy to extend and always at the forefront of statistical knowledge. Moreover, code written for R can be run on any computational platforms such as .net, J2EE etc.

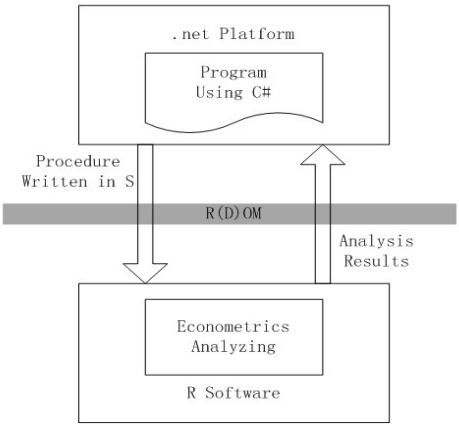


Figure 3. Interaction between .net and R

In our system, we can use R to do the time series analysis work on the .net platform. Using S to program at client and conveying it to software R for computing, at last the result will be return to C#. R(D)COM interface which can be used for C# to call R enable one to use R as a computational engine that also can render graphics. . Fig. 3 shows the interaction between .net and R.

1) Libraries

First of all, we should construct the R(D)OM by installing package rscproxy. For using econometrics analysis we need, packages urca, uroot and tseries are necessary. RODBC is package for communicating with DBMSs.

Following COM references should be added to the C#.net project, as well as the namespaces to the class.

```
using RServerManager;
using StatConnectorCommonLib;
using STATCONNECTORSRVLib;
```

After initialize the connector class of R, use EvaluateNoReturn() to casting and extracting data from the R environment. To add libraries, use method as following.

```
StatConnector sc = new StatConnectorClass();
sc.Init("R");
sc.EvaluateNoReturn("library(urca)");
sc.EvaluateNoReturn("library(RODBC)");
```

2) Data Management

Both text file and database work can be imported into R, using the following methods.

```
sc.EvaluateNoReturn("m1=read.table(file.choose())");//text files
sc.EvaluateNoReturn("library(RODBC)"); //database w/r
sc.EvaluateNoReturn("channel <-odbcConnect(\'d\', uid=\'ripley\')");
sc.EvaluateNoReturn("sqlFetch(channel, \'t\', rownames = \'state\')");
sc.EvaluateNoReturn("data(t)");
string r = "sqlSave(channel, t, rownames = \'s\', addPK = TRUE)";
sc.EvaluateNoReturn(r);
```

3) Econometrics Analysis

It is very convenient to use R for econometrics analyzing. The following codes are used to do ADF test with lag 1.

```
string rcode = "adft.out = unitroot(uscn.spot, trend=\'c\',"
+"statistic=\'t\',method=\'adf\', lags=1)";
sc.EvaluateNoReturn(rcode);
```

As other examples, the major codes are “coint.rc = coint(uscn.ts,trend="c",lags=4)” is used for co-integration test and “spectrum(SOI, span = sqrt(2 * length(SOI)))” is used for spectrum analysis.

4) Graphics

Graphics Tool is a very important component of R environment. The most commonly used graphics function is plot(). For example, if x and y are vectors, plot(xy) will produce a scatter.

By using graphics tool, almost all kinds of statistics graphics can be displayed.

B. Evaluation

We have evaluated this system by a series of tests. For example, the copper futures price data of SHFE and LME from 1 April to 20 May in the year 2008 are chosen to predict the price trend of 21 May in the guided market. Fig. 4 shows the prediction of copper price trend in guided market LME on 21 May, 2008. The leading time turns out to be 1 hour.

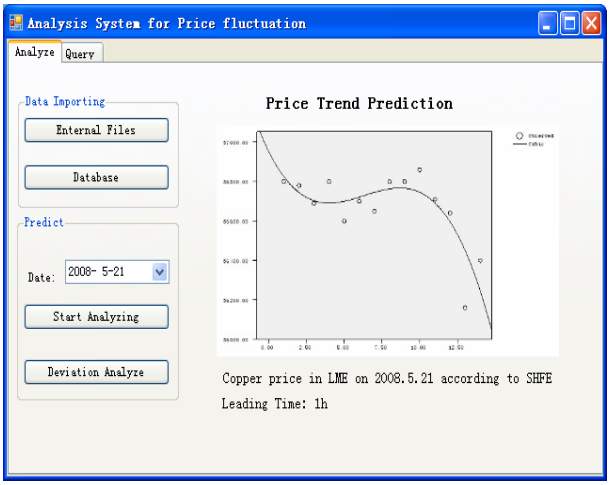


Figure 4. Prediction of Copper Price Trend on 21 May, 2008

To test the effectiveness of the prediction, we compare the observed data and the prediction data as Fig. 5 shows.

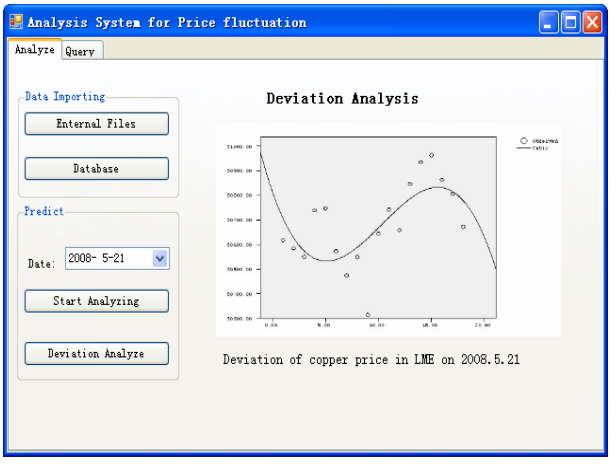


Figure 5. Deviation of Copper Price in LME on 21 May, 2008

The prediction of price trend and peak position is essentially consistent with the observed data.

By simple operation, this analysis system has accurately found the correlation between two price fluctuation data and well predicted the futures price trend in the guided market.

V. Conclusion

Our research applies spectrum analysis to international futures price fluctuation prediction, offering a new thinking to quantitative study in this field. The method is not only used to predict price fluctuation of the futures mentioned in this paper but also other futures and provides theoretical support to other price fluctuation researches. Meanwhile, we realize the price fluctuation correlation model by using information system and introduce R for statistics calculation on programming platform, making the prediction so simple and convenient.

Mechanism of futures price fluctuation prediction and arbitrage is still at the starting stage. More analysis methods can be introduced into the price fluctuation study because of the imperfect of spectrum analysis used in our research. Higher precision of price data which will make the prediction more accurate and more study on the analysis system for arbitrage program or even transaction robot are all further research work.

Acknowledgment

This research was supported by National High-tech R & D Program (863 Program) of China (No.2008AA04Z127) and Shanghai Leading Academic Discipline Project (No.B210).

References

[1] O. M. Williams, “Energy derivatives: Preliminary views on energy derivatives trading and CFTC oversight,” GAO *Reports*, vol.7, pp.1-20, 2007.

[2] C. Floros, D. V. Vougas, “The efficiency of Greek stock index futures market,” *Managerial Finance*, vol. 34(7), pp.498-519, 2008.

[3] M. A. Razak, Bacha O I, “Pricing efficiency of the 3-month KLIBOR futures contracts: an empirical analysis,” *Applied Financial Economics*, vol. 19(5), pp.445-463, 2009.

[4] E. Daal, J. Farhat, P. P. Wei, “Does futures exhibit maturity effect? New evidence from an extensive set of US and foreign futures contracts,” *Review of Financial Economics*, vol. 15(2), pp.113-128, 2006.

[5] M. D. Mackenzie, H. Mitchell, “Power ARCH modeling of commodity futures data on the London Metal Exchange,” *The European Journal of Finance*, vol. 7(1), pp.22-38, 2007.

[6] C. B. Hyndman, “Gaussian factor models—futures and forward prices,” *IMA Journal of Management Mathematics*, vol. 18(4), pp.353-369, 2007.

[7] J. Wang, “Research on the international linkages of sino-US soybean oil futures & spot markets and dynamic forecast,” *Journal of Northwest A&F University (Social Science Edition)*, pp.31-35, 2008.

[8] X. Deng, B. Ma, “The study of price return and spillover effects between interest rate futures market and spot market in Taiwan of China,” *Taiwan Research Quarterly*, pp.63-69, 2009.

[9] W. Ye, “Research on the linkages of price fluctuation in international zinc futures market — Set SHFE and LME as an example,” *Times Finance*, vol.(7), pp.30-32, 2008.

[10] T. Zhang, “A study of the correlation between metal futures in LME and SHFE,” *Times Finance*, vol.(5), pp.40-42, 2007.

[11] H. Ye, *Correlation Analysis and Prediction System for Price Fluctuation of International Futures*, Shanghai: Fudan University, 2009.

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 112-115

Extraction Algorithms for Using a Regular Grid DEMs

Kun Hou[1,2], Jigui Sun[1], Wei Yang[2,3], Tieli Sun[2], Ying Wang[3], and Sijun Ma[4]
[1] Jilin University/College of Computer Science and Technology, Changchun, China
Email: houk431@nenu.edu.cn
[2] Northeast Normal University/School of Computer Science and Information Technology, Changchun, China
[3]Changchun Normal University/College of Computer Science and Technology, Changchun, China
[4]Jinggangshan University/Art School, Jinggangshan, China
Email: yangw474@nenu.edu.cn, viviyw@163.com

***Abstract*—Algorithms are presented which use a digital elevation model (DEM) to extract topographic patterns. The algorithms have been applied to extracting natural topographic automatically from the regular grid DEM of a testing area. The experimental results have proved that the automatic extracting algorithms are effective.**

***Index Terms*—Algorithm, GIS, DEMs, topographic pattern**

I. Introduction

Because it is a fundamental problem in digital terrain analysis, the extraction of extracting drainage networks plays an important role in many applications of GIS, such as hydrologic analysis, mineral deposition, land erosion, pollution diffusion analysis, etc. [1-4]. The basic and important problem of extracting drainage networks is to determine flow direction of every cell in a raster DEM. The most popular and commonly used algorithm is known as D8 [5-8], but the continuity of drainages does not extend across pits and flats.

In this paper, we present a method to handle pits and flats with heuristic information. Heuristic information is often applied to finding the shortest path. While searching for the outlet, the proposed method not only checks the eight adjacent cells of pit or flat, but also takes the general trend of the DEM into considerations.

II. Overview Of The Methods Used For Drainage Delineation

A. Conventional Methods and Iimprovement

Numerous methods have been proposed for automated drainage recognition in grid DEMs. The basic D8 (deterministic eight-neighbor method) algorithm is probably the most popular method. However, it has a number of deficiencies. In this method, the flow direction of each cell in a raster DEM is determined by the comparison between the cell's elevation and its eight adjacent neighbors, the maximum drop direction was identified as the flow direction. The generation of parallel flow lines in flat areas that restricts the formation of concentrated channel flow has been identified as a limitation of this method. In particular, pits and flats, both real and spurious, in raster elevation data present challenges for the automated derivation of fully connected drainage patterns.

Pits and flats are areas where the cell's flow direction cannot be determined with reference to its eight neighbouring cells. A pit is defined as a point none of whose neighbors have lower elevations, flats are areas of level terrain (slope =0). Pits and flats can be naturally occurring but are generally viewed as spurious features that result from elevation data capture errors [9, 10] or errors introduced during DEM generation [11]. Pits and flats can be found in all DEMs but are particularly pronounced in low relief areas or in the case of flats.

For ensure each cell in a raster DEM has flow direction, many methods are proposed to assign flow direction of either pits or flats. One of most common methods is proposed by Jenson and Domingue (1988). The method works in two stages. In the first stages, DEM filling, all pits are identified and filled to their pour points. The pits are thus turned into flats. For flow tracing along both natural and resulting flat areas, iterative flow direction assignment through flat areas, the second stage was used. Cells in the flat areas are assigned flow directions that point to one of its neighboring cells having a flow direction not pointing back to the selected cell, until the entire flat area has been consumed.

Outlet breaching modified DEM filling by assuming that the pits are created by an elevation overestimation at the obstruction [12]. However, DEM filling assumes that pits are created by an underestimation of elevation values within the pits. Outlet breaching involves identifying the breach points for each pit and lowering the elevation values in these areas, subject to user-specified tolerances. An advantage of outlet breaching is that it typically involves making fewer alterations to a DEM than filling and much of the original interpolated elevation values, and therefore flow directions within a pit, are preserved.

Imposing relief across flats is a method for flat areas. Flats in a raster DEM are not exactly level in nature because DEM can't represent elevation difference less than vertical resolution. The method introduces a relief algorithm to impose a finite gradient over the flats [13]. This method requires less computer memory compared to the algorithm proposed by Jenson and Domingue (1988). First, this algorithm searches the perimeter of a flat for valid downslope cells (exit points). These exit points are

used as a source and neighboring cells within the flat are assigned a small fixed elevation value slightly higher than the outlet cell(s). The process is repeated iteratively until the entire flat has been consumed, thereby creating a gradient across flats based on Euclidian distance from the outlets. In more advanced forms of relief imposition, a secondary shallow gradient is imposed to force flow from the higher terrain surrounding a flat. This second gradient ensures that flow within the flat is consistent with the topography surrounding the flat surface and avoids parallel flow paths across flats that can be witnessed in the methods proposed by Jenson and Domingue (1988) and Martz and Garbrecht (1993).

The parallel drainage lines were found to be unrealistic and a new method was developed for flow tracing, in which the pits were handled in a more realistic way [14]. Spurious pits and real pits were identified and treated separately. The method involves defining a main flow path through the flat to an outlet, and directing the flow paths for nearby points in the flat towards the main path. However, because the main flow path is defined as a straight line, it could cut through an area of higher elevation [15-20].

B. Heuristic Search

The primary search strategy of the methods described above is uninformed search (also called blind search). The term means that they have no additional information about states beyond that provided in the problem definition. All they can do is generate successors and distinguish a goal state from a nongoal state. The uninformed search strategies can find solutions to problems by systematically generating new states and testing them against the goal. Unfortunately, these strategies are incredibly inefficient in most cases. Informed search (or heuristic search) strategies which use problem-specific knowledge beyond the definition of the problem itself can find solutions more efficiently than the uninformed search strategies. Information about the state space can prevent algorithms from blundering about in the dark [21].

In heuristic search, a node is selected for expansion based on a heuristic function, $f(n)$. Heuristic functions are the most common form in which additional knowledge of the problem is imparted to the informed search algorithm. Traditionally, the node with the most suitable evaluation is selected for expansion, because the evaluation measures cost to the goal.

Heuristic function evaluates the importance of the node. It evaluates nodes by combining $g(n)$, the cost to reach the current node and $h(n)$, the cost to get from the current node to the goal node.

Since $g(n)$ gives the path cost from the start node to current state, and $h(n)$ is the estimated cost of the cheapest path from current node to the goal node, we have

$f(n)$ = estimated cost of the cheapest solution through the current node

Thus, if we are trying to find the cheapest solution, a reasonable thing to try first is the node with the most suitable value of $g(n)+h(n)$. It turns out that this strategy is more than just reasonable: provided that the heuristic function $f(n)$ satisfies certain conditions, the search is both complete and optimal.

III. PROPOSED METHOD

The proposed method is based on heuristic search. The main purpose of the method is to handle pits and flats with heuristic information in one procedure.

A. Calculating the Incipient Flow Direction Using the Basic Algorithm

One of the keys to deriving hydrologic characteristics about a surface is the ability to determine the direction of flow from every cell in the raster DEMs. A 3-by-3 moving window is used over a DEM to locate the flow direction for each cell. The elevation difference (ΔE) from the centre cell of the window to one of its eight neighbours is calculated as:

$$\Delta E = \frac{E_0 - E_i}{\phi(i)} \tag{1}$$

Where $\phi(i)$ =1 for 2, 4, 6, and 8, (E, S, W, and N neighbours) and $\phi(i)$ = $\sqrt{2}$ for 1, 3, 5, and 7 (NE, SE, SW, and NW neighbours). If ΔE is less than zero, a zero flow direction is assigned to indicate that the flow direction is undefined. If ΔE is greater than zero and occurs at one neighbour, the flow direction is assigned to that neighbour. If ΔE is greater than zero and occurs at more than one neighbour, the flow direction is assigned to the maximum ΔE, the lowest neighbour cell. If the flow direction is still undefined, a zero flow direction is assigned to indicate that the flow direction is undefined. The cell with an undefined flow direction is then processed by the algorithm proposed by this article.

B Finding the Optimal Outlet of Depresssions and Flat Areas Using the Proposed Method

After calculating the incipient flow direction, the cells without flow direction (pit and flat cells) need advanced processing with the proposed method. The proposed method is based on heuristic information. The method would require three data structure, closed list, open list and an array. The closed list stores every checked node, the open list keeps the unchecked nodes, and the array stores the cells without flow direction.

The proposed method consists of a sequence of procedures, which are illustrated in Fig. 1. The procedures operate on the elevation matrix, flow direction matrix and flow accumulation matrix.

The proposed method uses the heuristic information value to evaluate the node .The heuristic information consists of two elements, actual cost and estimated cost, according to:

$$f(i) = g(i) + h(i) \tag{2}$$

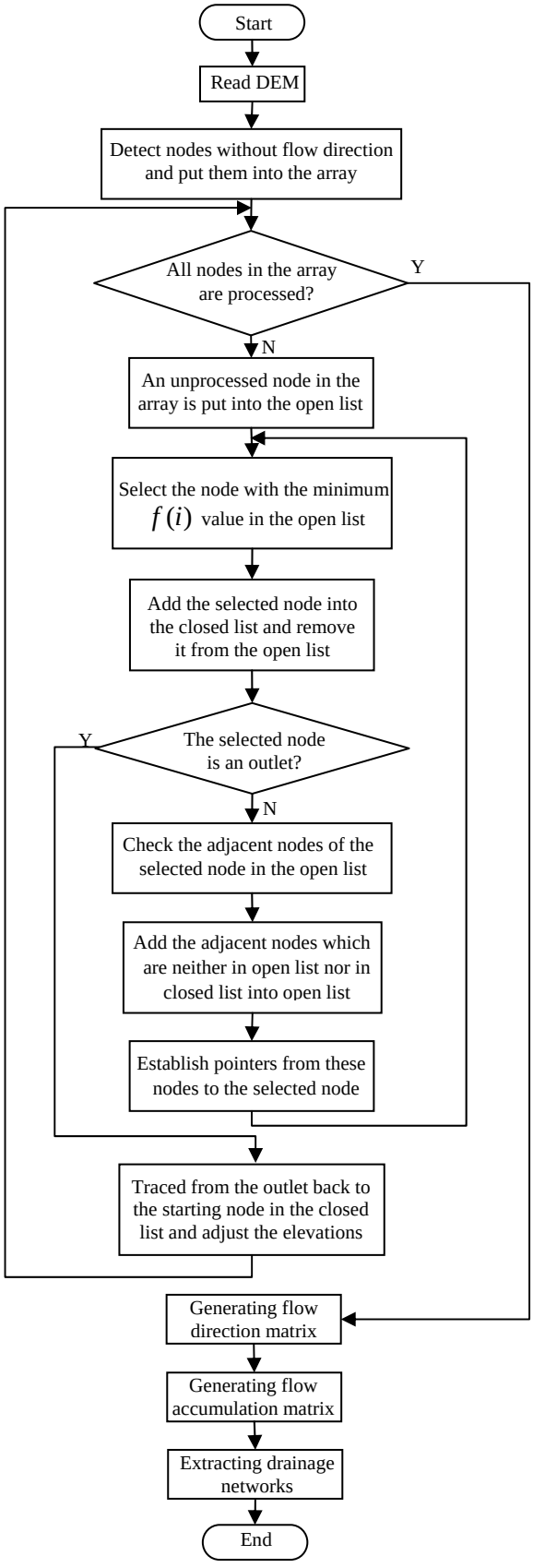


Figure 1. Schematic diagram showing the proposed method.

$$g(i) = E_i - E_s \tag{3}$$

$$h(i) = \frac{\sum_{k \in S_i} E_k}{\|S_i\|} \tag{4}$$

Where $f(i)$ is heuristic information value of node i, $g(i)$ is the difference from the node i to the starting node, $h(i)$ is estimated cost, the arithmetic average elevation of the cells in n-by-n (n=3, 5, 7, …) window centered the node i, we choose 5 in this article, E_s is the elevation value of the starting node s, E_i is the elevation value the node i, E_k is the elevation value the node k, S_i is a set which contains all the nodes in n-by-n window centered the node i, k is a element of set S_i, and $\|S_i\|$ is the number of nodes in the set S_i.

IV Experiment And Results

To examine the suitability and performance of the proposed algorithm, we implement it in Pascal and make experiments on actual DEM data. For presentation purpose all raster maps are converted into vector format.

The study area is "1-degree" DEM which was provided by the US Geological Survey [22]. The 1-degree DEM is 1:250,000-scale Digital Elevation Models and covers an area of 1° in longitude by 1°latitude. The basic elevation model is produced by the Defense Mapping Agency (DMA) using cartographic and photographic sources. The 1-degree DEM consists of a regular array of elevations referenced horizontally on the geographic coordinate (latitude/longitude) system of the World Geodetic System 1984 Datum. Elevations are in meters relative to mean sea level. The DEM contains 601 rows by 1201 columns and spacing of the elevations along and between each profile is 3 arc-seconds.

Fig. 2 shows the drainage networks generated using ArcGIS 9.2. The major drainage could be delineated satisfactorily but the spurious parallel flow channels can be seen and the detailed information is lost. It is obvious in the selected rectangle and circle region in Fig. 2. It is due to the error in selecting the flow direction of pits and flats when more than one possible direction exists.

Fig. 3 shows the proposed method for extracting drainage network from the same DEM data. It can be seen from Fig. 3, continuous drainage network and major drainage are delineated.

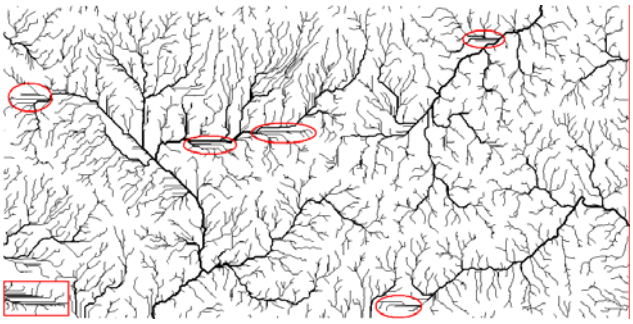


Figure 2. Drainage networks generated using ArcGIS 9.2.

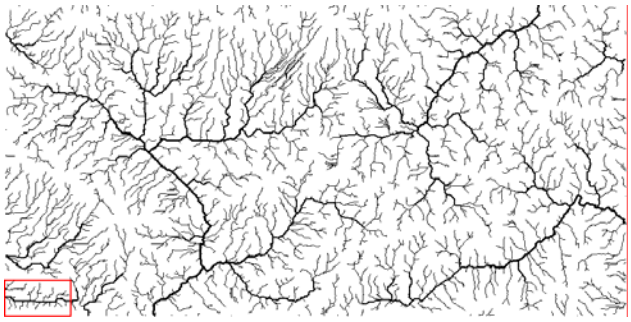


Figure 3. Drainage networks generated using the proposed method.

The detailed information can be retrieved and the spurious parallel flow lines can not be found. In order to avoid these cases, the proposed algorithm takes heuristic information for hydrologic analysis and tries to capture the topography and assign the optimal flow direction to the spurious pits and flats.

V Conclusion And Discussion

This paper presents a new method to assign the direction of the flow through pits and flats by searching for minimum value of heuristic function. The proposed algorithm has been tested with actual DEM data, and the experimental evidence suggests that the proposed algorithm can extract drainage networks better and fully than the basic method. To compensate inadequate searching information of other methods, the proposed method finds the outlet with heuristic information. Furthermore, the proposed method can handle pits and flats effectively in one procedure.

Acknowledgment

The work is supported by Science Foundation for Young Teachers of Northeast Normal University.

References

[1] D. M. Wolock, and G. J. McCabe, “Comparison of single and multi-flow direction algorithms for computing topographic parameters in TOPMODEL,” *Water Resources Research*, 31 (5), pp. 1315–1324, 1995.

[2] T. G. Freeman, “Calculating catchment area with divergent flow based on a regular grid,” *Computers & Geosciences*, 17 (3), pp. 413–422, 1991.

[3] Z. Li, Q. Zhu, and C. Gold, *Digital Terrain Modeling: Principles and Methodology*, Boca Raton: CRC Press, 2004, pp. 323.

[4] J. F. O’Callaghan, and D. M. Mark, “The extraction of drainage networks from digital elevation data,” *Computer Vision, Graphics, and Image Processing*, 28, pp. 323-344, 1984.

[5] S. K. Jenson, and J. O Domingue, “Extracting topographic structures from digital elevation data from geographic information system analysis,” *Photogrammetric Engineering and Remote Sensing*, 54 (11), pp. 1593-1600, 1988.

[6] F. Kenny, B. Matthews, and K. Todd, “Routing overland flow through sinks and flats in interpolated raster terrain surfaces,” *Computer & Geosciences*, 34, pp. 1417-1430, 2008.

[7] L. W. Martz, J. Garbrecht, “The treatment of flat areas and depressions in automated drainage analysis of raster digital elevation models,” *Hydrological Processes*, 12, pp. 843-855, 1998.

[8] L. W. Martz, and J. Garbrecht, “Automated extraction of drainage network and watershed data from digital elevation models,” *Water Resources Bulletin*, 29 (6), pp. 901–908, 1993.

[9] W. Zeng, and R. L. Church, “Finding shortest paths on real networks: the case for A*,” *International Journal of Geographical Information Science*, 23 (4), pp. 531-543, 2009.

[10] J. Lindsay, and I. F. Creed, “Sensitivity of digital landscapes to artifact depressions in remotely-sensed DEM,” *Photogrammetric Engineering and Remote Sensing*, 71 (9), pp. 1029–1036, 2005.

[11] L. E. Band, “Topographic partition of watersheds with digital elevation models,” *Water Resources Research*, 22 (1), pp. 15–24, 1986.

[12] L. W. Martz, and J. Garbrecht, “An outlet breaching algorithm for the treatment of closed depressions in a raster DEM,” *Computers & Geosciences*, 25, 835–844, 1999.

[13] L. W. Martz, J. Garbrecht, “Numerical definition of drainage network and the sub-catchment areas from digital elevation models,” *Computers and Geosciences*, 18 (6), pp. 747–761, 1992.

[14] A. Tribe, “Automated recognition of valley lines and drainage networks from grid digital elevation models: a review and a new method,” *Journal of Hydrology*, 139, pp. 263–293, 1992.

[15] J. Richard, “Algorithms for using a DEM for mapping catchment areas of stream sediment samples,” *Computers & Geosciences*, 28 (9), pp. 1051-1060, 2002.

[16] M. F. Hutchinson, “A new procedure for gridding elevation and stream line data with automatic removal of spurious pits,” *Journal of Hydrology*, 106, pp. 211–232, 1989.

[17] J. Fairfield, and P. Leymarie, “Drainage networks from grid digital elevation models,” *Water Resources Research*, 27(5), pp.709-717, 1991.

[18] D. G. Tarboton, “A new method for the determination of flow directions and contributing areas in grid digital elevation models,” *Water Resources Research*, 33, pp. 309-319, 1997.

[19] R. Jana, T. V. Reshmidevi, P. S. Arun, and T. I. Eldho, “An enhanced technique in construction of the discrete drainage network from low-resolution spatial database,” *Computer & Geosciences*, 33, 717-727, 2007.

[20] CatchmentSIM,http://www.toolkit.net.au/catchsim/Overview/PFSMethod/PFSMethod.htm, last access: 13 March 2007, 2007.

[21] S. J. Russel, and P.Norvig, *Artificial Intelligence – A Modern Approach*, Pearson Education, Inc., USA: New Jersey, 2003.

[22] USGS, http://edcftp.cr.usgs.gov/pub/data/DEM/250, last access: 13 March 2007.

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 116-119

Study on the Failure Identifiability of Disaster Detection System

Xiaoyu Wang[1], and Kaiquan Wang[2]
[1]Changzhou Iinstitute of Litht Industry Technology,Changzhou 213164, China
Email: wxy730@163.com
[2]Jiangsu Polytechnic University,Changzhou 213164, China
Email: wkq@jpu.edu.cn

***Abstract*—There is a serious potential risk in the petrochemical production process, the Disaster Detection System(DDS) is of key equipment to monitor the production status, prevent and control the potential risk. However, the Initiating failure and Inhibiting failure would occur in the period of the Disaster Detection System(DDS) working. In order to identify the two failures by DDS itself, the paper proposed a mathematic method for finding the fault test codes based on the theory of Bayle derivation, and verified it on the FCC (Fluidized Catalytic Cracking Unit) of a petrochemical company. The results evidence that according to the property of the Boolean difference, the codes are available to fined DDS failures; in order to identification of DDS failure it is necessary to properly design the combinational logic unit for ensure it has the fault test codes; if the DDS failure that had been judged to be identification, the fault test codes should always be input to DDS timely to implement self fault test so that its failure may be able to find. Through such means the DDS functions should be achieved correctly.**

***Index Terms*—Disaster Detection System; Initiating failure; Inhibiting failure; Identification; Fluidized Catalytic Cracking Unit**

I. INTRODUCTION

There is a serious potential risk in the petrochemical production process, the Disaster Detection System(DDS) is of key equipment to monitor the production status, prevent and control the potential risk. However, the initiating failure and inhibiting failure would occur in the period of DDS' working. The former because of "undetected", mistakenly believe that a critical condition is safety so that neither recognize the dangers nor prevention the accidents in time; the latter because of "false detection", in contrast consider the safety state as dangerous, forced interruption of production, caused the cut-off losses. [1] In order to achieve the overall security of the production process, DDS must have self-testing capabilities to identify the two types of fault.

The core component of DDS in petrochemical production (such as Disrubution Control System-DCS, Emergency Shutdown System-ESD) is a combination logic unit, so the problem must be selution in study of DDS failure identifiability is that whether the information carrying two types of fault can be correctly detected in the output of the unit. Specifically, for a tertain combination logic unit, can we found out some representatives test code while these code be input to the unit then the failure signal may be detected in the outpu. If such a test code exists, the DDS failures can be identification. [2] [3]

The approach of seeking the relationship of code between input and output in combinational logic unit was proposed firstly by prf. Akers. [4] Based on approach and Boolean difference operation, this paper will study the identificability of DDS' failure.

II. PRINCIPLE OF DDS FAULT IDENTIFICATION [5]

A. The properties of Boolean difference

Assume that there is a combinational logic unit, and its initial input Boolean variable is recorded as $x_1, x_2, ...x_n$, output Boolean function denoted by $f(x_1, x_2, ...x_n)$, then

$$\frac{df(X)}{dx_i} = f(x_1, x_2, ..x_{i-1}, x_i, x_{i+1}, ..x_n) \oplus f(x_1, x_2, ..x_{i-1}, \overline{x_i}, x_{i+1}, ..x_n) \quad (1)$$

known as Boolean difference $f(X)$ to x_i.

Boolean difference has two important properties.

Property 1: If $f(x_1, x_2, ..x_{i-1}, 0, x_{i+1}, ..., x_n) = f(x_1, x_2, ..x_{i-1}, 1, x_{i+1}, ..., x_n)$ then

$$\frac{df(X)}{dx_i} = 0$$

There is no effect on the output while x_i change, so that the failure at x_i is unpredictable.

Property 2: If $f(x_1, x_2, ..x_{i-1}, 0, x_{i+1}, ..., x_n) \neq f(x_1, x_2, ..x_{i-1}, 1, x_{i+1}, ..., x_n)$ then

$$\frac{df(X)}{dx_i} = 1$$

There is effect on the output while x_i change, so that the failure at x_i is predictable.

B. Calculation of failure test code

1) Logical unit failure mode

Combinational logic unit failures in two ways:

Failure Mode 1 - Fixed 1 fault $(s-a-1)$. While the logic unit has AND gate and Boolean algebraic expression is

$$Z = x_1 x_2 x_3 \tag{2}$$

Then any one of element (such as x_1) take 1, Z will not change, that is, expression (2) can not find the failure mode 1 of x_1, so that x_1 is Fixed 1 fault $(s-a-1)$ of Boolean algebraic expression Z. Similarly, x_2, x_3 is also Fixed 1 fault $(s-a-1)$.

Failure Mode 2 - Fixed 0 fault $(s-a-0)$. While the logic unit has OR gate and Boolean algebraic expression is

$$Z = x_1 + x_2 + x_3 \tag{3}$$

Then any one of element (such as x_1) take 0, Z will not change, that is, expression (3) can not find the failure mode 2 of x_1, so that x_1 is Fixed 0 fault $(s-a-0)$ of Boolean algebraic expression Z. Similarly, x_2, x_3 is also Fixed 0 fault $(s-a-0)$ of Boolean algebraic expression Z.

2) Calculation of single failure test code

If $T(a_1, a_2, ..., a_n)$ is the test code of Fixed 0 fault in measured channel $X(x_1, x_2, ..., x_n)$, then the formula established:

$$X(x_1, x_2, ..., x_n) \times \frac{df(X)}{dx_i}\Big|_{T(a_1, a_2, ..., a_n)} = 1 \tag{4}$$

Where $a_i = 0$ or 1.

If $T(a_1, a_2, ..., a_n)$ is the test code of Fixed 1 fault in measured channel $X(x_1, x_2, ..., x_n)$, then the formula established:

$$\overline{X(x_1, x_2, ..., x_n)} \times \frac{df(X)}{dx_i}\Big|_{T(a_1, a_2, ..., a_n)} = 1 \tag{5}$$

Where $a_i = 0$ or 1.

Application the Property 1 and Property 2 of Boolean difference, it would be appraised whether the Fixed 1 fault $(s-a-1)$ and Fixed 0 fault $(s-a-0)$ may be measured on the channel $X(x_1, x_2, ..., x_n)$.

Applications formula (4) and (5) , the single failure test code of the Fixed 1 fault $(s-a-1)$ and Fixed 0 fault $(s-a-0)$ on the unit $X(x_1, x_2, ..., x_n)$ would be calculated. For example, added the single test code $T(a_i)$ to the input of the measured unit, if the unit output $Z = 1$, then there is no failure of $(s-a-1)$ in $X(x_1, x_2, ..., x_n)$, otherwise there is.

Similarly, the multi-fault test code $T(a_1, a_2, ..., a_n)$ can be calculated, not repeat them due to space limitations.

III. EXAMPLE

There is a petrochemical company using ESD as DDS in its FCCU-II sets. [6][7]The main logic diagram of the combination logic unit shown in Fig.1.[8] Let's examine the identification of the DDS' failure.

A.Failure identifiability of combination logic unit

There are three kinds of basic logic unit in fig.1 (Fig. 2).[9][10]

For fig. 2 (a), because of $f(X) = x_1 x_2 + x_1 x_3 + x_2 x_3$, then

$$\frac{df(X)}{dx_1} = (x_1 x_2 + x_1 x_3 + x_2 x_3) \oplus (\overline{x_1} x_2 + \overline{x_1} x_3 + x_2 x_3)$$
$$= \overline{x_2} + \overline{x_3} + \overline{x_2 x_3}$$

So that the fault test expressions:

$$x_1 \frac{df(X)}{dx_1} = x_1(\overline{x_2} + \overline{x_3} + \overline{x_2 x_3}) \tag{6-1}$$

$$\overline{x_1} \frac{df(X)}{dx_1} = \overline{x_1}(\overline{x_2} + \overline{x_3} + \overline{x_2 x_3}) \tag{6-2}$$

By symmetry,

$$x_2 \frac{df(X)}{dx_1} = x_2(\overline{x_1} + \overline{x_3} + \overline{x_1 x_3}) \tag{7-1}$$

$$\overline{x_2} \frac{df(X)}{dx_1} = \overline{x_2}(\overline{x_1} + \overline{x_3} + \overline{x_2 x_3}) \tag{7-1}$$

$$x_3 \frac{df(X)}{dx_1} = x_3(\overline{x_1} + \overline{x_2} + \overline{x_1 x_2}) \tag{8-1}$$

$$\overline{x_3} \frac{df(X)}{dx_1} = \overline{x_3}(\overline{x_1} + \overline{x_2} + \overline{x_1 x_2}) \tag{8-1}$$

Similarly, for fig. 2 (b) and fig. 2 (c), may also find thier fault test expressions. Choosed different value of x_i ($x_i = 1$ or 0) and fixed others in these pairs of expressions (for example in expression (6-1)and (6-2)), if the value of these pairs expression do chuange, then x_i is the single fault test code, and this set of variables is called combinational fault test code.

Tab.1 lised out all of the combinational fault test code of the logic unit of fig.2 (a), fig.2 (b) and fig.2 (c).

B.Faults identification of combination of logic unit

Based on the analysis above, the fault of the DDS can be identification as follow.

In Figure 1,

$$x_A = (x_1 x_2 + x_1 x_3 + x_2 x_3)\overline{x_4} + x_B = x_{26}\overline{x_4} + x_B$$

$$x_B = (x_5 + x_6 + x_7 x_8)\overline{x_9} + x_C + x_D = x_{27}\overline{x_9} + x_C + x_D$$

$$x_C = (x_{10} x_{11} + x_{10} x_{12} + x_{11} x_{12})\overline{x_{13}} + x_{24} = x_{28}\overline{x_{13}} + x_{24}$$

$$x_D = (x_{14}x_{15} + x_{14}x_{16} + x_{15}x_{16})\overline{x_{17}} + x_{24} = x_{29}\overline{x_{17}} + x_{24}$$

$$x_E = x_{18}x_{19}\overline{x_{20}} + x_{25} = x_{30}\overline{x_{20}} + x_{25}$$

$$x_F = x_{21}x_{22}\overline{x_{23}} + x_{24} + x_{25} = x_{31}\overline{x_{23}} + x_{24} + x_{25}$$

where $x_{26} = x_1x_2 + x_1x_3 + x_2x_3$

$$x_{27} = x_5 + x_6 + x_7x_8$$

$$x_{28} = x_{10}x_{11} + x_{10}x_{12} + x_{11}x_{12}$$

$$x_{29} = x_{14}x_{15} + x_{14}x_{16} + x_{15}x_{16}$$

$$x_{30} = x_{18}x_{19}$$

$$x_{31} = x_{21}x_{22}$$

Because of

$$\frac{df(X)}{dx_{28}} = (x_{28}\overline{x_{13}} + x_{24}) \oplus (\overline{x_{28}}\,\overline{x_{13}} + x_{24}) = \overline{x_{13}}\,\overline{x_{24}}$$

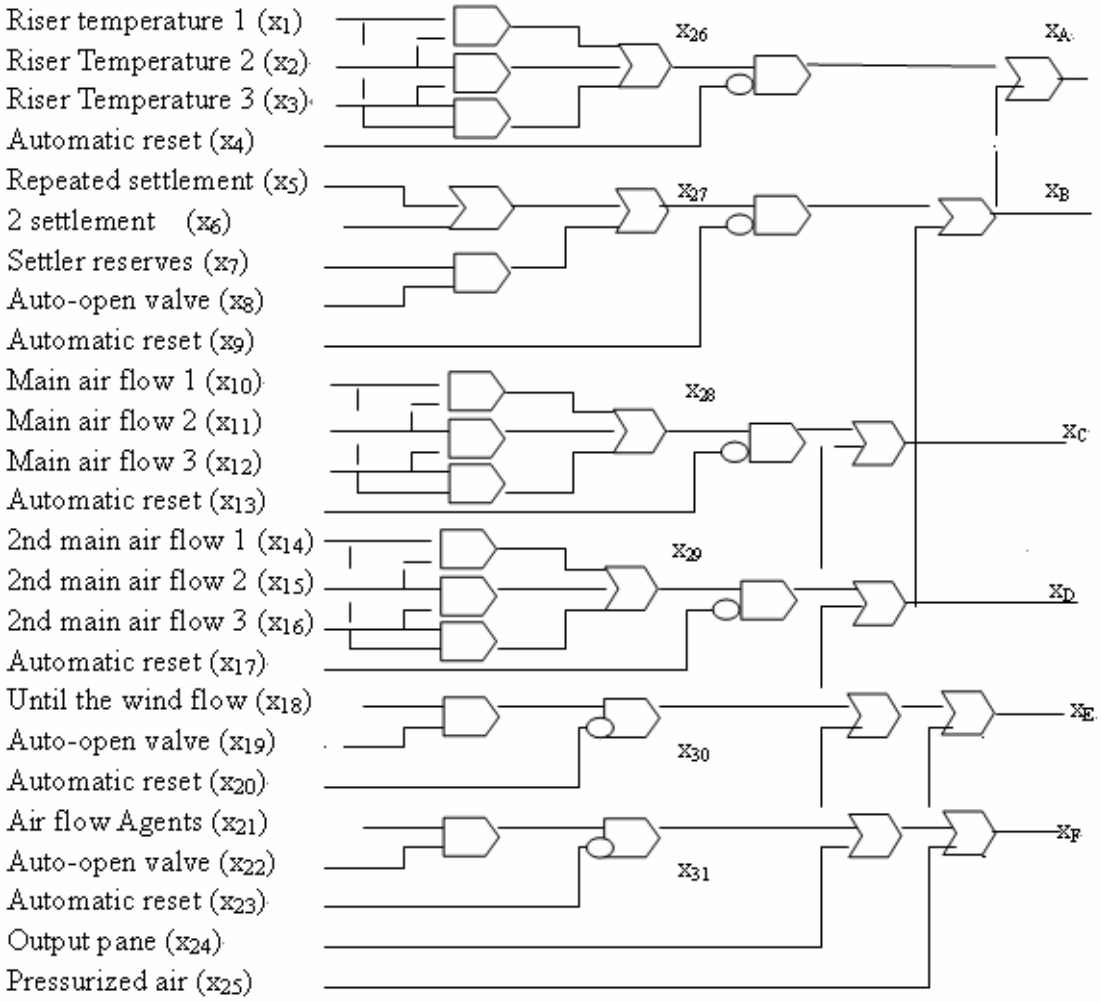


Fig.1 Main logic diagram of combination logic unit

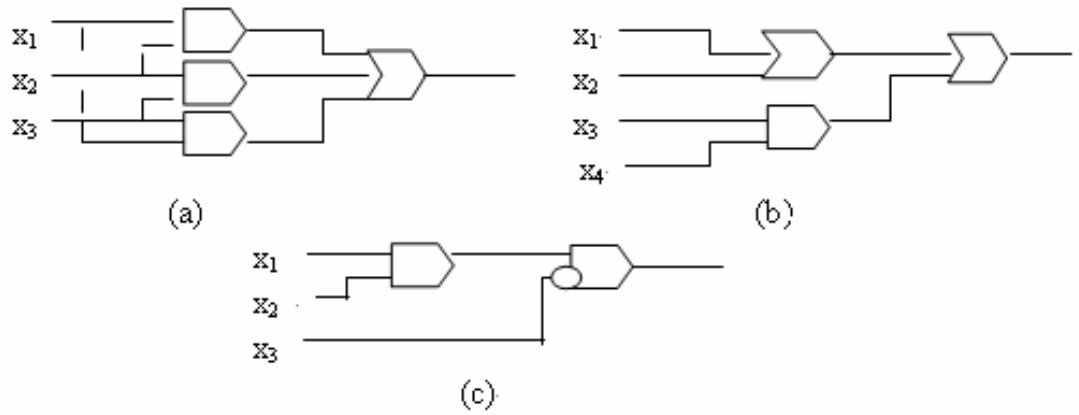


Fig.2 Basic logic unit in fig.1

Tabal.1 Combinational fault logic test code of FCC unit

Fig		x_1	x_2	x_3	x_4
Fig.2(a)	**s-a-0 fault**	{0,1,0},{0,0,0},{0,0,1}	{1,0,0},{0,0,1},{0,0,0}	{1,0,0},{0,1,0},{0,0,0}	
	s-a-1 fault	{1,1,0},{1,0,0},{1,0,1}	{1,1,0},{0,1,1},{0,1,0}	{1,0,1},{0,1,1},{0,0,1}	
Fig.2(b)	**s-a-0 fault**	{0,0,0,1},{0,0,1,0},{0,0,0,0}	{0,0,0,1},{0,0,1,0},{0,0,0,0}	{0,0,0,1}	{0,0,1,0}
	s-a-1 fault	{1,0,0,1},{1,0,1,0},{1,0,0,0}	{0,1,0,1},{0,1,1,0},{0,1,0,0}	{0,0,1,1}	{0,0,1,1}
Fig.2(c)	**s-a-0 fault**	{0,1,0}	{1,0,0}	{0,1,0},{0,0,0}{1,0,0},{1,1,0}	
	s-a-1 fault	{1,1,0}	{1,1,0}	{0,1,1},{0,0,1},{1,0,1},{1,1,1}	

then

$$x_{28}\frac{df(X)}{dx_{28}} = (x_{10}x_{11} + x_{10}x_{12} + x_{11}x_{12})\overline{x_{13}}\,\overline{x_{14}}$$

$$\overline{x_{28}}\frac{df(X)}{dx_{28}} = (\overline{x_{10}} + \overline{x_{11}})(\overline{x_{10}} + \overline{x_{12}})(\overline{x_{11}} + \overline{x_{12}})\overline{x_{13}}\,\overline{x_{14}}$$

Similarly, we have

$$x_{29}\frac{df(X)}{dx_{29}} = (x_{14}x_{15} + x_{14}x_{16} + x_{15}x_{16})\overline{x_{17}}\,\overline{x_{24}}$$

$$\overline{x_{29}}\frac{df(X)}{dx_{29}} = (\overline{x_{14}} + \overline{x_{15}})(\overline{x_{14}} + \overline{x_{16}})(\overline{x_{15}} + \overline{x_{16}})\overline{x_{17}}\,\overline{x_{24}}$$

$$x_{30}\frac{df(X)}{dx_{30}} = x_{30}\overline{x_{20}}\,\overline{x_{25}}$$

$$\overline{x_{30}}\frac{df(X)}{dx_{30}} = \overline{x_{30}}\,\overline{x_{20}}\,\overline{x_{25}}$$

$$x_{31}\frac{df(X)}{dx_{31}} = x_{31}\overline{x_{23}}\,\overline{x_{24}}x_{25}$$

$$\overline{x_{31}}\frac{df(X)}{dx_{31}} = \overline{x_{31}}\,\overline{x_{23}}\,\overline{x_{24}}\,\overline{x_{25}}$$

So that the combinational fault test codes of DDS may be computede which are not listed in this paper becaose of too numerous.

IV. CONCLUSION

(1) The method of fault test codes for identification DDS failures are proposed and be applied in an examples;
(2) The results indicate that

a) According to the property of the Boolean difference, the codes are available to fined DDS failures;

b) In order to identification of DDS failure it is necessary to properly design the combinational logic unit for ensure it has the fault test codes;

c) If the DDS failure that had been judged to be identification, the fault test codes should always be used in input of DDS timely to implement self fault test so that its failure may be able to find. Through such means the DDS functions of monitor the production status, prevent and control the potential risk should be achieved correctly.

REFERENCES

[1] Wang Kaiquan, On the Model of Petrochemical Process Safety Control[J]. Journal of Jiangsu Institute of Petrochemical Technology, Vo1. 11 No.4 Dec. 1999:20-22

[2] Tang Yonghong, system reliability, fault tolerance cut off, and [M]. Chongqing: Chongqing University Press, 1990 edition; 58-79

[3] Yang Yinzhe, the application of intelligent control and research [J]. Scence and Technology Innovation Herald, Vo1. 9 No.22. Nov. 2008:38-42

[4] Gu Li-ping, the use of reliability technology [M]. Beijing: Mechanical Industry Press, .1992 first edition :55-69

[5] Yang Junhua, Fault Diagnosis of the Digital Logic Circuit Bast on Boolean Difference Calculus[J], Journal of University of Electronic Science and Technology of China, Vo1. 14 No.4 Dec. 2005:44-47

[6] Wang kaiquan,. Selection of Assessment Method for Safety Monitoring System[J]. Industrial Safety and Environmental Protection, Vo1. 29 No.1.Jan.2003:230-25

[7] Wang kaiquan,. Study on the Safety system Reliability[J]. Journal of Shanghai Maritirne University, Vo1. 22 No.3 Sep. 2001:169-172

[8] Thannamani G, Narendran TT. Assessment of availabilily of' a fluid catalytic craking uint through simulation[J]. Relialbility Engineering and System Safety, Vo1.20 No.47.1995: 207-220

[9] Wang kaiquan, The Analysis of the Redundant Safety System Reliability[J] Journal of Petrochemical University, Vol.16 No.1 Mar. 2003:64-67

[10] Wang kaiquan, Petrochemical Process Safety and Availability Analyses[J]. Journal of Petrochemical University, Vol.11 No.3. Sep. Mar. 1998:52-54

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 120-123

Provably Correct Aspect-Oriented Self-monitoring Programs

Li Li[1], and GuoSun Zeng[2]
[1]Department of Computer Science and Technology, Tongji University, Shanghai 201804, China
Email: snopy-xj@163.org
[2]Department of Computer Science and Technology, Tongji University, Shanghai 201804, China
Email: gszeng@sina.com

***Abstract*—In-lining monitor into software to produce self-monitoring programs is a powerful, flexible method to enforce runtime monitoring over untrusted code. The preconditions of its application are formal proofs which show that these monitor inlined programs will comply with the intended security policies and monitor has no bad effect on original programs. But at present, the existing methods can only prove that the programs comply with the safety policies. Thus a new method is proposed in this paper. Firstly, we encode security policies and develop monitor by means of Aspect-oriented programming. Secondly, adopting Alternating Transition System and Alternating-Time Temporal Logic, an abstract program structure for monitor inlined program is constructed, and the formulas, which describe the correct executions of programs, are defined according to security policies. Finally, by checking the validity of formulas on abstract structure, the conclusions of whether the executions of the monitor-inlined program complies with the security policies and whether the monitor has some bad effect on original program can be get. This method can prove that monitor can execute broader classes of policies, including important classes of liveness policies. Its validity exemplified lastly.**

***Index Terms*—Monitor-inlined program; Correct; Aspect-oriented; Alternating-Time Temporal Logic**

I. INTRODUCTION

In-lining monitor into software to produce self-monitoring programs is an efficient, powerful method to enforce execution monitoring over untrusted code according to some security policies[1,2]. The general implementation of this method is that the untrusted code or the binary executables are automatically rewritten according to security policies, which may be defined by some kind of policy specification languages. The outputs of this rewrite process are monitor in-lined self-monitoring programs.

Though powerful and flexible as this method is, it must be certified correct. Correct implicates two abstract properties, which we call the soundness and transparency. Soundness means that the self-monitoring programs should be guaranteed to adhere to the security policies. Transparency indicates that the in-lined monitor code should not impair the functions of the original programs. Mobile [3] is a certifying in-lined reference monitoring system for the Microsoft .NET framework. It rewrites .NET CLI binaries according to a declarative security policy specification, producing a proof of policy-adherence in the form of typing annotations in an effect-based type system. These proofs can be verified by a type-checker to guarantee policy-adherence of code with in-line monitors.Aktuga et al. [4] designed a two-level class file annotation scheme using Floyd-style program logic for Java bytecode, characterizing two key properties: (i) that the program adheres to a given policy, and (ii) that the program has an embedded monitor for this policy. They sketch a simple in-lining algorithm, and show how the two-level annotations can be completed to produce a fully annotated program. This method establishes the *mediation* property, meaning that in-lined programs are guaranteed to adhere to the intended policy. Those two methods can only certify that a monitor in-lined program adheres to certain safety policies- police specifying that "nothing bad ever happens". Neither can establish the transparency property.

In this paper, we try to deal with the certification of self-monitoring programs in a different way. Firstly, we encode the security policies to build monitors by means of Aspect-oriented programming (AOP), and by virtue of the weaving process of AOP, monitor is in-lined into primary program. This process will produce an Aspect-oriented self-monitoring (AOSM) program. Secondly, based on the characteristics of AOSM program, adopting Alternating-Time Temporal Logic and Alternating Transition System, we abstract execution structure of an AOSM program, and define formulas which characterize the soundness and transparency properties of an AOSM program. Finally, by model checking the validity of formulas against program's structure, the conclusions of whether AOSM is correct can be got.

II. AOSM PROGRAM

AOP[5] have been proposed to deal with the code tangling and scattering problems driving from the apartment of crosscutting concerns. It especially realized in the AspectJ languages [6]. We now exemplify how to encode a security policy into Aspects to Produce AOSM programs.

Example 1 : There is a Separation of Duty policy, which claim that *critical ()* operation is performed only

Supported by the china 863 program under grant of 2009AA012201, the National Natural Science Foundation of China under grant of 90718015，the joint of NSFC and Microsoft Asia Research under grant of 60970155, the Ph.D. Programs Foundation of Ministry of Education under grant of 20090072110035.

AP-PROC-CS-10CN006

under the endorsement of both the *manager ()* and *accountant ()* operations. We encode this policy using AspectJ language. The code is shown in Fig. 1.

```
bool pm = false;
bool pa = false;
……
manager();
if （…） { accountant(); }
critical();
```

(a) Base Code

```
aspect Ma
{  Pointcut m()}:
   call (* manager())
   &&target(P);
   After m() { pm = true;}
}
```

```
aspect Cr
{ Pointcut c()}:
   call (* critical())
   &&target(P);
   before c()
   if (pm ∧ pa)
     { pm = false; pa = false; }
   else throw IRMException();
}
aspect Ac
   {  Pointcut a()}:
        call (* accountant())
        &&target(P);
        After a() { pa = true;}
   }
```

Figure 1. base code and aspect of example 1

The base code in block (a) is the primary program. It defines two global variables, *pa* and *pm*, and executes three security-related functions: *accountant(), manager()* and *critical()*. According to the policy, *critical()* can be executed only after both *accountant()* and *manager()* have been executed. To execute this policy, we create three aspects. First, *aspect Ma* defines *Pointcut m*, which is located at the function *manager().* The type of advice is *After*. Thus, the action defined by *aspect Ma* is that after *manager()* has finished, the variable *pm* is set to "*true*". Next, *aspect Ac* is defined similarly to *aspect Ma*, but sets *pa* to true after *accountant()* has finished. Finally, *aspect Cr* defines *Pointcut c()* at the function *critical().* The type of advice is *Before*, so the actions defined by *aspect Cr* take place before the execution of *critical().* The code fragment evaluates *pa* and *pm*; if both are *true*, it executes the *critical()* function and sets pa and pm to false. Otherwise, it throws an exception.

At compilation, the aspect weaver incorporates all three aspects into the base code to produce a AOSM. Figure 2 displays the code for the resulting program.

```
……
manager();
pm = true;
if （…） {accountant();
   pa = true;}
if (pm ∧ pa)
   { pm = false; pa = false;
     critical();}
else throw new IRMException();
```

Figure 2. Policy-adherence program after Aspect weaving

III. CORRECT VERIFICATION MODEL FOR AOSM PROGRAM

Based on the characteristics of AOSM programs, we construct a verification model to verify the correct property of monitor inlined programs.

A. Abstract structure

Alternating-Time Temporal Logic (ATL) and Alternating Transition System (ATS) [7, 8] are logical specification tools for open system. We specify an AOSM program as a Turn-Based Alternating Transition System (Turn-based ATS). The concrete definition given as follows:

Definition 3.1 An AOSM program structure (AOSM structure) can be abstracted as a Turn-based ATS. Expressed as tuple AOSM Structure=<Σ, Q,Π, π , σ , δ> with the follow components:

1. Σ is a set of players, which includes system components Aspect and BaseCode, as well as Environment.
2. Q is a finite set of state;
3. Π is a finite set of proposition;
4. Function π :Q→2Π is a labeling function, which maps each state q∈Q to a set π(q)$\subseteq$ Π. π(q) is the set of propositions true at state q.
5. Function σ: Q→ Σ map a state q to a player aq. Representing that at state q, it is the turn of player aq to choose the next execution steps of program. Nature number da(q)≥1 is moves available at state q for player a. We identify the moves of player a at state q with the numbers 1,……,da(q). For each state q∈Q, a move vector at q is a tuple < j1; : : : ; jk> such that 1 ≤ja≤da(q) for each player a. For other players b∈ Σ\ aq at state q, db(q)=1;
6. δ(q, J_a) is a transition function. When a_q choose action j_a, the state will transit to next state q`=δ(q, J_a) ∈Q.

B. Defination of correct property

We now describe soundness and transparency properties in ATL formulas in this section. So it needs to interpret the syntax of ATL at first:

Definition 3.2 An ATL formula is one of the following:

1. Proposition p，P∈Π;
2. ¬ φ or φ1∨φ2, where φ、 φ1、 φ2 are ATL formulas;
3. <<A>>○ φ , or <<A>>□ φ , or <<A>>◇ φ , or <<A>> φ1 u φ2 are ATL formulas, where A$\subseteq\sum$is a set of players, and φ、 φ1 and φ2 are ATL formulas.

The operator <<>> is path quantifier, and ○("next"), □("always"), ◇ ("eventually"), and U ("until") are temporal operators. <<A>>represent path chosen by players in set A. Quantifier<<>>also has a dual form 〖〗. While formals<<A>>φ means that the players in A can cooperate to make the φ true, the dual form 〖A〗 φ means that the players in A can not cooperate to make the φ false.

Based on above definitions of ATL, we now formulate the correct property of self-monitoring program on AOSM structure.

Definition 3.3 (Soundness) Soundness means that the self-monitoring program should adhere to the intended policy. That is to say, on an AOSM structure, all the path decided by the strategies of players Aspect and BaseCode should satisfy the characters φ of policy, no matter how

the Environment chooses. Expressed as an ATL formula φ s=<<Aspect, BaseCode>> φ.

Definition 3.4 (Transparency) Transparency means that monitoring code should not impair the characters, such as function or semantic, of original program. That is to say, on an AOSM structure, all paths decided by the strategy of BaseCode should satisfy the characters φ of original program, no mater how the other players chooses. Expressed as an ATL formula φt=<<BaseCode>> φ。

φt would be described differently according to the characters of original program:

Given AOSM structure and soundness and transparency ATL formulas of an AOSM program, we can conclude whether this program satisfy the correct property by model-checking the formulas against the abstract structure.

V. AN EXAMPLE

Example.2 A program module makes use of the point class, which has a method move, to move a point on canvas. The program interface contains a canvas, two numeric text fields where the user can fill in *x* and *y* coordinates, and press "ok" button to move the point to the specified location on the canvas. The program only reads the text fields; it does not write to them. Their values are therefore determined by the environment alone. We assume that the text fields only accept non-negative coordinates.

```
Class Point
 { int x, y;
   int scaleFactor=1;
   public int getX() {return x}
   public int getY() {return y}
   public coordinate(int a, int b)
    {this.x=a; this. y=b;}
   public void move (int nx, int ny)
   { x=nx*scaleFactor;
    y=ny*scaleFactor;}
}
……
   While（！Cancel）
   {   if (Ok)
         p.coordinate (getX(),getY())
         p. move (x , y)
   }
```

(Base Code)

```
aspect adjustscale
{
Pointcut m(Point p):
   execution(void
Point.move(int,int))
        &&targets(p);
 Before m()
        { if ((p.x<5)&& (p.y<5))
            {p.scaleFactor=10;}
        }
 After m()
            {p.scaleFactor=1;}
 }
```

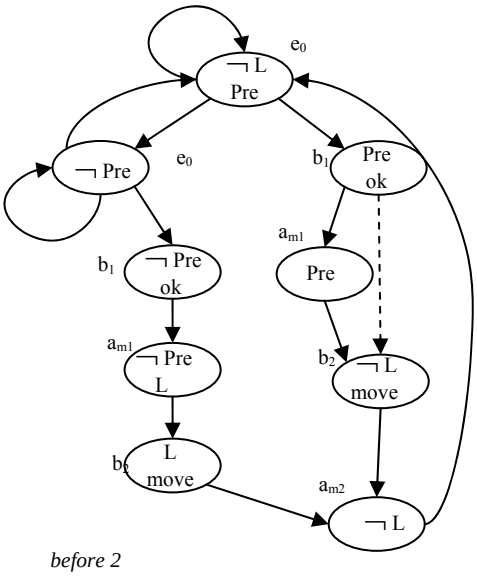
(Aspect)

Figure 3. Code of Example 2

For clarity of display, we add a constraint policy on this program's executions. When the point's location is too close to the origin, e.g. 0<*x*<5 and 0<*y*<5, the base code multiplies the coordinate by a significant factor, in this case 10. This factor is stored in the variable *scaleFactor*. Conversely, if $x \geq 5$ or $y \geq 5$, the coordinate is left alone. For the display to continue working properly, *scaleFactor* must be reset to 1 at the end of the program. This policy is encoded as an aspect named *adjustscale*. The code is shown in Fig. 3. Henceforth, the constraints are written as $(x, y) < 5$ and $(x, y) \geq 5$ for simplicity.

A. Abstract structure

Adopting the method proposed in section 4, we construct the PAP Structure of the code in Fig. 4. There are three players: Environment. BaseCode, and Aspect. Environment sets the inputs of the program and presses the button. BaseCode executes the move function. Aspect executes the constraint policy. Henceforth we use the characters e, b and a to represent the three players. Figure 5 represents our PAP Structure visually using state transition diagrams.




```
before 2
m1：if ((p.x<5)&& (p.y<5)) {p.scaleFactor=10;}
while (!cancke)
{0：if (ok )
1：p.coordinate (getX(),getY())
2：p. move (x , y)
}
After 2
m2：p.scaleFactor=1;
```

Figure 4. Abstract Structure of Example 2

On the bottom of Fig. 4, we label lines of code with 0, 1, 2, m1 or m2. On the top of Fig 4, each circle with label inside describe which propositions hold in that state , such as, at root state, propositions ¬ L and pre hold.. Those propositions are abstractions of the concrete program states, which are defined in Table 1. The symbol e, b or a, beside the circle indicates the active player at that state. For example, the symbol beside the root state is e_0, meaning that Environment is the active player at program point *0*.

Assume that the program is at point *0*, and that both coordinates (*x*,*y*) are currently greater than or equal to 5. The Environment can choose between entering coordinates $(x, y) \geq 5$, entering coordinates (*x*, y)<5, or pressing the ok button.

If the ok button is pressed, the program will execute the right-hand subtree. BaseCode gets the values of the coordinates, then attempts to advance the program

TABLE I.
DESCRIPTION OF PROPOSITION VARIABLES

proposition variable	meaning	Proposition variable	meaning
Pre	x or y≥5	¬ pre	0＜x and y＜5
L	scaleFactor=10	¬ L	scaleFactor=1
ok	button is pressed	move	Execute move function

counter to point *2*. Aspect interrupts the flow, moving the program to point *m1*. Since $(x, y)\geq 5$, Aspect does not increase the *scaleFactor* value and returns control to BaseCode at point *2*. After the move function is finished, Aspect interposes itself again and sets *scaleFactor* to 1.

If the Environment enters coordinates in the range $0<(x, y)<5$ and then presses the button, Aspect would set scaleFactor to 10 at point m1 but change its value back to 1 after the move function is complete.

B. Correct formulas and verification

- soundness formulas

1) According to the policy, the coefficient *scaleFactor* should be reset to 1 eventually. This is a liveness policy(policy states that nothing irremediably bad happens, or that good things will happen eventually.). Thus, we have the $\langle\langle\rangle\rangle\Box$ (move $\rightarrow \langle\langle$ a, b $\rangle\rangle \Diamond \neg$ L. Whenever the move function is finished, Aspect and BaseCode should cooperate to restore *scaleFactor* to its normal value before the program terminates.
2) According to the security policy, program execution should always obey the following logic: if $(x,y)<5$, then the *move* function is executed with *scaleFactor*=10; if $(x,y)\geq 5$, then the *move* function is executed with *scaleFactor*=1. Thus, we have formulas:
   a) $\langle\langle\rangle\rangle\Box$ ((¬ pre ∧ ok) → ⟨⟨a, b⟩⟩□ ¬ (¬ L ∧ move): whenever $(x,y)<5$ and the button is pressed, on all the path choose by Aspect and BaseCode, there is no state satisfy *move* with *scaleFactor*=1
   b) $\langle\langle\rangle\rangle\Box$ ((pre ∧ ok) → ⟨⟨a, b⟩⟩□ ¬ (L ∧ move): whenever $(x,y)\geq 5$ and the button is pressed, on all the path choose by Aspect and BaseCode, there is no state satisfy execute *move* with *scaleFactor*=10.
   c) $\langle\langle\rangle\rangle\Box$ ((¬ pre ∧ ok) → [[e]]◊(L ∧ move): whenever $(x,y)<5$ and the button is pressed, Environment cannot force the program to execute *move* with *scaleFactor*=1.
   d) $\langle\langle\rangle\rangle\Box$ ((pre ∧ ok) → [[e]]◊(¬ L ∧ move): whenever $(x,y)\geq 5$ and the button is pressed, Environment cannot force the program to execute *move* with *scaleFactor*=10.

- transparency formulas

Recall that the original function of BaseCode is to move a point on canvas. This function should not be stopped by any behavior of Aspect. Thus, we have the following formula:

3) $\langle\langle\rangle\rangle\Box$ (ok → ⟨⟨b⟩⟩◊move): Whenever the button is pressed, BaseCode must execute *move* eventually.

We can successfully verify those six formulas on program' abstract structure. Hence, we conclude that the program in Fig. 4 has the *correct* property.

Ⅵ. CONCLUSIONS

This paper focuses on the correct verification of AOSM programs. Comparing with similar research, this method can verify: 1) whether self-monitoring program comply with a wide range of security policies, not limited to "safety" ones, 2) whether monitor code has some side effect on original program. 3) Have less limitations on policy definition.

REFERENCES

[1] úlfar Erlingsson. The Inlined Reference Monitor Approach to Security Policy Enforcement. PhD thesis, Cornell University, Ithaca, New York. 2004.

[2] J. Ligatti, L. Bauer, D. Walker, "Run-Time Enforcement of Non-safety Policies," ACM Transactions on Information and Systems Security, 2009, vol. 12, pp.19:1-19:41.

[3] Kevin W. Hamlen, Greg Morrisett, and Fred B. Schneider. "Certified in-lined reference monitoring on .NET," In: Proceedings of the 1st ACM Workshop on Programming Languages and Analysis for Security, Ottawa, Canada. june 2006, pp. 7–15.

[4] Irem Aktug, Mads Dam, Dilian Gurov. "Provably correct runtime monitoring," Elsevier: The Journal of Logic and Algebraic Programming. 2009, vol. 78, pp.304–339.

[5] Kiczales Gregor, Lamping J., Medhdhekar Anurag et al. "Aspect-Oriented Programming," In: Proceedings of the 11th European Conference on Object-Oriented Programming, Finland. 1997, 1241pp.220–242.

[6] Kiczales Gregor, Hilsdale E., hugunin J., kersten M., palm J., Griswold W. "An overview of AspectJ," In Proc. of the European Conference on Object-Oriented Programming. 2001, pp.327–353.

[7] Rajeev. Alur, Thomas A. Henzinger, Orrna Kupferman. "Alternating-Time Temporal Logic," Journal of the ACM. 2002, vol.49, pp.672–713.

[8] Rajeev. Alur, L.de Alfaro, R. Grosuz etc. "JMOCHA: A Model Checking Tool that Exploits Design Structure," In: Proceedings of the 23rd International Conference on Software Engineering. 2001(ICSE'01), pp.8-35

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 124-128

A New Information Hiding Method for Image Watermarking Based on Mojette Transform

Cui Zhang[1], Junyu Dong[1], Jining Li[2], and Florent Autrusseau[3]
[1] Computer Science Department, Ocean University of China, Qingdao, China
Email: zhangcui1985@163.com, dongjunyu@ouc.edu.cn
[2] Weifang Public Security Bureau, Weifang, China
Email: wfjjljn@163.com
[3]École Polytechnique de l'Université de Nantes, Nantes, France.

***Abstract*—This paper describes a new image watermarking method for which is suitable for both copyright protection and information hiding. The presented method is based upon the morphological mathematics properties of the Mojette Transform[1,2] and the Mojette Phantoms[3]. The main properties of the Mojette transform are briefly introduced and the concept of linked phantoms which depicts the null space of the operator is presented. In this paper the Mojette Phantoms can be used not only as the embeded watermark, but also can be used as the mark which is inscribed with some certain information, e.g. Chinese characters. Corresponding embedding and extractions of either the mark or the hidden message are then described. Finally, experimental results are presented in the last section.**

***Index Terms*—Mojette Transform; copyright protection; Mojette Phantoms;watermarking.**

I. INTRODUCTION

Inscribing invisible marks into an image has a variety of applications such as copyright, steganography or data integrity checking. There are many techniques can be used for watermarking, such as DCT , DWT and so on . In [3], the authors presented a method which use the Mojette Phantoms as the mark and hide the mark in the Mojette domain. However, the mark used in [3] is meaningless. In this paper we present a new method that can hide watermark which bear certain meanings. In particular, we implement the proposed method so that Chinese characters contain designed meanings can be hidden in the input image.

THE MOJETTE TRANSFORM

A. Direct Mojette Transform (MT)

The Mojette Transform is a kind of discrete Radon transform[4]. This transform needs many angles such as $\theta = arc\tan\left(\frac{q}{p}\right)$, where ($p$,$q$) are integers restricted to PCD(p,q) = 1. For an image at each angle, we can acquire a group of projections. Every projection is called a bin. The process of the transform can be described by the following $M_{p,q}$ operator :

$$M_{p,q} f(k,l) = \sum_{k=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} f(k,l)\Delta(b+kq-lp) \text{ where}$$

$$\Delta(b) = \begin{cases} 1, if b = 0 \\ 0, if b \neq 0 \end{cases}, \quad (1)$$

and $M_I f(k,l)$ is denoted as the set of I projections: $M_I f = \left\{ M_{pi,qi} f; i \in \{1,2,...I\} \right\}$. In another way, given an image and an angle, its projections are acquired by adding together the gray value of the pixels along the straight lines of m+qk-lp=x. Figure 1 shows an example.

B. Inverse Mojette transform

The inverse Mojette transform is the process of inverse the projections into the origin gray values of pixels. There have many ways to implement the inverse Mojette transform and we usually use the simplest one: first, select a projection which contains only one pixel; second, select the projections which contain this pixel and subtract the gray value of this pixel from these projections. Then do the above steps repeatedly until all the pixels are retrieved. This is shown in Figure 2 .

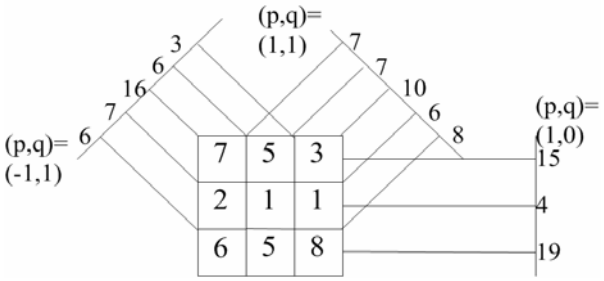


Figure 1. The direct Mojette transform for a 3 × 3 support and 3 projections.

C. Mojette phantoms

The Mojette phantoms is an image which has the projections for all the given angles are zeros. The virtue of the Mojette phantoms is when adding the Mojette phantoms into an image the image's projections for the given angles will not change. Figure 3 show the properties of the Mojette phantoms.

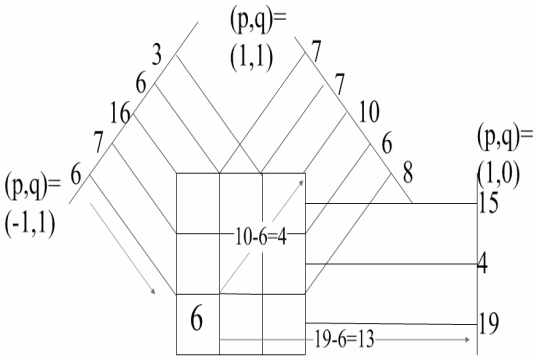


Figure 2. The process of the inverse Mojette transform.

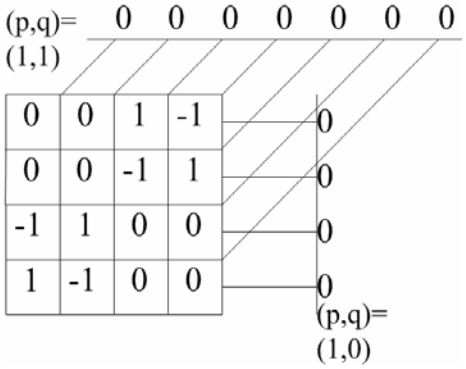


Figure 3. the Mojette phantoms

THE WATERMARKING ALGORITHM

D. Watermark processing

The Mojette phantoms can be embedded into the original image as the watermark. However, the original Mojette phantoms can not carry any meaningful information. In contrast we present a new method which allows the watermark composed by the Mojette phantoms to be able to hide meaningful information. The idea is to let the Mojette phantoms have a small size, and the watermark comprises multiple blocks of the Mojette phantoms.

The specific process is as follow:

Step 1: Design the information to be carried in the watermark.

Step 2: According to the information of the watermark, we construct a watermark by the phantoms(Figure 3) and a block whose pixel gray values are all zeros. We merge the blocks of the phantoms and the zero blocks as every block is a pixel. If there is the information we fill the watermark by the phantoms blocks. Otherwise it is filled by the zreo blocks.

Figure 4 shows how to design the watermark. An example is showed in figure 5. We can see meaningful information is hidden in the watermark. In this way we can hide meaningful watermark in the original image and achieve information security. For example, the information hidden in the image can be the name of the owner or the number which has a special meaning.

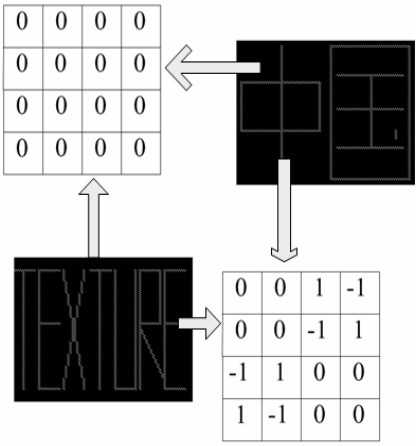


Figure 4. How to design the watermark. The areas we want are filled by the phantoms blocks otherwise they are filled by zero.

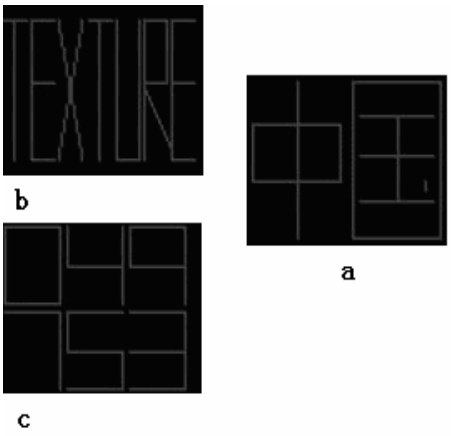


Figure 5. A set of watermarks composed of the Mojette phantoms with meaningful information.(a).a watermark containimg two chinese characters meanimg China (b) a watermark which is an English word ‘texture’ (c) a watermark which is the code of a traffic police office

E. Embedding scheme

For a given image and the information of the watermark, we should design the block size of the Mojette phantoms and the angles which are used to take projections. The number of the angles is denoted by *I*. We embed the watermark in the following steps:

Step 1: For the origin image,we segment it into small blocks and denote the size of the block by $n \times n$. The purpose of this produre is to facilitate the calculation of the Mojette transform.

Step 2: For each block, we take the Mojette transform at the given *I* angles. At every angle, all the projections can be regarded as a set of projections. So we acquire *I* projections.

Step 3: According to the first (*I*-1) angles, we can create the correlative Mojette phantoms whose size is $n \times n$. We can obtain the watermark by splicing the small blocks of the Mojette phantoms and the zero blocks.

Step 4: Embedding the watermark[5] in every block of the origin image. Then we can get the watermarked image.

Step 5: Segment the marked image into small blocks same as the origin image.

Step 6: For every blocks of the marked image, we take the Mojette transform at the given *Ith* angles. The (*I*+1)th projection is acquired .

Figure 6 shows the process of embed watermark.

Once the embedding scheme is finished, we can transmit the projections.

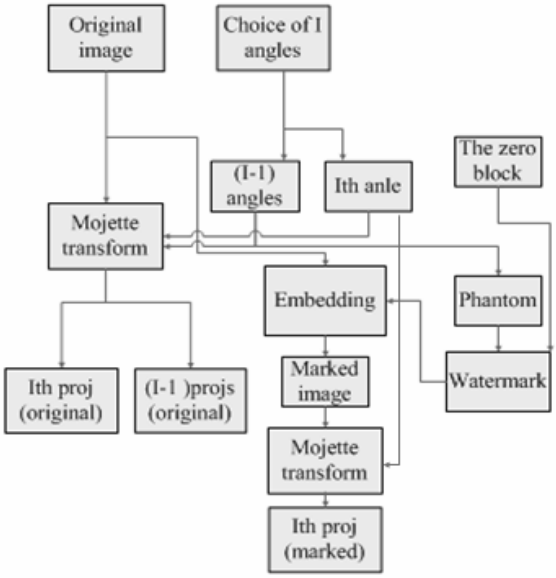


Figure 6. Eembedding watermark process

F. Decoding algorithm

When the receiver receives the (*I*+1) projections, the origin image and the marked image can be retrieved according to the projections and the given angles . The first *I* projections are used to obtain the origin image, and the marked image is generated by the first *(I-1)* projections and the last projection. The specific process can be described by the following steps:

Step 1: For the first *I* projections, take the inverse Mojette transform to obtain the original image based on the size of the block,.

Step 2: For the first (*I*-1) projections and the last projection, take their inverse Mojette transforms to generate the marked image.

Step 3: Subtract the original image from the marked image to retrieve the watermark.

Figure7 shows the process of extract watermark.By comparing the extracted watermark and the original watermark, we can find whether the projections have been changed.

EXPERIMENTAL RESULTS

We use a set of car images and watermarks in Figure 5 as the input, and thus we can get the marked images. The watermark is composed of a number of blocks as depicted in Figure 3. The angles used to take projections are (-2,1),(1,1),(1,0) and those used to gain the phantoms are (1,1),(1,0). Figures 8,9,10 show a set of experiment results.

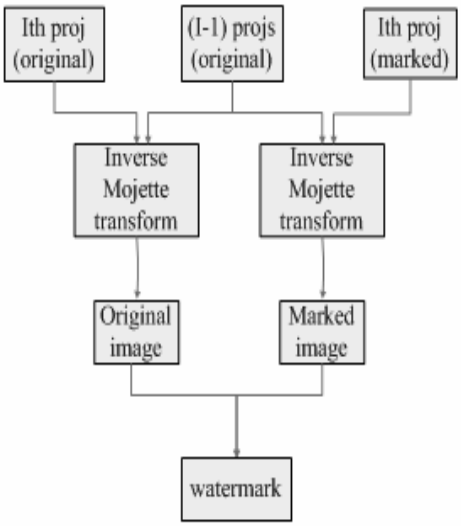


Figure 7. How we extract watermark

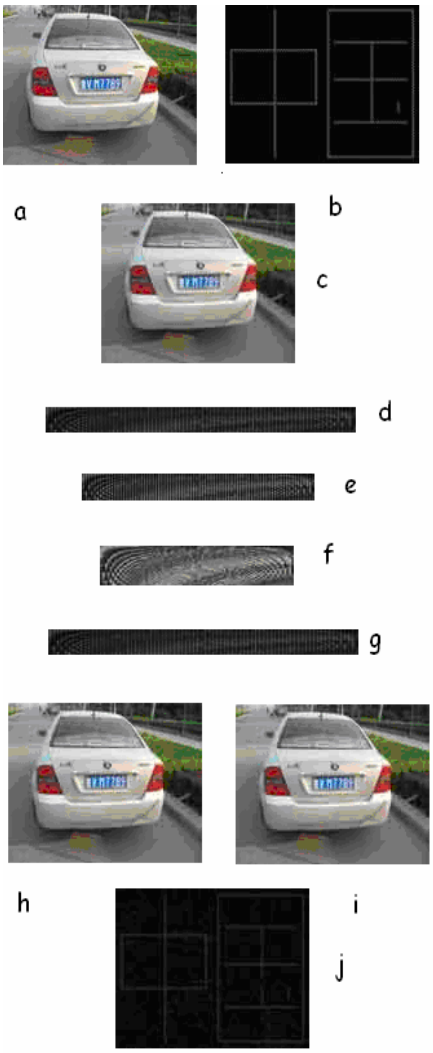


Figure 8. (a)the original image.(b) the watermark.(c)the marked image (d)the projections of the original image at angle (-2,1).(e) the projections of the original image at angle (1,1).(f) the projections of the original image at angle (1,0).(g) the projections of the marked image at angle (-2,1). (h) the retrieved origin image.(i) the retrieved marked image. (j) the extracted watermark.

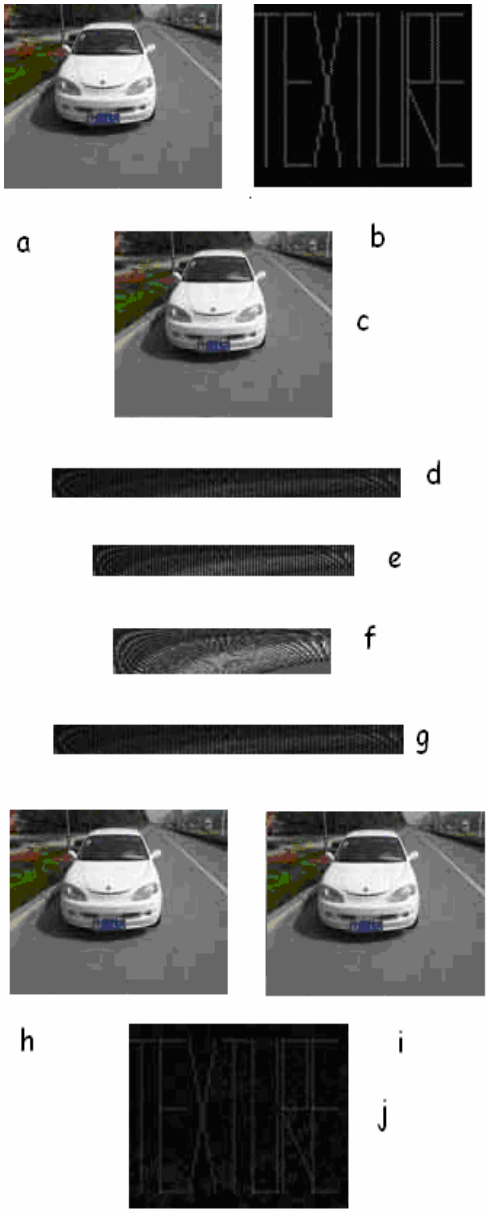


Figure 9. (a)the original image.(b) the watermark.(c)the marked image (d)the projections of the original image at angle (-2,1).(e) the projections of the original image at angle (1,1).(f) the projections of the original image at angle (1,0).(g) the projections of the marked image at angle (-2,1). (h) the retrieved origin image.(i) the retrieved marked image. (j) the extracted watermark

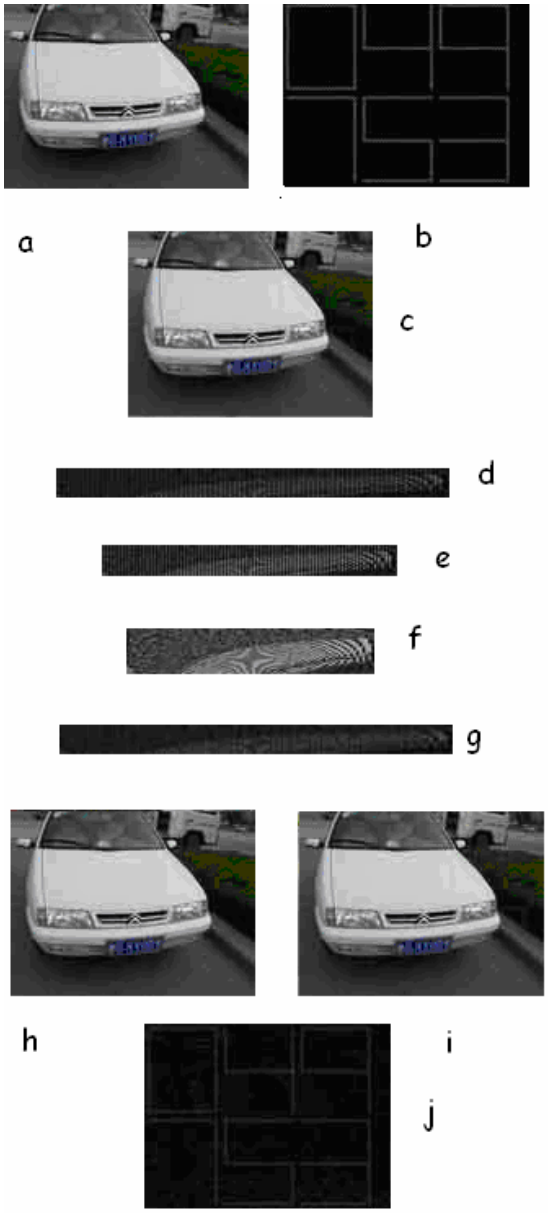


Figure 10. (a)the original image.(b) the watermark.(c)the marked image (d)the projections of the original image at angle (-2,1).(e) the projections of the original image at angle (1,1).(f) the projections of the original image at angle (1,0).(g) the projections of the marked image at angle (-2,1). (h) the retrieved origin image.(i) the retrieved marked image. (j) the extracted watermark

CONCLUSION

A new watermarking method based on the mojette transform that can hide special information in the original image is presented in this paper. Experimental results show it is effective once the meaningful watermark is successfully constructed. However, it should be noted that this method might have problems if complex information is required to be hidden as efficient construction of this kind of watermark can be difficult. watermark..

ACKNOWLEDGMENT

The work was supported by a grant from National Nature Science Foundations of China (No. 60702014).

REFERENCES

[1] N. Normand, JP. Guédon, “La transformée Mojette : une représentation redondante pour l'image”, Comptes rendus Académie des sciences de Paris, theoretical computer science section, p. 124-127. January 1998.

[2] F. Autrusseau, J. V. Guedon, and Y. Bizais. Mojette cryptomarking scheme for medical images. Volume 5032, pages 958–965. SPIE, 2003.

[3] F Autrusseau, JV Guedon Image Watermarking for Copyright Protection and Data Hiding via the Mojette Transform Proceedings of SPIE, 2002

[4] J. RADON. Ober die Bestimmung von Funktio-nen durch ihre Integralwerte l¨angs gewisser Mannig-faltigkeiten. Computed Tomography, 27, 1983.

[5] F. Autrusseau, A. Saadane, D. Barba “ Psychovisual Approach for Watermarking", SPIE Electronic Imaging : Security and watermarking of multimedia contents, January 2001, San Jose, CA.

[6] M. Kutter and F. A. P. Petitcolas, “A fair benchmark for image watermarking systems” , Proc. SPIE Electronic Imaging Security and watermarking of multimedia contents, San Jose, CA, Vol. 3657, p. 226-239, January 1999.

[7] JP. Guédon, B. Parrein, N. Normand, “ Internet distributed image databases", Integrated Computer Aided Engineering, Vol.8, p. 205-214, 2001.

[8] M. Katz, “Questions of uniqueness and resolution in reconstruction from projections”, Springer Verlag, 1977.

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 129-132

Modeling Seismic Wave Propagation Using Graphics Processor Units (GPU)

Zhangang Wang[1], Suping Peng[1], and Tao Liu[2]
[1]State Key Laboratory of Coal Resources and Mine Safety, China University of Mining and Technology, Beijing, China
Email: millwzg@163.com
[2] School of Earth and Space Sciences, Peking University, Beijing, China
Email: liuluot@126.com

***Abstract*—The main drawback of the seismic modeling in 2D viscoelastic media on a single PC is that simulations with large gridsizes require a tremendous amount of floating point calculations. To improve computation speedup, a graphic processing units (GPUs) accelerated method was proposed using the staggered-grid finite difference (FD) method. The geophysical model is decomposed into subdomains for PML absorbing conditions. The vertex and fragment processing are fully used to solve FD schemes in parallel and the latest updated frames are swapped in Framebuffer Object (FBO) attachments as inputs for the next simulation step. The simulation program running on modern PCs provides significant speedup over a CPU implementation, which makes it possible to simulate realtime complex seismic propagation in high resolution of 2048*2048 gridsizes on low-cost PCs.**

***Index Terms*—seismic, wave propagation, finite difference, viscoelastic media, model, GPU**

I. INTRODUCTION

The main drawback of the FD method is that simulations with large model spaces or long nonsinusoidal waveforms can require a tremendous amount of floating point calculations and run times. In recent years the spatial parallelism on high-performance PC clusters makes it a promising method for seismic numerical computing in realistic (complex) media [1, 2].

Modern GPUs have now become ubiquitous in desktop computers and offer an excellent cost-to-performance-ratio, exceeding the performance of general purpose CPU by many times, due to their powerful and fully programmable parallel processing architectures. In the past few years, programmability of GPU and increased floating point precision has allowed GPU to perform general purpose computations [3, 4].

In this article, we have proposed a new implementation under the current GPU architecture to improve the real-time FD simulation of wave propagation in viscoelastic media, including domain decomposition and processing techniques.

II. THEORY

Early researches on FDM for elastic wave modeling in complex media gradually formulated finite difference schemes based on a system of high-order coupled elastic equations [5-8]. In such studies, however, the earth's viscosity has been ignored and synthetic seismograms fail to model attenuation and dispersion of seismic waves. Day and Minster [9] made the first attempt to incorporate anelasticity into a 2-D time-domain modeling methods by applying a Pade approximant method. An efficient approach [10] was proposed based on the rheological model called "generalized standard linear solid" (GSLS), which was shown to explain experimental observations of wave propagation through earth materials [11]. Then a staggered grid finite difference technique was proposed based on the GSLS to model the propagation of seismic waves in 2D/3D viscoelastic media [12], for staggered-grid operators are more accurate than standard grid to perform first derivatives for high frequencies close to the Nyquist limit [8].

A. Staggered-grid Finite difference formulation

The first-order velocity-stress equations of viscoelastic wave propagation are given by

$$\begin{cases} \dfrac{\partial \sigma_{xy}}{\partial t} = \mu \dfrac{\tau_\varepsilon^s}{\tau_\sigma}(\dfrac{\partial v_x}{\partial y} + \dfrac{\partial v_y}{\partial x}) + \gamma_{xy} \\ \dfrac{\partial \sigma_{xx}}{\partial t} = \pi \dfrac{\tau_\varepsilon^p}{\tau_\sigma}(\dfrac{\partial v_x}{\partial x} + \dfrac{\partial v_y}{\partial y}) - 2\mu \dfrac{\tau_\varepsilon^s}{\tau_\sigma} \cdot \dfrac{\partial v_y}{\partial y} + \gamma_{xx} \\ \dfrac{\partial \sigma_{yy}}{\partial t} = \pi \dfrac{\tau_\varepsilon^p}{\tau_\sigma}(\dfrac{\partial v_x}{\partial x} + \dfrac{\partial v_y}{\partial y}) - 2\mu \dfrac{\tau_\varepsilon^s}{\tau_\sigma} \cdot \dfrac{\partial v_x}{\partial x} + \gamma_{yy} \\ \dfrac{\partial \gamma_{xx}}{\partial t} = -\dfrac{1}{\tau_\sigma}[\gamma_{xx} + \pi(\dfrac{\tau_\varepsilon^p}{\tau_\sigma} - 1)(\dfrac{\partial v_x}{\partial x} + \dfrac{\partial v_y}{\partial y}) - 2\mu(\dfrac{\tau_\varepsilon^s}{\tau_\sigma} - 1) \cdot \dfrac{\partial v_y}{\partial y}] \\ \dfrac{\partial \gamma_{yy}}{\partial t} = -\dfrac{1}{\tau_\sigma}[\gamma_{yy} + \pi(\dfrac{\tau_\varepsilon^p}{\tau_\sigma} - 1)(\dfrac{\partial v_x}{\partial x} + \dfrac{\partial v_y}{\partial y}) - 2\mu(\dfrac{\tau_\varepsilon^s}{\tau_\sigma} - 1) \cdot \dfrac{\partial v_x}{\partial x}] \\ \dfrac{\partial \gamma_{xy}}{\partial t} = -\dfrac{1}{\tau_\sigma}[\gamma_{xy} + \mu(\dfrac{\tau_\varepsilon^s}{\tau_\sigma} - 1)(\dfrac{\partial v_x}{\partial y} + \dfrac{\partial v_y}{\partial x})] \\ \dfrac{\partial v_x}{\partial t} = \dfrac{1}{\rho}[\dfrac{\partial \sigma_{xx}}{\partial x} + \dfrac{\partial \sigma_{xy}}{\partial y} + f_x] \\ \dfrac{\partial v_y}{\partial t} = \dfrac{1}{\rho}[\dfrac{\partial \sigma_{xy}}{\partial x} + \dfrac{\partial \sigma_{yy}}{\partial y} + f_y] \end{cases} \quad (1)$$

The moduli π, μ can be expressed as

$$\pi = v_{p_0}^2 \rho R^2 \sqrt{\frac{1}{1 + \dfrac{iw_0\tau_\sigma}{1 + iw_0\tau_\sigma}\tau_\varepsilon^p}}, \quad \mu = v_{s_0}^2 \rho R^2 \sqrt{\frac{1}{1 + \dfrac{iw_0\tau_\sigma}{1 + iw_0\tau_\sigma}\tau_\varepsilon^s}} \quad (2)$$

Where v_{P0}, v_{S0} denote the phase velocity at the centre frequency of the source(w_0) for P- and S-waves, respectively. The symbol R denotes the real part of the complex variable. The constants of stress relaxation times for both P-and S-waves, can be calculated by quality factor Q and angular frequency w.

A second-order centered difference scheme is applied to approximate the time derivatives, and a fourth-order staggered scheme with centered differences to approximate the spatial derivatives. From Eqs.1, for examples, $\sigma_{xy}^{n^+}$, $r_{xy}^{n^+}$ and v_x^{n+1} can be approximated as:

$$\sigma_{xy}^{n^+}(i^+,j^+)=\sigma_{xy}^{n^-}(i^+,j^+)+\frac{\tau\cdot\pi(i^+,j^+)\tau_\varepsilon^n(i^+,j^+)}{\tau_\sigma(i^+,j^+)}\left(\frac{\partial v_x^n(i^+,j)}{\partial y}+\frac{\partial v_y^n(i,j^+)}{\partial x}\right)+\frac{\tau}{2}\left(r_{xy}^{n^-}(i,j)+r_{xy}^{n^+}(i,j)\right) \quad (5)$$

$$r_{xy}^{n^+}(i^+,j^+)=\left(1+\frac{\tau}{2\tau_\sigma(i^+,j^+)}\right)^{-1}\left(\left(1-\frac{\tau}{2\tau_\sigma(i^+,j^+)}\right)\cdot r_{xy}^{n^-}(i^+,j^+)-\frac{\tau\cdot\mu(i^+,j^+)}{\tau_\sigma(i^+,j^+)}\left(\frac{\tau_\varepsilon^s(i^+,j^+)}{\tau_\sigma(i^+,j^+)}-1\right)\cdot\left(\frac{\partial v_x^n(i^+,j)}{\partial y}+\frac{\partial v_y^n(i,j^+)}{\partial x}\right)\right) \quad (6)$$

$$v_x^{n+1}(i^+,j)=v_x^n(i^+,j)+\frac{\tau}{\rho(i^+,j)}\left(\frac{\partial\sigma_{xy}^{n^+}(i^+,j^+)}{\partial y}+\frac{\partial\sigma_{xx}^{n^+}(i,j)}{\partial x}+f_x^{n^+}\right) \quad (7)$$

Where *i, j, k, n* are the indices for the three spatial directions and time, respectively. τ denotes the size of a timestep.

The discretion of the spatial differential operator, for example,is given:

$$\frac{\partial v(i^+,j)}{\partial x}=\frac{1}{l}\left(c1\cdot(v(i+1,j)-v(i,j))+c2\cdot(v(i+2,j)-v(i-1,j))\right) \quad (8)$$

Where l denotes grid spacing and *c1, c2* denote the differential coefficients.

B. PML absorbing boundary condition

In order to simulate an unbounded medium, an absorbing boundary condition (ABC) must be implemented to truncate the computational domain in numerical algorithms. Since the highly effective perfectly matched layer (PML) method [13] for electromagnetic waves was proposed, the PML has been widely used for finite-difference and finite-element methods [14]. The effectiveness of the PML for elastic waves in solids and the zero reflections from PML to the regular elastic medium were proved [15].

III. Implementation on GPUs

A. Overview

To compute the staggered-grid FD and PML equations on graphics hardware, we divide the geophysical model into regions represented as textures with the input and output parameters. Fig.1 shows the division of a 2D model into a collection of regions. Two sets of alternating textures binding with a FBO are used to represent the parameters for computing frame at time t and t+1 respectively. All the other input-only variables are stored similarly in 2D textures. To update FBO textures at each time step, we render quadrilaterals (regions) mapped with the textures in order, so that a fragment program running at each texel computes the FD equations. After each frame, FBO attachments are swapped and the latest updated textures are then used as inputs for the next simulation step.

B. Domain decomposition

The computational domain is divided into an interior region and PML regions. The outgoing waves are absorbed by the PML via high attenuation. For 2D model, the size of the interior region is N1*N2, denoting the interior part of the whole textures. PML region covers 8 blocks and can be divides three types based on the computing process: PML1, processing absorbing condition both in x and z direction, with the size NBC*NBC; PML2 only in z direction and PML3 only in x direction. The vertex processors will transform on computational domains and the rasterizer will determine all the pixels in the output buffer they cover and generate fragments.

In our approach, velocity and stress fields are staggered in spatial and time domain. Therefore, the velocity field is updated after the stress field is updated and the priority of subdomains processing is interior region, PML2 (3), PML3 (2), PML1.

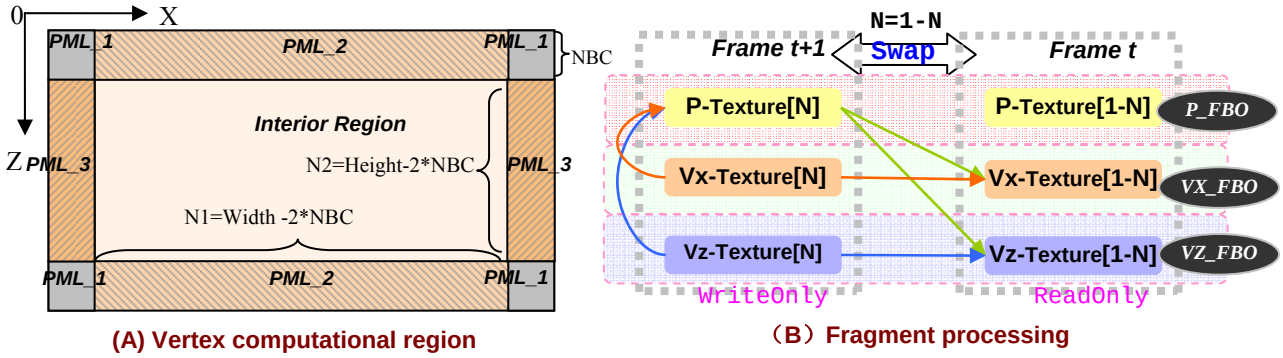


Fig.1. Implementation of staggered grid FD method with PML conditions on GPUs. Vertex processing switches computing regions and render quadrilaterals (regions) mapped to textures; fragment program run at each texel for computing frame at time t and t+1. FBO attachments are swapped and the latest updated textures are then used as inputs for the next simulation step (loops go on).

C. Data representation

For 2D viscoelastic FD schemes, computation on interior region at each frame is to get P_X, P_Z, P_{XZ}, V_X, V_Z, R_X, R_Z and R_{XZ}; computation on PML regions needs two extra components for P_X, P_Z, P_{XZ}, V_X or V_Z on each of 4 absorbing boundaries. Considering the time t and t+1, there are actually (8+5*2*4)*2 = 96 arrays storing these components to run in the whole process.

In order to reduce to the number of textures and to switch computational domain quickly, 16 textures are used to store interior stress-velocity parameters, called *whole- texture* with the size width*height, width and height denoting grid-point numbers in x and z direction respectively. It equals the size of the rendering viewport. The input seismic parameters Vp, Vs, Qp, Qs, ρ are also stored in *whole- texture*. 5*2 extra textures, called *x-direction-texture*, are used to express all velocity and stress parameters in x absorbing direction; and the size of each texture is (2*NBC, height), NBC defining the gridpoints of the absorbing boundary. In our work, NBC is 10. Simultaneously, there are 5*2 extra arrays in z absorbing direction, called *z-direction-texture*, and the size is (width, 2*NBC).There are actually another 20 arrays at time t+1.

From above analysis, the memory requirements *MEMREq* for 2D viscoelastic modelling is obtained by,

$$MEMREg = (N^2 * C + N * NBC * D) * Fz / 1024^2 \quad (11)$$

where the total number *C* of the whole-texture is 21 and the total number *D* of x-and z- direction- texture is 40. *Fz* of float size equals 4 bytes.

TABLE I
MEMORY REQUIREMENTS FOR 2D VISCOELASTIC MODEL

Gridsize(N^2)	Memory Requirements
512^2	22.56M
1024^2	87.12 M
2048^2	342.25 M
4096^2	1356.50 M

Table1 shows 2D viscoelastic model which is only less than 2048*2048 gridsizes can be supported on current GPUs directly, since graphics video memory size is about 256-512M in general.

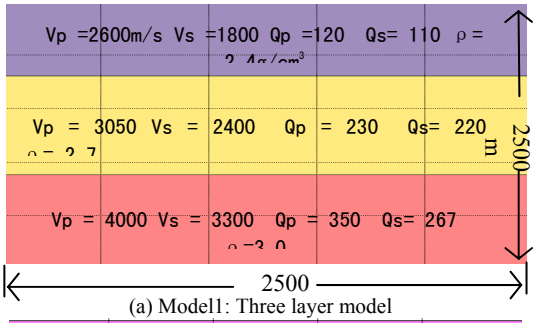
D. Fragment processing

The results of each step are written to a two-dimensional memory location called framebuffer. But the characteristics of the GPU prohibit the usage of the buffer for reading and writing operations at the same time. Fortunately, the OpenGL extension FrameBuffer Object (FBO) allows reusing the framebuffer content as a texture for further processing in a subsequent rendering pass and the "ping-pong" method allows two buffers swap their role after each write operation [3]. In order to generate and store intermediate results for each simulating step on GPUs, our strategy is to bind each texture that is rendered to onto one attachment on an assigned FBO; two textures binding on the same FBO are used to represent one stress-velocity component at time t and t+1, respectively. After each frame, FBO attachments are swapped and the latest updated textures are then used as inputs for the next simulation step while N = 1- N (N = 0, 1).

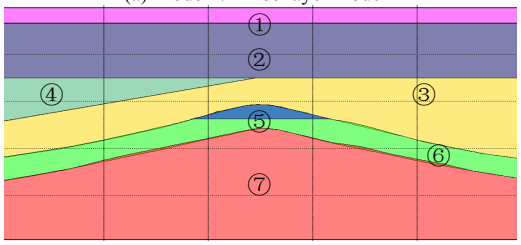
IV. NUMERICAL EXAMPLES

We have experimented with our GPU accelerated method implemented using C++/OpenGL and Cg shader language on a PC with Intel Core(TM)2 Duo E7400 2.8GHz, 2G main memory. The graphics chip is NVidia GeForce 9800GT with 512M video memory and 550MHz core frequency. The OS is Windows XP. All the results related to the GPU are based on 32-bit single-precision computation of the rendering pipeline. For comparison, we have also implemented a software version of the seismic simulation using single-precision floating point, and we have measured their performance on the same machine.

Two samples are used. One is a three-layer model (model1) and another is similar to Kelly model [14] (model2). The source is located at the top of the medium (Fig.2) and propagates downwards ricker wave with a dominant frequency of approximately 50 Hz.



(a) Model1: Three layer model



①Vp=1200	Vs=800	Qp=80	Qs=60	ρ =1.4
②Vp=2440	Vs=1330	Qp=100	Qs=90	ρ =2.2
③Vp=3210	Vs=2389	Qp=219	Qs=190	ρ =2.3
④Vp=3500	Vs=2173	Qp=200	Qs=170	ρ =2.5
⑤Vp=3225	Vs=1800	Qp=294	Qs=220	ρ =2.5
⑥Vp=3822	Vs=2980	Qp=490	Qs=340	ρ =2.6
⑦Vp=4785	Vs=3800	Qp=450	Qs=400	ρ =3.4

(b) Model2: Modified Kelly model

Fig.2. two test models: both are the same size of 2500m*2500m.

The computation runs 2000 steps on both GPU and CPU, respectively. The results for the two models are plotted in Fig.3. The time of viscoelastic wave simulation is defined as a function of the grid size. In order to compare the performance, GPU based method do not transfer the velocity-stress field or the parameters between the main memory and the graphics memory. The time spent on advecting and rendering the regions is negligible with respect to the simulation. The time of staggered-grid FD method is only dependent on the grid size, independent on geology structure complexity, which is shown in both GPU (Fig. 3a) and CPU (Fig.3b) based implementation.

The performance of the GPU based implementation increases significantly with the increasing number of the gridpoints. The speedup factor, which is defined as the ratio of CPU's time to GPU's time on the same grid size, increases from about 2 (128^2) to 50 (2048^2). That is mainly due to the advantages of GPU's parallel

computing, and the staggered grid FD method is amenable to an implementation by the rendering pipeline. For the simulation only runs in several minutes, geophysicists can use the new method to visualize the each frame of the computing process in real time, which helps to analysis the wave propagations easily.

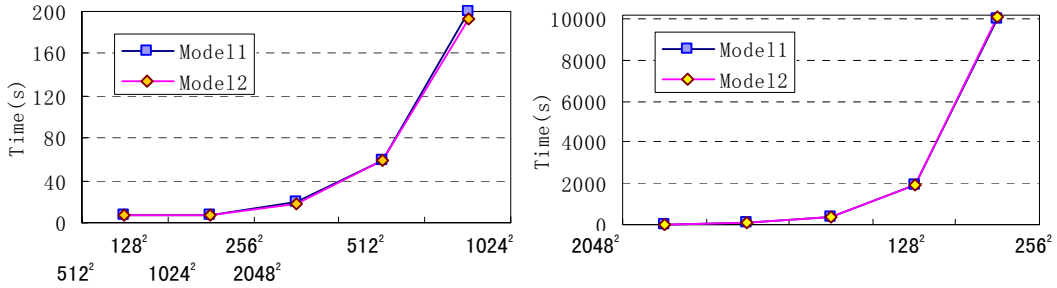


Fig.3. Comparative Analysis of GPU and CPU based staggered-grid FD method: (a) GPU based method using model1 and model2, (b) CPU based method using model1 and model2.

On the other hand, GPU based method improves more requirements of graphics card to use the programmable shaders. Now, only the latest nVidia or ATI card (Geoforce 6 above) can support the method. In additon, the available grid size is limit by the GPU texture maximum size. Now, the texture maximum size is 4096^2, so difference grid will not surpass the maximum size and the simulation accuracy is restricted. Graphics video memory size is about 256-512M in general, which limit the computing resources and only less than 2048^2 grid sizes can be supported directly.

V. Conclusion

We have presented a GPU accelerated staggered-grid finite difference method to simulate the seismic propagation in 2D viscoelastic media model with complex earth structure geometry. The physical model is divided into subdomains for PML absorbing conditions and the implementation is totally implemented using the current fragment and vertex processing stages on GPUs. The simulation program runs on Modern PCs with popular low-cost GPUs, yet it is much faster than CPU-based simulation. Our experimental results have shown that the GPU version is significantly faster (for example, 3-50 times faster) for large grid sizes of 2D Models. The method makes it possible to simulate realtime complex seismic propagation in high resolution of 2048*2048 gridsizes on low-cost PCs.

Acknowledgment

The work was supported by China Postdoctoral Foundation (No.20090450466) and the National Key Basic Research and Development (973) Program of China (Grant No. 2010CB732002, 2009CB724601, 2005CB221505).

References

[1] T. Bohlen, “Parallel 3-D viscoelastic finite difference seismic modelling,” *Computers & Geosciences*, vol. 28, no.8, pp.887–899, October 2002

[2] W. Sun, J.W. Shu, and W. Zheng, “Parallel Seismic Propagation Simulation in Anisotropic Media by Irregular Grids Finite Difference Method on PC Cluster,” Gervasi et al. (Eds.): *ICCSA 2005*, LNCS 3483, 2005, pp.762 – 771

[3] J. Owens, D. Luebke, N. Govindaraju, M. Harris, J. Kruger, A. Lefohn and T. Purcell, “ A survey of general-purpose computation on graphics hardware,” In *Eurographics 2005*, State of the Art Reports, pp. 21-51.

[4] H. Mark, “Mapping Computational Concepts to CPUs,” *GPU Gems2, Addison-Wesley*. NVIDIA, 2005, Chapter 47, pp.493-508

[5] J. Virieux, “P-SV wave propagation in heterogeneous media: velocity-stress finite-difference method,” *Geophysics*, vol. 51 issue 4, 1986, pp.889–901.

[6] K.R Kelly, “Numerical study of love wave propagation,” *Geophysics*, vol.48, issue 7, 1983, pp.833–853

[7] R. W. Graves, “Simulating seismic wave propagation in 3D elastic media using staggered grid finite differences,” *Bull. Seism. soc. Am.*, vol. 86, 1996, pp.1091–1106

[8] G. Kneib and C. Kerner, “Accurate and efficient seismic modeling in random media,” *Geophysics*, vol. 58, issue 4, 1993, pp.576–588

[9] S.M. Day and J.B. Minster, “ Numerical simulation of wavefields using a Pade approximant method,” *Geophysical Journal of the Royal Astronomical Society*, vol. 78, 1984, pp.105–118

[10] H. Emmerich, M. Korn, “Incorporation of attenuationinto time-domain computations of seismic wave fields”. *Geophysics*, vol. 52, issue 9, 1987, pp.1252–1264.

[11] H.P. Liu, D.L. Anderson and H. Kanamori, “Velocity dispersion due to anelasticity: implications for seismology and mantle composition,” *Geophysical Journal of the Royal Astronomical Society*, vol. 47, 1976, pp.41–58

[12] J.O. A , Robertsson, J.O. Blanch, and W.W. Symes, “Viscoelastic finite-difference modeling,” *Geophysics*, vol. 59, issue 9, 1994, pp.1444–1456.

[13] J.P. Berenger, “A perfectly matched layer for the absorption of electromagnetic waves,” *Journal of Computational Physics,* vol. 114, 1994, pp.185–200

[14] F. D. Hastings, J. B. Schneider, and S. L. Broschat, “Application of the perfectly matched layer (PML): absorbing boundary condition to elastic wave propagation,” *J. Acoust. Soc. Am*, vol. 100, 1996 , pp.3061–3069

[15] W.C. Chew and Q.H. Liu, “Perfectly matched layers for elastodynamics: a new absorbing boundary condition,” *Journal of Computational Acoustics*, vol. 4, 1996, pp.72–79

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 133-135

Explore Diigo for the First Time Applying to Learning Hitting the Target

Bo Yuan, and Bin Tan
School of Information Science and Communication, Jinggangshan University,Ji-An Jiangxi ,China
E-mail: bobi612@163.com

***Abstract*—Diigo is that Web2.0 times being pushed the pragmatism network implement bestowing favor on fully , is able to carry out mark , the classed implement on network information , group group and other material. And the main body of a book have analysed the Diigo function and have handled technological process detailedly, have introduced that it applies to learning hitting the target.**

***Index Terms*—Diigo; Web2.0; Socialization network implement**

Knowledge-base economy times competition embodies the competition being information resources still more , face vast information of Internet ocean, will have become once personal key being hit by being able to base self on sure ground or not in competition the information how effectively to distinguish , to gain , to store , manage these networks[1]. As network brand-new one socialization implement, individuality, high-effect characteristic precipitated by Diigo all has provided a effective approach to personal studying and the ability to innovate culture.

I. DIIGO SUMMARIZES

the Diigo team concentrating efforts on be that the consumer provides FOAK and pragmatic online services ". "This name comes from Diigo in "Digest of Internet Information, Groups and Other stuff. "(Internet information digest (annotation and commentary) , group group and their his) abbreviation. [2]

Diigo is that a "socialization annotates and comments on " website. Have been tied in wedlock the socialization bookmark , excerpt , annotation and commentary , label, and full text search, sharing and interactive function easy to use, not only Diigo be that knowledge shares platform as well as socializing a individual big and powerful of workers implement , but also entire Web will become a medium write , participating in and moving mutually with the development of it. Earth , Google Home) all arrange Google on the afterwards! It is a social nature of online one function very big and powerful bookmark implement , it is unlike the centre warehouse , good Ei net , drawer etc. is similar to website but, fairly not bad annotation Web page, shares collect, data assemble arrange and so on. In famous CNET of in the world magazine Internet 10 big "the well liked go into implements " judging out in 2006, Diigo row fourth! Four implements (Google Scholar , Google Book , Google Earth , Google Home) that Google is chosen as are all arranged on the afterwards!

II. MAIN DIIGO FUNCTION AND CHARACTERISTIC

A. Annotation serves Diigo on the net providing social nature(F1) , this is one of main Diigo characteristic.

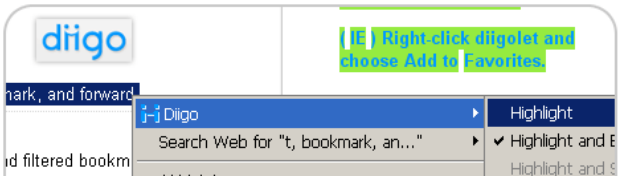


Figure 1. Online Notes

The user when glances over the homepage, easy to lose in the massive network information, thus causes had glanced over the information not to be able the effective retention in the middle of the cerebrum. In the polling message process, the idea which or the opinion produces is unable promptly the record retention. And, regarding some value strong information, some users cannot discover and understand its connotation. In view of network user existence's these demands, through Diigo Highlight and Sticky Note (Gao Liang and labelling) the function, can very good be able to realize. Said with the CNET words: "if in the middle of a 3000 character's article, you only needs a few words merely, then can know in the website makes the symbol is a how practical matter" . Diigo lets Web page become a book , becomes the distance between having shortened "write " , thereby immense "seeing that" and "telling" really. [4]

The consumer makes use of the Diigo Highlight and Sticky Note , location adds bright some Duan characters, and the idea adding mark , writing self down arbitrarily in Web page. One process , your idea are not bad by this the permanency preservation covers a page as with a net here going ahead. That is ,no matter when and where, you all can see what had been composed Gao Liang and mark as long as the computer links a network. The consumer can set up mark do mind enjoying and , be set up for private content, can only be seen that by consumer individual, the consumer is therefore not bad free rein's announce impressions to some things , take the judgement thinking deeply that the ability draws exercise thereby. But those shared mark, the whole world Diigo consumer is all not bad see and participate in exchange, have strengthened network interactive nature thereby very good. Face a computer especially specifically for every day , require that the large amount of the people reading , arranging ,

studying to inspect with the Diigo Highlight function may assist the data accomplishing a great quantity to collect with HSKP.

B. Network-rization bookmark(F2)convenient and rapid

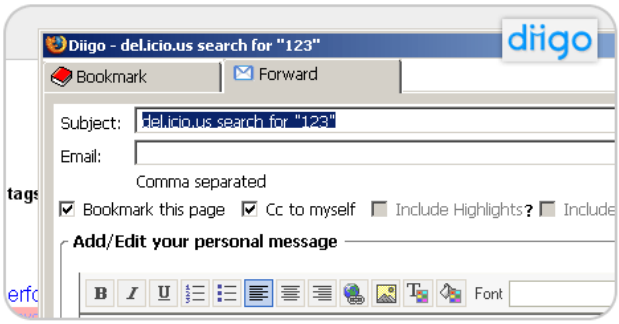


Figure 2. Bookmark Network

Other network bookmark is compared with each other and , Diigo owns the Web page especially big and powerful montage function , it is not only not bad the permanency field has preserved the characters that the consumer collects , photograph information , up to the minute Diigo have increased by to video frequency and Flash to cut out and collect a function. Be rare being more, the consumer institute shares the share and reconnaissance function quilt that information can by Diigo collected , marked by you other. [3]

C. Plentiful visitor platform big and powerful

Be just your any annotation Web page being able to immediately transform into the plentiful visitor announces. The form being able to direct it on immediately your content being interested in passing the right key as long as you have assembled diigo , your to have a wide knowledge of a visitor is announced & not; ,Diigo supports plentiful various type visitor , includes: WordPress blog , blogger blog , live journal blog , Typepad Blog , Movable Type Blog , Windows Live Spaces , Drupal Blog. Certainly, you also can announce them on the plentiful visitor that Diigo certainly brings along , you also can add self plentiful visitor at the same time , your plentiful visitor being interested in is announced on your own plentiful visitor.

D. The unique group forms a forum

The Diigo group group forum allows the consumer topic of conversation specifically for self is interested in to establish the sign discussing that the group , the kind setting up a group draw not only, discussion attracting more people thereby participating in a group, the consumer returns back to the classification being able to be based on a group or the function making use of a group to search for comes to find the topic of conversation being interested in, and carry out restoration on the person. Another , Diigo have provided one kind of any member in being more nimbly that way , it allow a group certainly fixing a sign , the average member that the group forms at this time has accommodated oneself to same as the director , have owned the liberty being more vast thereby in being able to participate in highly autonomously to the forum, bring space into play.

Diigo still has more function , Bi Ru: Customization search for an implement and unique content searches for a menu. The Diigo group forum function, the allowable Diigo forum user of group initiates and participates in to discussing middle. Diigo has provided a aggregation warehouse , not bad at any time renewal and so on.

As always with a conversion to PDF, authors should *very carefully* check a printed copy.

III. DESIGNING MENTALITY OF THE SYSTEM

First, enter www.diigo.com , strike join diigo, self information, accomplishes last strike sign in , imports the consumer and the password , goes in ashore according to demanding to register. Its effect (pal , share) except being able to look like space , page of face being able to collect self liking it or often going to, can edit page of face, if Gao Liang display , short comment, can add mark moreover at the same time. And then, time be loaded with diigo installation arrives at a computer: Diggo website home page most underneath "tool" item, in page of face opening up, in striking the browser as follows, pursuing the what be shown button time be loaded with Diggo implement strip(F3), and assembling as the plug in connector to you.

Figure 3. Diggo Toolbar

The Microsoft IE browser is still red fox , finishes assembling a hereafter, need the weight open once (be just switching with the browser , open up) again again, this moment having appeared on your browser navigation fence pursue the what be shown Diggo blip as follows , can begin to have used Diggo now. (F6)

Figure 4. Diggo logo

Assemble the day afer tomorrow , turn on any website, only be capable to do the content demonstrating on self browser , demonstrating as long as the mouse bid is only capable to do on readjusting oneself to a certain extent to the hubble-bubble review so long as somebody discusses to website. [5]

IV. USE DIIGO TO EDUCATE THE TECHNOLOGY SPECIAL FIELD COLLEGE GRADUATE TO LEARN

The new exploring with Internet new technique idea suggesting that , that ken , Diigo having entered people move towards the more and more many service with web2.0 characteristic is a representative among them. Diigo has provided a rich society exchange platform to our person who provides a useful implement simultaneous to a large stock of information also. Believe

that studying being going to be to us provides very big help in studying if we lead Diigo into a graduate student. For this purpose, we have attempted to lead Diigo into the location education technology.

Have introduced first, Diigo to everybody , have let everybody know that implement very well, the trial recommending everybody uses that implement.

Secondly, think that broad to how to being able to arouse everybody pays close attention to, the author makes use of Diigo to have founded a group "who educates the special subject that technology special field hot spot problem assembles " for this purpose. And then, spicilegium has arranged a few hot topic of conversation shown solicitude for by the special field educating a technology , as well as developing the trend data such as: Educate technology six general trends ? t to hope that USA educates technology event top10 to wait a minute in 2007(F7).

Figure 5. Diggo thematic group

Everybody is hit by thirdly, a logon consumer in Diigo , is enters the group "who educates the special subject that technology special field hot spot problem assembles " founded by author , is joins that set , is is in progress to data collected by author to browse and then. In this special subject, the author information to self feeling interest has made Gao Liang display , has assumed and a little annotation and review to here. Everybody has labeled sum annotation when browsing a data very easy to be able to see priority done by author right away , therefore, everybody has shown self viewpoint out while reading , has announced various different view , has perfected and special subject content at the same time also unceasingly to having added necessary complement , having enriched further in the set while everybody discovers new relevance data information.

Fourth, has exchanged for making everybody be in progress more , the author has opened group forum function. In group forum, the author has built a "prospect problem forum educating a technology to obtain employment" with member of group from fixing the denotational form , everybody has carried out restoration on this problem announcing self view view. At the same time, some schoolmates have cared for also with regard to self or the problem feeling uncertain has built here the pertinent forum, not only such give bigger free space to schoolmates , but also the group who uses us is especially unique.

Fifth ,The author has suggested that everybody establishes the pertinent special subject , provides schoolmates with studying together according to self go into direction fifth, then , many schoolmates has kept mark all in others's special subject, have done a little comment. Schoolmates have expanded knowledge face like this even , have made to found person self has gained new complement resource.

Sixth, has done a summary by that a period of time sigmatism , the author using the Diigo valuation to everybody and using Diigo which to have got rise in the field of the day afer tomorrow.

A. Overall evaluation

Participating in this time of experiment's is 16 college graduates who educate a technology altogether , that the valuation to Diigo is allotted for Three Estates of the Realm , difference is: Be fond of being just like, and not good. Pursue what be shown finally as follows, regarding it being good among them is 13 people , accounts for 81%; Regarding it being the same is 2 people , accounts for 13%; Think that not good 1 people, accounts for 6%(F8).

Figure 6. Diigo divided into three levels of evaluation

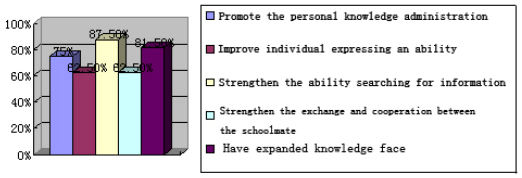


Figure 7. Learning Evaluation

A. The effect studying is appraised[6]

The author will use the Diigo queen to be five principal aspect to the partition studying the effect effect here,will pursues what be shown as follows(F9).

REFERENCES

[1] The Chinese interconnection network develops status counting to inquire into http: //Www.cnnic.net.cn. 2008-2-12.

[2] Diigo government website www.diigo.com.

[3] http://williamfan.blogbus.com/logs/3876551.html.

[4] Li Ying , Li thicken , explore for the first time the Diigo application in teaching. Shanghai normal university, 2007.

[5] Reformation , baysalt lecture Li being thickened by , studying management system Moole and the new course, March 29 , 2007.

[6] Tang Yue Ming , Wu Jing Song , Li Hong An county , Diigo application in studying in the graduate student are probed into.

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 136-139

Agent-Based Service-Oriented Dynamic Integration ERP Architecture

Zuoming Huang[1], Jincheng Zhang[2], and Qiushi Cong[2]
[1]Audit information engineering key laboratory, Nanjing Audit University, Nanjing, PR China
zuomingh@nau.edu.cn
[2] Nanjing Audit University,Nanjing, PR China
zjc@nau.edu.cn

***Abstract*—The manufacturing enterprises of the 21st century are facing an environment where markets are frequently shifting, new technologies are continuously emerging, and competition is globally increasing. This paper proposes a service-oriented dynamic mobile integration architecture for batch customization ERP production. The combination of Web services and mobile agents provides a promising computing paradigm for efficient service selection and integration of inter-organization business processes Batch customization satisfies the personalized demand of users by production cost and time. In this way, development cost and time can be reduced and the flexibility and adaptability of ERP system can be enhanced. A prototype system is designed to demonstrate that MASOAERP can custom the collaborative ERP software under the opening, dynamic and changeable environment, realize the business enterprise alliance resources service of the dynamic state integration.**

***Index Terms*—ERP Architecture; Web Services; Mobile Agent**

I. INTRODUCTION

The manufacturing enterprises of the 21st century are facing an environment where markets are frequently shifting, new technologies are continuously emerging, and competition is globally increasing. The rapidly changing needs and opportunities of today's global market require unprecedented levels of interoperability to integrate diverse information systems to share knowledge and collaborate among organizations. The traditional ERP can only solve the problem of business process integration inside of the enterprise, but for the resource integration and knowledge sharing among the alliance enterprises it has severe limitation. In the future, the ERP system production under the Internet computing environment is based on a platform, which is fruitful in ERP service resource and has open, dynamic and changeable software architecture. The challenge provides a scarce opportunity for our studying collaborative ERP system theory, method and technology, the study emphasis of ERP system software is changing from C/S to new system platform. The constructing technology of ERP system based on new middleware platform shall make a breakthrough. In the project supported propose a mobile agent-based service-oriented integration architecture for batch customization ERP production, construct service-oriented dynamic architecture ERP model under the internet environment, and discuss the customization production implement of collaborative ERP system.

II. RELATED WORK

Currently the technologies of enterprise resource integrations, such as CORBA,XML, Agent and UDDI, are insufficient for the support of the dynamic enterprise alliance resource integration, especially for the support of the inhomogeneous cross-platform information system are insufficient and cannot satisfy the customized production of cooperative ERP based on the Internet computation. New technologies based on single Web, such as Web services and mobile Agent, also cannot fulfill the needs of the dynamic enterprise alliance resource integration[1]. Reference[1] presented the cooperative model of the enterprise inner work-flow, in the proposed model, agent-based technology provides the workflow coordination at both inter- and intra-enterprise levels while Web service-based technology provides infrastructures for messaging, service description and workflow enactment. Reference[2] designed and developed the distributed service discovery and negotiating system for the implementation of B2B E-business with the web service and intelligent agent technology. Reference[3] presented a mobile agent platform based on the integration of the mobile agent computing paradigm with Web Services, WebMages (Web-based Mobile Agent System), which providing a lightweight agent run-time environment inside a web container, thus adding agent functionality to existing Web Servers. Reference[4] presented an integrated framework based on multi-agent collaboration and case-based reasoning that can resolve collaboration issues in the supply chain. Reference[5] indicated the deficiencies of the web service, namely insufficient for the abundant representation methods and the attractive strategy, and presented intelligent service and web service autonomy mechanism based on semantic cooperation. References [6]-[8] researched the web based information system integration. On the other hand, many researchers have researched the web based ERP system. Reference[9] presented and developed the web based ERP system, which solved the problem from the simple office automation to the complicated supply chain management for the management of the real world operation process. Chinese scholars have done some work also, for example,

Reference[10] analyzed the evolution trend of ERP, proposed the element model of the service oriented architecture and the ERP system integration strategy. Based on the traditional ERP architecture, it also proposed a kind of service oriented ERP integration model. The author also made in-depth study on the ERP production model, reference[11] proposed the ERP batch customized development model based on version management driven by enterprise requirement and exactly pointed out that the batch customization is the direction for the ERP production mode. Enlightened by the references above, through studying about the Service-Oriented Architecture and mobile agent technology, we propose the service oriented collaborative ERP customization based on mobile agent. The aim of our study is to provide a new implement project for the batch customization of collaborative ERP product under the internet computing environment by utilizing the application resource intra- and extra- enterprise to dynamically integrate and aggregate the service component according to the enterprise demand and the dynamic model.

III. AGENT-BASED SERVICE-ORIENTED ARCHITECTURE COLLABORATIVE ERP SYSTEM

The "description, publishing, finding, binding and interaction" enterprise service architecture expressed by SOA and the "establishment, execution, freezing, migration and extinction" of software lifecycle of mobile agents fulfill the characteristics of the application dynamic service integration in the open Internet environment. Mobile agent-based, and service-oriented collaborative ERP software solution has to meet with following requirements: 1) The whole collaborative ERP software system should be built by a series of "rough grain" component, which are mobile agent-based enterprise business service component, inner-enterprise application services and inter-enterprise service cooperative agents, and should be loosely coupled. The functions are expressed in the form of rough grain services, and each kind of service can express its business value clearly. The rough grain service providing form makes the large-scale software service possible, and provides the agility of construction and reconstruction of ERP systems in the open, dynamic and uncertain environment of Internet. 2) Self-description of application service units of each layer can be made in a standard way. The descriptions should be understandable by enterprise cooperative agents, and processable by computers. Dynamic publishing, finding, matching, binding and management services can only be carried out according to these descriptions. 3) Lifecycle management systems of enterprise system service components, in charge of the test, simulation, and optimized assembly and migration, should be provided. 4) Lifecycle management systems of enterprise services, in charge of dynamic creation, state migration, performance monitoring, and service termination, should be provided. 5) Management systems of enterprise applications, in charge of registration, access, controlling, and service relationship termination, etc. should be provided. 6) Enterprise application solution based on process integration should be provided. Flexible information sharing mechanism should be provided.

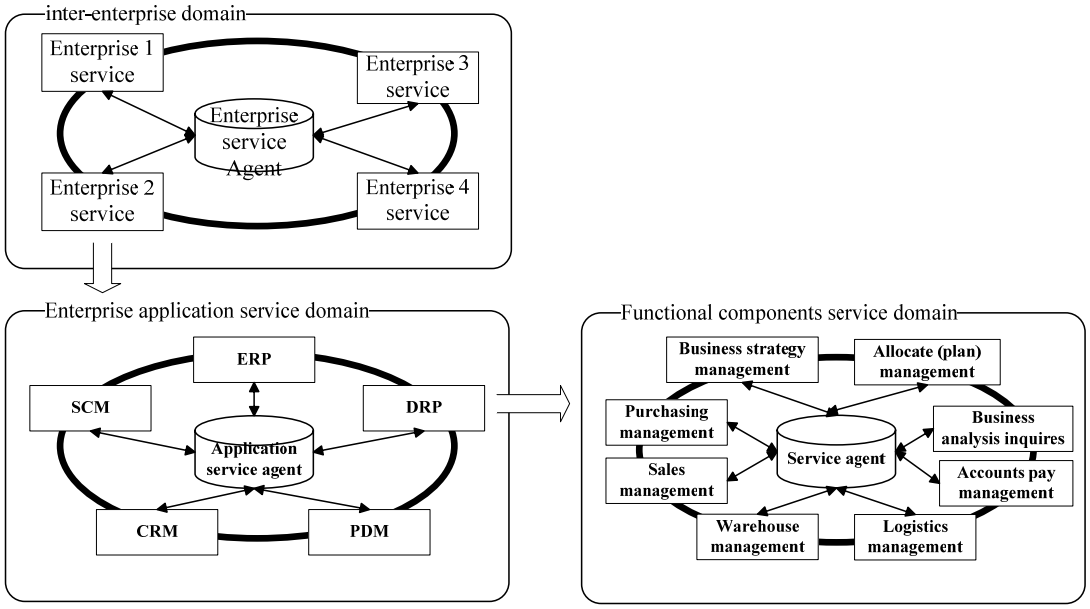


Figure 1. The agent-based service-oriented cooperative ERP software model

In the open, dynamic and uncertain environment of Internet, the architectural structure of the mobile agent-based, and synthesized service-oriented cooperative ERP software is depicted as figure 1. The services are divided into three layers -- inter-enterprise web services, intra-enterprise application services and composite application business component services. Accordingly, the three layers' services can be described by three domains, namely, inter-enterprise service domain, enterprise application service domain and business component service domain. The services are managed layeredly, and layers of the services indicate the constraint and

dependency between services. Inter-enterprise web services, intra-enterprise application services, and composite application business component services are of the first, second and third layers, respectively. Inter-enterprise services are top-layered, while the application business component services are bottom-layered. And higher-layered services are dependent on the implementation of lower-layered services.

In the enterprise service domain, On the one hand, business users can release their services to the enterprise service agents through their own self-description enterprise Web services, register enterprise Web services in the business services agents, publish enterprise service functions and interfaces. Enterprise service agents are responsible for the coordination and management work, such as information Enterprise service registration, positioning, bundling services and so on. On the other hand, when enterprises users request for service, the users register in the enterprise service agents first, after the success of registration, the enterprise service agent match the demand service by the petitioner, and sent the service information to the service requester, the service request Mobile Agent visit to the service providers, service providers and service bundle with the petitioner to provide them services. Here the following detailed descriptions：

(1) Enterprise Service Agent is a set of server groups.

(2) Mobile Agent-based enterprise Web services, strip heterogeneous characteristics of local resources through standard protocols to provide transparent services. Mobile Agent-based business services have reciprocal (P2P) authority, that is, to meet a variety of agreements and norms under the premise of equal use of resources. All Mobile Agent-based enterprise Web services, in accordance with business needs and business model of dynamic work flow services, share Resources, collaborate Mutually, serve each other and Complete a common task. When Mobile Agent-based Web services rely on the relationship between them, we can gain the control and Management through mutual consultation and cooperation.

(3) Mobile Agent-based business services are independently responsible for their own resource scheduling, and dynamically interact with the business services agents ,to make them understand real-time enterprise resource information services and dynamically change management decision-making , coordinate resources "compete" resources "exclusive" and other conflict resources, in accordance with specific circumstances.

(4) The Mobile Agent-based business services to other business services and information through the use of authority, can only be authorized by business service agent. Once two Mobile Agent-based business services establish connections, all matters can be completed by coordination between the two sides without intervention of enterprises services agent.

Depending on the given task’s requirement, the cooperation between two mobile agent-based enterprise service can takes the direct/active model or direct/active model. Using the direct/active non-terminate model, the service requester mobile agent sends active non-terminate service request to the service provider mobile agent, asking the provider to actively provide service to the requester when the conditions are met. When the provider has received the service request, it will make direct/active service promise to the requester. Immediately the conditions are met, the provider will provide service actively, and return the result accordingly. Using the direct/active terminate model, the service requester mobile agent sends active terminate service request to the service provider mobile agent, asking the provider to actively provide service to the requester when the conditions are met. When the provider has received the request, it will make active terminate service promise to the requester. Immediately the conditions are met, it will provide service actively, and return the results accordingly, and the cooperation is terminated.

In the enterprise application service domain, application service agent is established. The agent is responsible of the query, detection, binding coordination, and other enterprise application services. Enterprise applications (such as PDM, DRP, CRM, SCM) are encapsulated using mobile agent-based web services, and registered at application service agents. Implementing the Enterprise Application Integration (EAI), the working mechanism is similar to the inter-enterprise services, except it is confined intra-enterprise.

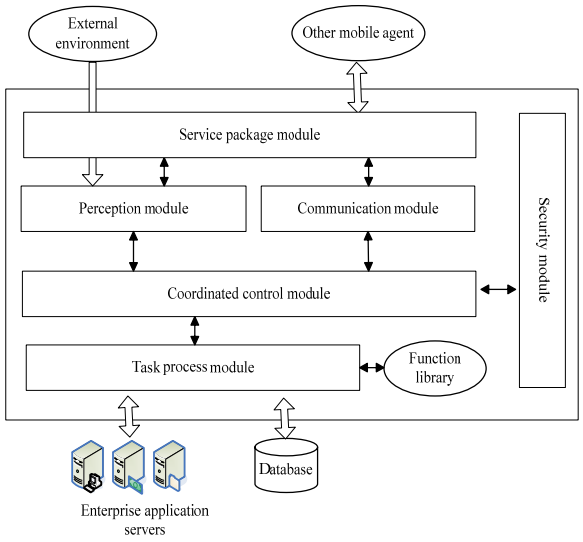


Figure 2 Based on MA function structure

Business component is the software component that can finish some enterprise business function, such as production plan management, storage management, order management and plant control. In the collaborative ERP software production based on mobile agent and service-oriented integration framework, the business component is the service component encapsulated by web service based on mobile agent technology and is different from the traditional business component. In the business component service domain, the agent based component function structure is composed of service encapsulation

module, communication module, apperception module, control and cooperation module, task processing module, which is shown as Figure2.

(1) Task processing module. Based on current knowledge and certain rules to judge the reasoning, the implementation of specific function or functions of the method. By this way, the task is accomplished.

(2) Communication module. It mainly responsible for the link between the Mobile Agent. It can not only transmit other Agent request / response signals to the task processing module, but also transmit collaboration / consultations messages of task processing module to other mobile Agent.

(3) Sensing module. It is used to sense the stimulation from external environment, and transmits the corresponding information to the control module. Firstly, control and coordination module do the filting and abstract process. The result is formatting meaningful symbols which corresponds the objects in external environment. Then it search corresponding method in the function to match. So the corresponding decision-making is got.

(4)Control and coordinated module. It is central role in all modules. All modules are managed by it. Through the coordination of the results, the final decision of the implementation of specific tasks, storage and management of all processing tasks and records the task of state.

(5) Service encapsulation module. It will encapsulate the Mobile Agent into independent service. It has the characteristics of SOA framework. And it can access through a standard interface so that the different components of services call each other. As independent service resources for the provision of in the network to other components, it can provide service to other components of a more flexible relationship between loosely coupled.

(6) Security module. To meet the business components with high reliability and security requirements, we add security module into application business components. The multi-level authentication and security mechanisms are increased. Business services component of the integrity of the information is ensured.

Ⅳ.CONCLUSIONS

Under the opening, dynamic and changeable Internet environment, the resource sharing and integration has become the challenge of the computer software technology. This paper presented the MASOAERP, collaborative ERP software model based on mobile Agent service, to construct dynamic collaborative ERP software model based on service under the opening Internet environment, and solve the ERP resource integration under the opening, dynamic and changeable Internet environment. In addition, we discuss the ERP product custom-built batch implementation under the opening Internet computing environment. The prototype system implemented according to the model will prove that the service-oriented integration framework can construct the collaborative ERP system, implement the resource service of the dynamic enterprise alliance, and construct the new web based collaborative ERP software production model.

ACKNOWLEDGEMENT

This research was supported by the Jiangsu Provincial Natural Science Foundation of China under Grant 07KJD520089, and supported by Nanjing Audit University in China under Grant NSK2007/A02.

REFERENCES

[1] Shuying Wang, Weiming Shen, Qi Hao,An agent-based Web service workflow model for inter-enterprise collaboration[J] . Expert Systems with Applications ,31 (2006) 787–799

[2] Raymond Y.K. Lau, Towards a web services and intelligent agents-based negotiation system for B2B eCommerce[J] . Electronic Commerce Research and Applications, 2006

[3] I.E. Foukarakis, A.I. Kostaridis, C.G. Biniaris, D.I. Kaklamani, I.S. Venieris, Webmages: An agent platform based on web services[J] . Computer Communications, 2006

[4] Hassan Artail, Elie Kahale, MAWS:Aplatform-independent framework for mobile agents usingWeb services[J] . Parallel Distrib. Comput. , 66 (2006) 428–443

[5] OhByung Kwon, Multi-agent system approach to context-aware coordinated web services under general market mechanism[J]. Decision Support Systems, 41 (2006) 380–399

[6] Ibrahim Cil, A new collaborative system framework based on a multiple perspective approach: InteliTeam[J]. Decision Support Systems, 39 (2005) 619– 641

[7] David Zhengwen Zhang, Anthony Ikechukwu Anosike, Ming Kim Lim, Oluwaremilekun Mowanuola Akanle， An agent-based approach for e-manufacturing and supply chain integration[J]. Computers & Industrial Engineering, 51 (2006) 343–360

[8] Shijia Gao, Huaiqing Wang, Yingfeng Wang, Wenqi Shen, Siubun Yeung, Web-service-agents-based family wealth management system[J]. Expert Systems with Applications, 29 (2005) 219–228

[9] C.D. Tarantilis , C.T. Kiranoudis, N.D. Theodorakopoulos， A Web-based ERP system for business services and supply chain management: Application to real-world process scheduling[J]. European Journal of Operational Research，(2006)

[10] Yao Yuan , Service oriented architecture enterprise resource planning system application model and integration strategy [J]. Computer integrated manufacturing system, 10(2006)1570-1577

[11] HUANG Zuo-ming1,2 , XUE Heng-xin2, GUI Liang-jun. ERP Batch Customized Development Model Based on Version Management Driven by Enterprise Requirement[J]. Computer science,3(2006)

[12] Petrie C, Bussler C. Service agents and virtual enterprises: a survey[J]. IEEE Internet Comput 2003，7(4)：68–78.

[13] Matskin M, Kungas P, Rao J, Sampson J, Peterson SA. Enabling Web services composition with software agents[C]. Proceedings of the ninth IASTED international conference on internet & multimedia systems & applications (IMSA 2005), 2005 August 15–17, Honolulu, Hawaii, USA.

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 140-143

Applications of Computer Control System in Automobile Comprehensive Performance Test

Zuming Xiao[1], Zhan Guo[1], Bin Tan[2], and Bing Zhu[2]
[1] Dept. of Physics, Jingdezhen Comprehensive College,Jingdezhen Jiangxi ,China
e-mail: teachergz@163.com
[2]School of Information Science and Communication, Jinggangshan University,Ji-An Jiangxi ,China
e-mail: jatanbin@163.com

***Abstract*—With the development of auto detection technology, applications of computer control system in automobile comprehensive performance test has developed into a new stage. Also the paper elaborates automobile comprehensive performance testing automatic control system's designing mentality, system composition, test process, and the characteristics of the system and function.**

***Index Terms*—Computer Control System, Detection of the cooling air flow，Sensors, Automobile Inspection Station，Automobile Comprehensive Performance Test，Speed parameter detection**

I. INTRODUCTION

Along with the development of modern transportation, car quantities is increase rapidly, traffic safety of vehicles, environmental protection, energy conservation and emission reduction has become a very serious problem. Automobile comprehensive performance testing is an important content of implementation of vehicle management for transportation departments at all levels. The current domestic examination of the automobile synthesize function is only based on manual inspection or half-automatic inspection. So how to realize the full-automatic inspection of the automobile synthesize function, that is quite attentioned by the corporation and academy who are very paying attention to the development of automobile inspection equipments at percent. Making use of modern electronics technique, computer and its network correspondence techniques, to develop the completely automatic vehicles overall performance inspecting line has become one inevitable trend.

II. SUMMARY OF SYSTEM

Automobile comprehensive performance test line is the testing system as a comprehensive use of modern detection technology, electronic technology, computer technology for automobile implementation non-dismantlement testing and diagnostic testing. It can diagnose and test all kinds of vehicle performance parameter indoors , and detect malfunction status providing reliable basis for the comprehensive and accurate evaluation of performance and technique of car's situation. Automobile comprehensive performance lines can undertake the test of vehicle performance, reliability and safety of economy, environmental management etc, and can undertake the test of vehicle maintenance quality inspection and heavy-duty truck technology condition, using the test evaluation on the lines of industrial PC and weighing, brake, speed and exhaust gas, lighting, sound level as combined testing equipment, then form digital signal by adjusting and gathering each station equipment's signal, and transmit to mainframe automatically for host store, and form a complete test report according to the test results, and can provide function of inquiring statistical data and analysis, so as to realize the automation management of motor vehicle safety testing and information sharing with superior departments and related business department[1].

III. DESIGNING MENTALITY OF THE SYSTEM

In order to succeed in the breakthrough of performance testing equipment assembly and online method, distributed computer test system will be applied in automobile comprehensive performance testing process. Then we will connect various testing equipments of respective operation modes and output data by computers, employing master computer and slave computer, to form a whole, then realize the concentrated automatic control and inspection the real-time data acquisition and processing, with the completion of each stage detection task rapidly, accurately and automatically[2].

System testing lines includes one master computer and three slave computer and a zero station machine, each slave computer is responsible for testing one operating station, and each operating station has at least two testing points. The first operating station mainly detects speed, tail gas and chassis. The second operating station mainly inspects braking performance by the axle load measurement platform and brake testboard.The third one operating station mainly measures engine performance, side slip, lighting, sound level, chassis dynamometer and other items.

System composition diagram is shown in graph 1.

Process of system inspection: Input and transmission of vehicles' information can be carried on in zero operating station, then single and system can collect judging and print detection results.

Master computer is command and control center of all lines, and it compile machine information transmitted from zero station machine to form command word, then notify slave computer, and simultaneously summarize, sort out, calculate observing data from slave computer.

AP-PROC-CS-10CN006

After the comprehensive performance testing, the final data is outputted to the printer, to print the report list and deposit in database for future query. It possesses at the same time high efficiency allowing several cars tested[3].

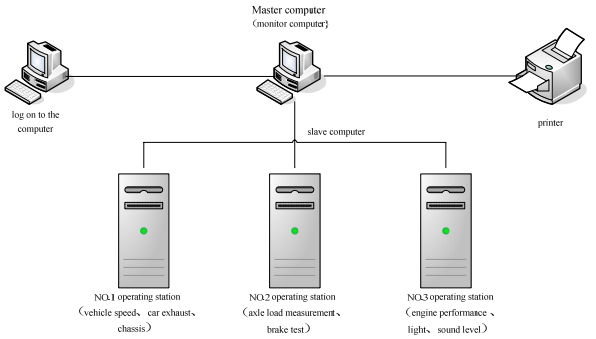


Figure 1. System composition diagram

IV. REALIZE THE FUNCTION

A. *log on the subsystem*

Logging system is the start of computer networking detection system, mainly for vehicle's information registration, to provide the necessary parameters for testing.

B. *on-line scheduling function*

After logging in vehicles, the vehicles scheduling and test of projects will be accomplished by the function of on-line scheduling for the log-in the outer-inspection automobile.

C. *auto detection function*

Auto detection function is the main part of the testing system, and it completes actual testing of all operations. After tested automobiles is logged in database through zero station machine, wagon examiners only pass vehicle testing lines' respective testing each operating station according to the instructions of information from LED. Then the drivers operate according to the tips on display, and vehicle detection of operation process can be completed through automatically detection system. At last, test report is automatically printed after testing[4].

D. *master controlling function*

Master controlling function are accomplished by the test line server 's master controlling IPC,usually by high-performance IPC,and total control function is vehicle detection's control and dispatch, data synthesis and storage,and test report is usually printed in the master controlling module.

E. *data management and retrieval function*

Data management and retrieval functions can possess statistics, retrieval of the vehicles' information from the master program's data files according to license plate number, serial number, vehicle owner's unit and types of projects.

F. *standard setting function*

Standards setting function the system can adapt to the state standards or change of local standards by modifying all testing items standard at any time.

G. *faults auto diagnosis and analysis function*

Computer networking detection system should be able to discover our own faults, and analyze and suggest the parts of the failure to help users for trouble-shooting. Namely, when the system is running, faults diagnosis' of mart CARDS, photoelectric switches, each sensor or circuit, etc should be suggested so as to ensure the normal running of the detection system[5].

V. HARDWARE COMPOSITIONS

A. *Master computer*

Master computer of the system, developed by China 486DX PC/industrial control machine,is the machine for steel box, double fan structure in the positive pressure, and is high-luminance shockproof, magnetic, high reliability, very suitable for industrial field application.Also,it adopts the machine double screen mode, letting the master controlling personnel understand fully at one glance of the screen display (The screen display all kinds of lines on the detection data, and vice screen online tests vehicles' lined up) . The master computer includes printer, one communication controller control eight and UPS power supply, etc[6].

B. *zero station machine*

Zero station machine is the AST-486PC machine, composited by a machine, a screen and UPS power. It is mainly responsible for appearance test results and input of vehicle’s natural information, and the natural information detection and transition of all data through RS232 standard communications mouth to PC.

C. *slave computer and outer equipment*

According to the test items the whole line is setted by three slave computer, each slave computer composited by 8031 single chip microcomputer, the LED display, photoelectric switches and red and green light, warning light and project input device as external devices. They are controlled and commanded by the slave computer. Then the slave computer directly control the detection equipment's operation according to the master computer manipulation, then remove inspection data and conveys to master computer in a finite format .

VI. THE SOFTWARE COMPOSITION

The device software includes two parts, namely, master computer software and， slave computer software is compiled by C language, and slave computer software uses assembly language.

A. *master computer software*

Master computer software includes the following modules: standard equipment, equipment calibration module, module testing data retrieval module, the inspection of main procedure module and data

communication modules[7]. Master control flowchart is shown in Figure 2.

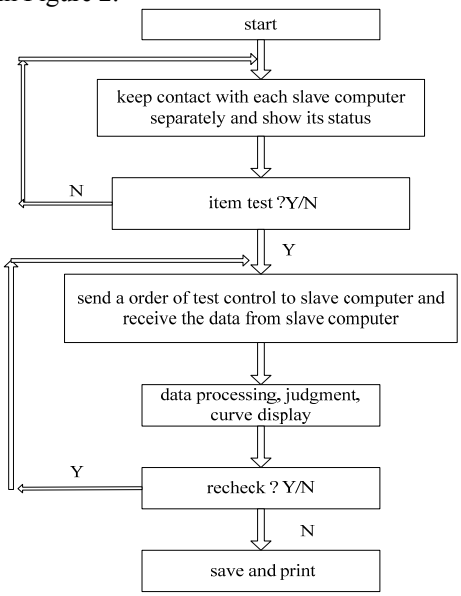


Figure 2. Master control program flowchart.

B. zero station machine software

The design has applied software under the support of the Windows operating platform software, including testing program modules, appearance inspection registration module, communication module, etc.

C. slave computer software

Slave computer software includes the following modules: control test, data acquisition and processing, sensor nonlinear processing, data communication, etc. Slave computer control flowchart is shown in Figure 3.

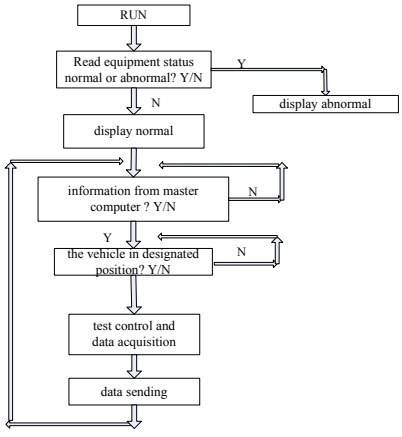


Figure 3. Slave computer flowchart.

VII. APPLICATION

Automobile comprehensive performance testing system consists of a number of subsystems, including comprehensive performance detection engine to become the focus of vehicle inspection and maintenance, for example to test automotive engine cooling air flow testing,

To overcome the car's internal heat and strong electromagnetic interference, the system selected high-temperature reflective photoelectric sensor to detect the signal through the heat shield signal lines for transmission to the larger shaping circuit, and then into the CPLD, through the CPLD programming, multi-frequency signal path for testing.

Through the wind tunnel experiment calibration, the establishment of wind speed, local wind and the relationship between the impeller rotation frequency:

$$V_i = g(f)$$
$$Q_i = \frac{1}{n} A\xi V_i, i = 1,2,.....,n \quad (1)$$

Formula:

V_i-adoption of the first i-impeller at the wind speed:

f - Impeller frequency:

ξ_i-correction factor.

This can be obtained through the following formula the total air volume:

$$Q_i = \frac{1}{n} A\sum_{i=1}^{18} \xi_i V_i, i = 1,2,.....,n \quad (2)$$

The wind speed sensor in the wind tunnel laboratory method of correction factor ξ_i is: adjust the wind speed wind tunnel to record the different wind speeds of impeller rotation frequency, and then curve fitting to determine the correction factor of the impeller ξ_i, the final calibration The results as shown in the table1.

TABLE I.
IMPELLER CORRECTION FACTOR

ξ_1	ξ_2	ξ_{13}	ξ_4	ξ_5
8052.96	**8117.202**	**7791.434**	**7585.941**	**7477.287**
ξ_6	ξ_7	ξ_8	ξ_9	ξ_{10}
7612.982	**8675.792**	**8442.977**	**8345.645**	**8511.925**
ξ_{12}	ξ_{13}	ξ_{14}	ξ_{15}	ξ_{16}
8388.959	**8633.423**	**8206.463**	**8269.513**	**8070.589**
ξ_{17}	ξ_{18}	ξ_{19}		
7839.829	**8108.363**	**8330.85**		

For example a foreign model cars , automotive cooling fan is turned on, using this system test three consecutive times, the test out of an automobile traveling at different speeds of the air volume, air volume units of m3/h, reference values for foreign car companies in test values under the same conditions. Specific test data table 2.

TABLE II.
THE RELATIONSHIP BETWEEN SPEED AND WIND SPEED TABLES

Speed（km/h）	**0**	**50**	**80**	**110**
Test 1	**3010**	**3610**	**4700**	**5530**
Test 2	**3020**	**3590**	**4400**	**5640**
Test 3	**3050**	**3675**	**4600**	**5580**
Reference	**3048**	**3648**	**4560**	**5544**
Relative error	**0.7%**	**0.6%**	**-0.1%**	**-0.7%**

The above test data with the company's North American use of foreign test system test results are basically the same. The cooling air flow through the

system testing can be seen that the whole test system is simple, easy operation, good reliability, high precision

VIII. CONCLUSIONS

The computer control system is the efficient and modern means for vehicle performance testing. Through computer networks, it also can accomplish information resources sharing, hardware resources sharing and software resource sharing. Furthermore, it has realized the data exchange and information communication between check stations and road transportation management agency, check stations and check stations within the province. Not only has it strengthened the management department's supervision to testing service, but also has improved the self-testing level of check stations.

REFERENCES

[1] Xuhui MA, Hongmin.Application of distributed computer test equipment in the auto comprehensive performance te st.Instrument Technique and Sensor, 2000.

[2] Xing-jia JIANG, Cai-hong YAO. Design of Distributed Comprehensive Performance Intellectual Detection System of Auto mobile.Gansu Science and Technology, 2005.

[3] Xiang-mo ZHAO, Ma-jian. Distributed computer-net automatic testing system on vehicle general performances.Journal of Chang'an University, 2003.

[4] Auto Comprehensive Performance Management System for automatic detection, Plant Maintenance Engineering 1999, pp.76–83.

[5] Chaoyong-Yan,Bo-Jiang.Application and Research of Computer Administration System in Vehicle General Performance Inspecting Station, Modern Electronics Technique 2007. pp.126–141.

[6] Jialong-Hong,Shilun-Li, The Characteristics and Prospects of Automobile Integrated Performance Automatic Detecting System. Computer And Communications, March 2003.

[7] Yabin-Shi,Bin-Xu. Automobile Network Control Self-checking System. AUTO ELECTRIC PARTS , June 2005.

[8] Xinhou-Zhang,Yalan-Zhang. The application of the computer network to detecting safety performances of motor vehicles. JOURNAL OF GUANGXI UNIVERSITY OF TECHNOLOGY. March 2004.

[9] Changming-Han,Wei-Chen. Auto diagnosis station in net. COMMUNICATIONS SCIENCE AND TECHNOLOGY HEILONGJIANG , April 2001.

[10] Xingliang-Chen,Ji-Xu. Research of Automotive Performance Test System. COMPUTER ENGINEERING AND APPLICATIONS 2002

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 144-148

Policy-based Autonomic Mobile Network Resource Management Architecture

Fang Mei[1], Yanheng Liu[1], Hui Kang[1*], and Shuangshuang Zhang[1]
[1.]College of Computer Science and Technology, Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun 130012, China
Email: {meifang, yhliu, kanghui}@jlu.edu.cn, zss198906@163.com

***Abstract*—In this paper, we establish a policy-based autonomic mobile network resource management architecture, then put forward the policies definition and hierarchies of the autonomic management system. We also present the various specific policy forms and the working mechanism of autonomic mobile resource distribution processes. Through the simulation experiments, we prove that the autonomic mobile resource management architecture has a lot of distinct advantages and flexibility contrast to the methods of fixed resource reservation and utility maximization resource allocation algorithm.**

***Index Tems*—autonomic communication, mobile network resource allocation, policy, utility function**

I. INTRODUCTION

The rapid development of the multimedia application and the complexity of the network infrastructure management have made it a new research hot spot to seek for a network and communication architecture which have the capacity of self-management, self-configuration, self-optimization. The key problem of providing end-to-end QoS guarantee is how to manage the network resource effectively in large-scale network environment. Therefore, it is urgently needed to provide an autonomic method which can meet the multidimensional requirement (such as, Service flow types, User requirements, Network status, etc) to manage the network resource dynamically.

At present, it has gone deep into trending to adopt the dynamic and self-adapting method to do the resource allocation research which concerning the fixed network and wireless networks. However, the current outcomes are mainly limited in dynamic reservation[1,2] and adjustment algorithm[3,4], and to realize the dynamism and self-adaptability of the resource allocation through monitoring the QoS of each service in the domain periodically and execute relevant algorithms. In mobile environment, this kind of method that fixed the allocation and adjustment logic inside the network resource management module can not afford enough flexibility and intelligence due to the variability of the network and the mobility of the users.

The goal of this article is to establish an autonomic management architecture of the mobile network resource, to realize self-management, self-configuration, self-optimization by combining policies and utility computing, meanwhile, to improve the utilization rate of the limited resource and guarantee all kinds of QoS of real-time and non real-time services.

II. POLICY-BASED AUTONOMIC MOBILE NETWORK RESOURCE MANAGEMENT

The article will take IBM autonomic management system[5,6] and universal policy management architecture[7] of IETF as prototypes to present the policy-based autonomic mobile network resource management architecture, shows in Fig1, aiming at the characteristic and requirement of the wireless resource management in mobile environment.

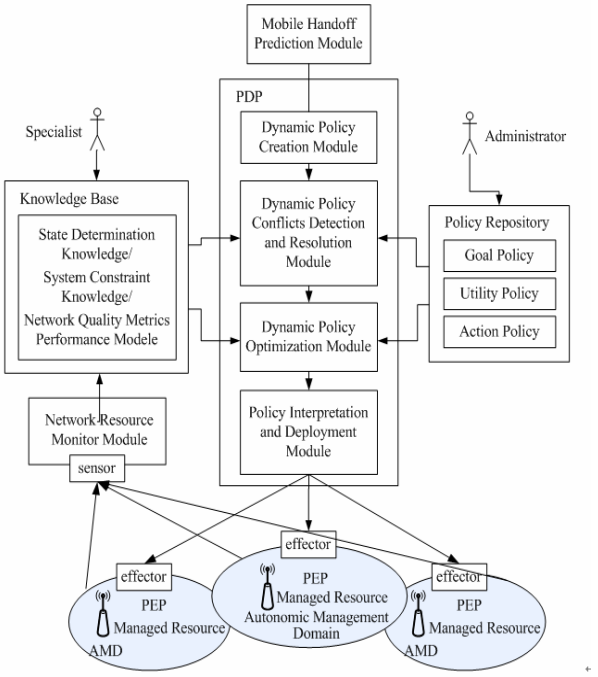


Figure 1. Policy-based Mobile Resource Autonomic Management Architecture

In this architecture, "network resource monitoring module" constitutes the monitor part of the autonomic management system, "policy decision point" constitutes the "analysis" part, all the policies in the "policy repository" made by the administrator constitute the "plan" part, "policy execution point" constitutes the "execution" part, all the experts knowledge in the knowledge base can support the function of these several parts above. These parts together form the "MAPE-K" control ring defined in the IBM Autonomic Computing concept; therefore, this architecture can meet the requirement of an autonomic management system.

III. AUTONOMIC RESOURCE ALLOCATION PROCESSES

A. Determine the SuperiorLevel Management Requirement

The service's QoS should be ensured according to the SLA signed by the user when wireless cell is providing services to mobile users. In SLA, QoE (Quality of Experience) describes the subjective feelings to the service properties of the mobile end users, it can be denoted by a kind of quantization way to indicate the end users' experiences and feelings about the service and network, and reflect the gaps between the current network and service's QoS and the expectation of the users[8].

$$QoE_i = f\left(KPI_{i1},\ldots,KPI_{in}\right) \qquad i=1,\ldots,K \qquad (1)$$

K represents the number of the QoE parameters used to measure the satisfaction degree of the users about the mobile services; n represents the number of the KPI parameters used to evaluate the *i*th QoE network performance parameter .

Wireless access network should also ensure to make the best use of the network resources to maximize the benefits of the wireless service providers at the same time when the quality of the users' service is ensured, and the benefits can be denoted by the Network Utility Function, NUF.

Therefore, the superior level management requirement of the PBMRAM framework can be represented as:

$$Max\left(QoE\right) \& Max\left(NUF\right) \qquad (2)$$

B. Goal-policy Description

Further, to transform the superior level management requirement into the policies of each level to serve the guidance effects. Goal policy describes the administrator's expectations and constrains in the performance and priority of the system. It can be classified to the following categories according to its effect in the system.

1) System Performance Goal Policy.

System performance goal policy describes the range of network performance parameters that each type of mobile service should reach after accessing the wireless cells according to the SLA users signed .Specific example is shown in Figure 2.

```
<GoalPolicy1> ::= <PerformencePolicyRule>
<PerformencePolicyRule>::= IF < PolicyCondition1 > THEN <PolicyAction1>
<PolicyCondition1>::= <ServiceType=ConvesationalClass>
<PolicyAction1>::=<MaxTransmissionRate≤2048Kbps> <MaxTransferDelay
    ≤100ms> <ResidualBER≤5*10^-2><GuaranteedBitRate≤2048Kbps >
```

Figure 2.Example of System Performance Goal Policy

2) System Performance Optimization Policy.

System performance optimization policy describes that it should whether or not to degrade the service level and compress the allocated network resource of the users according to the service category and the signed SLA when the resource of wireless access network is not enough; and whether or not to restore the once degraded service and increase the allocated resource when the network load come down and resource is recovered to be enough. Specific example is shown in Figure 3.

```
<GoalPolicy2> ::= <OptimizationPolicyRule1> |<OptimizationPolicyRule2>
<OptimizationPolicyRule1>::= IF <PolicyCondition2> THEN <PolicyAction2>
<PolicyCondition2>::= <ResouceAvailableStatus=inadequte>
    <ServiceType=BackgroundService>
<PolicyAction2>::=<ReduceResouceAllocation><RRAFlag=on>
    |<DecreaseQoSLevel><DQLFlag=on>
<OptimizationPolicyRule2>::= IF <PolicyCondition3> THEN <PolicyAction3>
<PolicyCondition3>::= <ResouceAvailableStatus=sufficient>
    <ServiceType=BackgroundService><RRAFlag=on>|<DQLFlag=on>
<PolicyAction3>::=<IncreaseResouceAllocation> |<IncreaseQoSLevel>
```

Figure 3. Example of System Performance Optimization Policy

3) Service Handle Priority Policy.

Service handle priority policy describes the priority of all the service types and users when more than one mobile users are applying to access to the wireless cell. Specific example is shown in Figure4.

```
<GoalPolicy3> ::= <PriorityPolicyRule1> |<PriorityPolicyRule2>
<PriorityPolicyRule1>::= IF <PolicyCondition4> THEN <PolicyAction4>
<PolicyCondition4>::= <ServiceType=InteractiveClass>
<PolicyAction4>::=<TrafficHandlingPriority=1>
<PriorityPolicyRule2>::= IF <PolicyCondition5> THEN <PolicyAction5>
<PolicyCondition5>::= <ServiceType=ConversationalClass>
```

Figure 4.Example of Service Handle Priority Policy

C. System Utility Analysis

Then we will generate the utility policy which describes the system performance optimization assessment model, by combining the state determination knowledge and the network quality parameters performance model instituted by the experts in the knowledge base. The utility functions in the utility policy are the quantization assessments of the user's QoE and the system's NUF.

1) Service Categories Utility Functions.

The users in mobile environment can be classified into four types: conversational class, stream class, interactive class and background class, each class is related to several rock-bottom parameters of network performance. To simplify the system modeling, here the utility functions of all the classes are hypothetic to be only related to the allocated bandwidth resource.

Now, conforming to document[9],we present the utility functions of each service classes used in utility policies in PBMRAM, and the maximum utility value of each class is 1 hypothetically.

a) Conversational Class Service. It is a real-time service, and its characteristics are like this. Its demand of data transmitting rate is not high, but it requires the relative small delay and jitter, and low bit error rate. Therefore, it is sensitive to the change of bandwidth, the minimum bandwidth is hypothetically $B_{minl,}$ when the broadband is b_1, the utility function $U_1(b_1)$ can be represented as:

$$U_1(b_1)=\frac{1+\mathrm{sgn}(b_1-B_{\min 1})}{2}\qquad (b_1\ge 0) \tag{3}$$

b) Stream Class Service. It is also a real-time service, its characteristics are that the demand of data transmitting rate is higher and bit error rate is lower, but it can bear larger delay and jitter, the speed commonly between a interval range. When the real bandwidth is b_2, the utility function $U_2(b_2)$ can be

represented as :

$$U_2(b_2)=\frac{1}{1+(1/\varepsilon-1)e^{-r_2 b_2}}\qquad (0\le b_2\le B_{\max 2})$$
$$r_2=2\ln(1/\varepsilon-1)/B_{\max 2} \tag{4}$$

The ε represents the lower utility value user get when b_2 is lower than B_{min2}, and it also determine the sensitive extent of the stream class service towards delay.

c) Interactive Class Service. Its characteristics are that the demand of bit error rate is lower, and the data transmitting rate can also be small or large according to specific service. It has certain demand of delay and jitter, but lower than conversational class service. When the real bandwidth is b_3, the utility function $U_3(b_3)$ can be represented as:

$$U_3(b_3)=\frac{1+\mathrm{sgn}(b_3-B_{\min 3})}{2}\frac{\ln(b_3/B_{\min 3})}{\ln(B_{\max 3}/B_{\min 3})}$$
$$(0\le b_3\le B_{\max 3}) \tag{5}$$

d) Background Class Service. Its characteristic are that it has less constrains of delay and has more elasticity towards data transmitting rate, when the bandwidth is low, service can be provided but the performance is not good enough. When the bandwidth is b_4, utility function $U_4(b_4)$ can be represented as:

$$U_4(b_4)=\frac{\ln(b_4+1)}{\ln(B_{\max 4}+1)}\qquad (0\le b_4\le B_{\max 4}) \tag{6}$$

2) Maximizing System Utility.

In mobile environment, the call that requested to access wireless cell can be sorted into two types: new call and handoff call. Too much interrupts will cause users' complaints because the interrupts switching is hard to accept by user, and will also cause negative impact to the entire system benefit.

Therefore, the utility function need to be defined extended as follows.

If it is a new call request and the required bandwidth the service can not be satisfied, the utility value of the service will fall down to 0, as it is defined before.

$$U_{ni}(b_i)=\begin{cases}0 & b_i\le B_{\min i}\\ u_i(b_i) & B_{\min i}\le b_i\le B_{\max i}\\ 1 & B_{\max i}\le b_i\end{cases} \tag{7}$$

If it is a handoff call request and the required bandwidth of the service can not be satisfied, the utility value of the service will get to be a negative number which indicates the negative impact to the entire benefit.

$$U_{hj}(b_j)=\begin{cases}-1 & b_j\le B_{\min j}\\ u_j(b_j) & B_{\min j}\le b_j\le B_{\max j}\\ 1 & B_{\max j}\le b_j\end{cases} \tag{8}$$

The proportion of each class service should be taken into consideration when calculating the entire utility of the wireless cell for that each class in the system may cause different extent influence to the cell's entire utility. The proportion is set by the system administrator. Suppose the proportion of the influence to system's entire utility that conversational class, stream class, interactive class, background class has is $Spec_i$, and $\sum_{i=1}^{4} Spec_i=1$,the whole capacity of link circuit of a wireless cell L is C, so the maximum prospective entire benefit NUF(L) of this cell can be represented as :

$$Max\,NUF(L)=Max\sum_{i=1}^{4} n_i\cdot Spec_i\cdot u_{ni}(b_i)+$$
$$Max\sum_{i=1}^{4} n_i\cdot Spec_i\cdot u_{hi}(b_i) \tag{9}$$

and meet

$$\sum_{i=1}^{4} n_i\cdot b_i+\sum_{j=1}^{4} n_j\cdot b_j\le C,$$
$$0\le b_i\le B_{\max i},\ 0\le b_j\le B_{\max j}$$

However, non-linear formula (9) is hard to solve and to calculate the bandwidth that each user should be allocated due to the limitation of the inequality .

In the framework of PRMRAM, there is no need to solve it positively, because the real bandwidth allocated to the user is realized by action policy, and the amount of the resource that the dynamic action policy gives is brought from the predicating information.

User can be accessed directly if the system's remaining capacity is enough. When the system's remaining capacity is not enough and if there are more than one user calling for access, it will be calculated using utility policy, to judge which user getting the remaining capacity will cause the maximum benefit to the entire system utility and user utility.

3) Users' QoE Maximization.

Another system expecting state in the superior level management requirement is to maximize the utility of individuals, which is the QoE maximization. It can be represented as:

$$Max(QoE_i)=Max\left(\frac{1}{T_i}\int_0^{T_i} u_i(t)\,dt\right)$$
$$u_i(t)=\frac{1}{k}\sum_{j=1}^{k} u_{ij}(b_{ij}) \tag{10}$$

The $QoE_i(T_i)$ *i*s the *i*th user's experience of the service quality in a period of time T_i. The $u_i(t)$ is the average user's service utility value measured K times in this period of time T_i. The b_{ij} is the user's allocated broadband value of the service measured *j*th time in this period of time T_i.

4) Utility-policy Description.

After the preceding analysis, we can arrive at the utility policy which describes the network performance assessment model. Specific example is shown in Figure 5.

<Utility Policy> ::= <Utility Policy Rule1>|<Utility Policy Rule2>
|<Utility Policy Rule3>
<Utility Policy Rule1>::= IF <Policy Condition6> THEN <Policy Action6>
<Policy Condition6>::= <Call Type=New Call Class>

$$<\text{Policy Action6}>::=<\text{Utility Function}= U_{ni}(b_i)=\begin{cases}0 & b_i \le B_{mini}\\ u_i(b_i) & B_{mini}\le b_i \le B_{maxi}\\ 1 & B_{maxi}\le b_i\end{cases}>$$

<Utility Policy Rule2>::= IF <Policy Condition7> THEN <Policy Action7>
<Policy Condition7>::= < Call Type=Handoff Call Class >

$$<\text{Policy Action7}>::=<\text{Utility Function}= U_{hi}(b_i)=\begin{cases}-1 & b_i \le B_{mini}\\ u_i(b_i) & B_{mini}\le b_i \le B_{maxi}\\ 1 & B_{maxi}\le b_i\end{cases}>$$

<Utility Policy Rule3>::= IF <Policy Condition8> THEN <Policy Action8>
<Policy Condition8>::= <Service Type=Conversational Class>
<Policy Action8>::=<Utility Function=

$$U_1(b_1)=\frac{1+\mathrm{sgn}(b_1-B_{min1})}{2} \quad (b_1\ge 0)>$$

Figure 5 Example of Utility Policy.

D. Action-policy Generation

The action policy which is finally deployed and executed in the mobile network equipment, is divided into two categories: Resource allocation policy, which is generated automatically by the "dynamic policy generation module" according to predicating information and has been done a thorough study by project team members[10]; the specific action policy set which are transformed from the high-level abstraction strategy. These policies work together to achieve the objective described by goal policy and mainly include:

- Security Authorization Policy, completing verification whether the user calling to access the wireless cell is legal or not;
- QoS Configuration Policy, being configured in the equipments that support QoS parameter configuration at the bottom level, carrying out data collection for the "network resource monitoring module", and completing autonomic management system's monitoring functions;
- Admission Control Policy, enforcing the accessing or switching process when the user calls for getting the required resources for example bandwidth through the dynamic policy;
- Resource Adaptation Policy .When the system resources are inadequate, the dynamic policy optimization mechanism produces the resource adaptation policy to adjust the accessed user resource allocation, for example, to extrude the bandwidth of the elastic services or the lower degrade user service.

IV. EXPERIMENT AND SIMULATION

Suppose the total accessed channel capacity of the wireless cell is 10Mbps, the overall average arrival rate of handoff call and new call is in line with the Poisson distribution whose parameter is λ, and the service mean duration meets the negative exponential distribution whose parameter is 1/μ=120s. The resource allocation and access control of the new call and handoff call are enforced under the PBMRAM framework.

System supports four types of service, the property is shown in Table 2, and the parameter ε is set to 0.001.Two wireless resource allocation methods used to compare with PBMRAM are Utility Maximizing Allocation (UMA) and Fixed Reservation Allocation（FRA）.

Table 2 Analysis of Each Service's Property Setting in the Model

Service Type	*Average Rate*	*Average Duration*	B_{min}	B_{max}	*Spref*
conversational class	64 Kbps	60s	64 Kbps	64 Kbps	0.4
stream class	256 Kbps	300s	64 Kbps	350 Kbps	0.2
interactive class	512 Kbps	120s	30 Kbps	1 Mbps	0.3
background class	64 Kbps	30s	0	100 Kbps	0.1

The new call dropping rate of these three wireless resource allocation methods is shown in Figure 6.

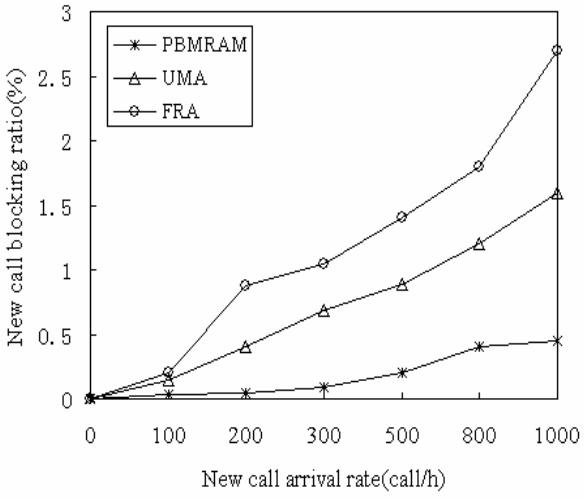


Figure 6. New call Blocking Rate of Three Wireless Resource Allocation Methods

The handoff dropping rate of these three wireless resource allocation methods is shown in Figure 7. The experimental results show that PBMRAM and UMA are significantly superior to FRA in performance; moreover PBMRAM has a variety of policy adaptive methods, making it better than the simple UMA in performance.

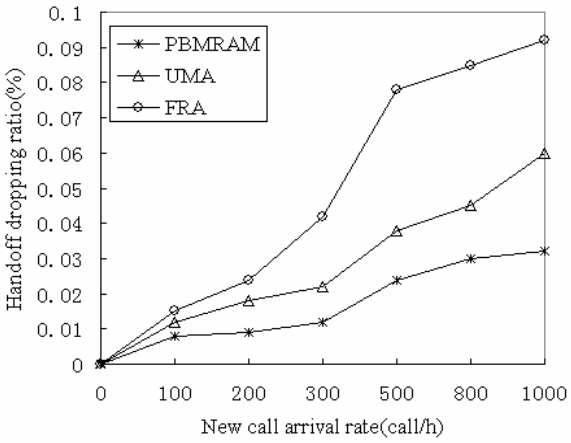


Figure 7. Handoff Dropping Rate of Three Wireless Resource Allocation Methods

Figure 8 shows the bandwidth occupancy of four kinds of mobile user services: VoIP, IPTV, Net games, and FTP download service, with time varying under the PBMRAM framework.

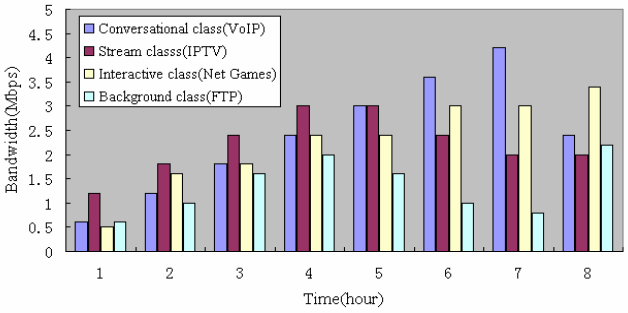


Figure 8. Bandwidth Occupancy of Four Kinds of Mobile User Services

We can see that in the initial case, the system load is small and all service requirements can be met. When the system is close to full load, if there are a variety of new services continuing to request accessing to the system, and the conversational class (Spref = 0.4) and the interactive class (Spref = 0.3) which have greater impact than other on the overall utility can continue to access, while the bandwidth of stream class service and background class service will be squeezed to reduce the bandwidth occupancy under the guidance of policy. The system will keep fully loaded, and background class service will not occupy the bandwidth continuously under the guidance of the policy until a large number of conversational services end or leave the cell.

V. CONCLUSION

The resources allocation of the wireless access network are achieved usually by dividing channel and adjusting threshold, and with the disadvantage that reserving fixed channel may result in the decline of the wireless network resource utilization rate, and distinguishing services by priority then queuing to access may lead to long waiting delay for some services. Furthermore, the division and adjustment of the channel also depends on the experience of administrators. The description of static policy rule itself is not enough to react to the effect on system after different kinds of services access. In this paper, the autonomic mobile resource management architecture is established. By means of the combination of policies and utility computing, mobile resource utilization rate is raised, making the system utility and user QoE to be maximized and guaranteeing QoS with a variety of real-time and non real-time services. Simulation result shows that the policy-based autonomic mobile network resource management architecture has obvious advantages and flexibility.

VI. ACKNOWLEDGEMENT

The authors thank for the discussions and suggestions from the fellows in their laboratory. This work is supported by the National Natural Science Foundation of China (grant no. 60973136) and the International Cooperation and Exchange Project (China-Canadian innovation and commercial development strategies) Special Project (grant no. 2008DFA12140) of the Ministry of Science and technology of China.

REFERENCES

[1] Huang, Y.-R., Lin, Y.-B., & Ho, J. M. Performance analysis for voice/data integration on a finite mobile systems[J]. IEEE Transactions on Vehicular Technology, 49, 367–378. March, 2000.

[2] Lai, F. S., Misic, J., & Chanson, S. T. Complete sharing versus partitioning: quality of service management for wireless multimedia networks[C]. In Proceedings of the 7th International Conference on Computer Communications & Networks, pp. 584–593. October, 1998.

[3] Mariana Dirani, Tijani Chahed: Framework for Resource Allocation in Heterogeneous Wireless Networks Using Game Theory[C]. EuroNGI Workshop 2006: 144-154 .

[4] Takayuki Osogami, Mor Harchol-Balter, and Alan Scheller-Wolf, "Robustness and performance of threshold-based resource allocation policies," Technical Report CMU-CS-04-112, 2004.

[5] Horn P. Autonomic computing: IBM's perspective on the state of information technology. IBM Corporation, 2001.

[6] Kephart JO, Chess DM. The vision of autonomic computing. IEEE Computer, 2003,36(1):41-50.

[7] A Westerinen, et al. "Terminology for Policy baesd Management". IETF RFC 3198, November 2001.

[8] David Soldani, Man Li, Renaud Cuny. QoS and QoE Management in UMTS Cellular Systems[M]. Wiley, August, 2006.

[9] Shi Lei, Changbin Liu, and Bin Liu. Network utility maximization for triple-play services[C], Computer Communications, Volume 3, Issue 10 , 2008: 2257-2269.

[10] Liu Xue-jie, Liu Yan-heng, Wei Da, Wen Zhe. Research on generation mechanism to dynamic policy of resources allocation for mobile IP handoff[J], Journal on Communications,2006,27(10): 108-115.

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 149-152

A Energy-efficient Scheduling Algorithm in mWSN

Xinlian Zhou[1],and Bin Xiao[2]
[1]School of Computer Science and Engineering, Hunan University of Science and Technology University, Xiangtan, china
E-mail:zhouxinlian68@sina.com
[2]School of Computer Science and Engineering, Hunan University of Science and Technology University, Xiangtan, china
E-mail:xiaobing90@163.com

***Abstract*—We research the topological control of the combination of the clustering hierarchical structure and sleep scheduling mechanism. Firstly, present a inner-cluster node scheduling algorithm solution of inner-cluster connectivity coverage problem, which can meet user's expected coverage,avoiding the influence of mobile node location. Based on coverage analysis theory, deducts smallest mobile nodes number, realizes inner-cluster nodes optimal scheduling, which only select k nodes with higher energy and nearer close to fixed node, others should be sleeping. Then, the whole network use TDMA to uniformly divide the time slots to avoid the interrupt of inter-cluster and inner-cluster communication. The scheduling makes the half of cluster realize parallel work, and cross-layer design. Simulation result display, EDG(Efficient Data Gathering) decreases data delay, and largely relieves the burden of cluster-head, and has apparent energy-saving effect, and considers node's mobility, can preferably suit to mobile wireless sensor network.**

***Index Terms*—mobile wireless sensor networks(mWSN), clustering, energy-efficient, TDMA**

I. INTRODUCTION

WSN is one new information acquisition and process method, and can gather the information of the sense-objects in covered area, with multi-hop wireless communication way, by deploying massive sensor nodes in monitored area[1]. In mobile WSN, mobile nodes has induction、computation and communication, acts as static node[2]. However, mobile nodes have self-configuration and self-organization. The challenge of mobile WSN, includes deployment、position、self-organization、topological control、coverage、connectivity、energy-saving and data process. Its application, not only monitor environment, but also trace goal、search and rescue.

Currently, many domestic and overseas scholars have researched the topological control of wireless sensor network. However, these researches mainly spread them around stationed WSN. Due to node mobility easily result in topology change, route path interrupt and invalidation, mainly algorithm can't apply to mobile wireless sensor network[3], e.g., Do-Seong Kim presented LEACH-M protocol, which can support mobility by enhancement of LEACH protocol[4].

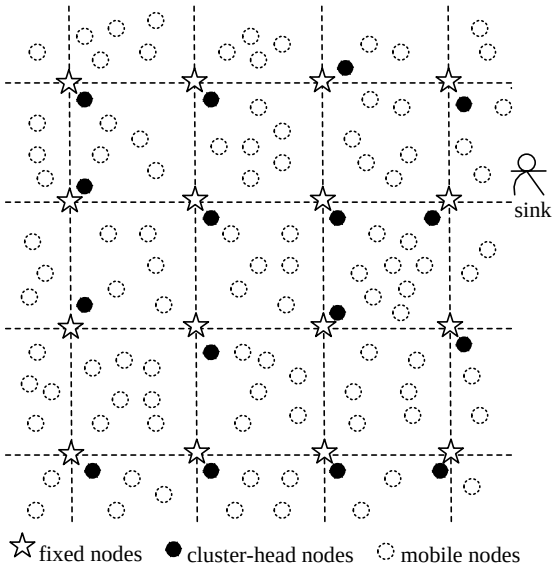
This paper uses the combination of the hierarchical topological control and the nodes sleeping schedule, to realize high efficiency and energy-saving, based on clustering mobile wireless sensor network. We mainly research two aspects:

(1)Inner-cluster nodes optimal scheduling. Introducing inner-cluster coverage think, research the selection of smallest inner-cluster mobile nodes number k, which meet user's expected sensing coverage scale φ, and realize energy-saving and better mobility of nodes.

(2) Using TDMA-Time Division Multiple Address, to uniformly scheduling time slots to whole network, to avoid inter-cluster and inner-cluster communication interrupt..

II. CLUSTERING MOBILE WIRELESS SENSOR NETWORK MODEL

Such as Figure 1, mobile wireless sensor network structure, n nodes are deployed in monitored area Q by the ways thereinafter. Except m fixed nodes deployed in square grid, there are n-m mobile nodes deployed in



Figur1.network structure model

Manuscript received January 1, 2009; revised June 1, 2010; accepted July 1, 2009.

AP-PROC-CS-10CN006

randomly uniform way. These fixed nodes' location is knowable, and distance among these is L (L less than the threshold value d_0 of free space model). Mobile nodes' location is unknowable.

Network model describes as follows:

(1) Mobile wireless network consists of numerous energy heterogeneous nodes, including m fixed nodes and n-m mobile nodes, fixed nodes higher energy than mobile nodes; number: 1,2,…..,m, m+1,……,n;

(2) The number of unique sink is *, and sink can change its location at random;

(3) Beforehand, fixed nodes are deployed in square grid with L width, mobile nodes are deployed at random, and move toward random direction;

(4) All mobile nodes can't obtain their location, no GPS or location algorithm;

(5) The time of all nodes in network is synchronous;

(6) Mobile nodes select nearest fixed node to form cluster. The fixed nodes select nearest node as cluster-head, relying on the remaining energy of inner-cluster mobile nodes. The formed cluster size and uniform location, all think about of the mobility of the cluster-head, avoiding the leave of cluster-head. It is responsible for collecting the inner-cluster sensing data, and aggregating these. The route among fixed nodes, is responsible for the transmission of information to sink.

III. DATA STRUCTURE AND MESSAGE DESIGN IN ALGORITHM

A. Data structure of sensor nodes

The described data structure of nodes in network model this paper researched(including fixed nodes and mobile nodes), such as following: suppose that all nodes store their local information in continuous space.

```
typedef struct
{
int ID;    /*the node's ID,  the fixed nodes and mobile nodes have uniform number*/
float E0;  /*the node's initial energy value*/
float Er; /*the node's remaining energy value*/
int Mrp_ID ; /*The ID referred to the fixed node*/
char state; /*the node state: the fixed nodes denoted by yellow ball 'Y', the mobile nodes initially denoted by white ball 'W', when state is changed, cluster-head denoted by black ball 'B', inner-cluster nodes denoted by gray ball 'G'*/
int hop;          /*the depth of nodes in route tree*/
float Tbegin;   /*the TDMA time slot starting time of node*/
float Tdelay;   /*the TDMA time slot value of node*/
int NumberOfChild;  /*the number of child node in collection tree*/
} NODEDATA;
Struct  NodelLocalData {
  NODEDATA  Nd;
  int parent;     /*indicator denotes the location of the parent node in tree*/
} NLD;
```

In this mobile wireless sensor network model, each fixed node and cluster-head has one neighborhood table NT[] in local, to store the remaining energy value, and the delay time to the fixed node and the state information of its neighborhood nodes.

```
typedef struct{
    int ID;      /*unique ID of neighborhood nodes in whole network*/
char state;      /* the state of neighborhood nodes*/
float dt;        /* the delay time of neighborhood nodes to fixed node*/
float Er;        /* the remaining energy value of this neighborhood node*/
} NeighborhoodTable;
NeighborhoodTable NT[]; /*all neighborhood nodes' information*/
```

B. Message design

Table 1 shows the main message format and its description, for this algorithm design, in which the first column is the message type to distinguish each type of message; the second column is message description; the third column is message format.

TABLE I. MESSAGE FORMAT

Message	Message Description	Message Format
Fixednode_Msg	Staticnode Broadcasting	('1',ID, T0)
Mobilenode_Msg	Mobilenode Broadcasting	('2',ID,Mrp_ID,Er, Δ tij)
KeyFixed_Msg	Key Fixednode Broadcasting	('3',Cid, T , Tdelay，k, MidList[])
Sink_Msg	Sink Broadcasting	('4', parentID, rund, hop)
Clusterhead_Msg	Cluster Head assign slot	('5',ID)
Data_Msg	Sensor Data	('6',Cid, ID,data)
Hello_Msg	Cluster Head Broadcasting Hello	('7', ID)

IV. EDG ALGORITHM

This paper presents the optimal schedule of inner-cluster mobile nodes and whole-network TDMA slot schedule, and achieves an energy-efficient EDG(Efficient Data Gathering).

A. EDG algorithm implementation steps

According to the previous algorithm, the main steps are summarized as follows:

The first step: in distributed clustering algorithm, each fixed node transmits Fixednode_Msg, mobile nodes accept Fixednode_Msg and compute the distance from all fixed nodes around itself, and select the nearest fixed node as a reference. Duration Tc

The second step: form the cluster and select the cluster-head, schedule the k mobile nodes. Duration TH

The referred and fixed nodes receive the Mobilenode_Msg of the round mobile nodes, if the Mrp_ID is equal to their own ID number, these mobile

nodes will be formed one cluster, their messages will be stored in NT[] of the fixed nodes.

The referred fixed nodes will choose greater energy than the average energy, and the nearest node to the cluster-head from the NT[].

Compute k value, and call the optimal schedule algorithm inter-scheduling(), select k working nodes.

The third step: data gathering, duration TD

while carry out data collection, call TDMAscheduling() algorithm to inner-cluster nodes with TDMA, allocate the initial work time with parallel division approach to k inner-cluster working nodes;

the members of inner-cluster can receive Hello_Msg within own slot from the cluster-head, then transmit the sensing data to cluster-head. The cluster-head will aggregate these data and forward them to fixed nodes.

B. EDG algorithm pseudo-code description

```
EDG()
{while( the time (Tc) is not expired)      /* Tc stage clustering begin */
{ if ( Nd.state = 'Y')      /*only for the fixed nodes*/
  {  Broadcast(Fixednode_Msg);  /*transmit the message containing its own ID and time stamp */
   }
  if (Nd.state ≠ 'Y')      /*only for mobile node*/
  {  Receive(Fixednode_Msg); /*receive up to 4 Fixednode_Msg message */
   Δtij=|ti1-t0|; /* compute the distance from different fixed nodes, approximately replace with delay time*/
   Mrp_ID=Choose min( Δtij); /* select the nearest fixed node as the referred fixed node*/
  }
}/* Tc stage end*/
while(the time (TH) is not expired)      /* selecting cluster-head TH stage begin*/
{ if(Nd.state ≠ 'Y') Broadcast(Mobilenode_Msg) to Mrp_ID;
/*all mobile nodes broadcast messages to round fixed nodes, including its own ID, Mrp_ID, remaining energy Er, Δti */
 if(Nd.state='Y')  /*the referred fixed node selects cluster-head*/
  { Receive(Mobilenode_Msg); /* continuous receive the messages from round mobile nodes, which satisfy the condition (Nd.ID be equal to Mrp_ID of message ). And these messages are stored in its own NT[].*/
 k=INT(lg(1-φ)/lg((rc*rc+2*rc*rs)/(rc+rs)(rc+rs))+0.5);
The array NT[] order from small to large by dt properties, the previous k records will be stored in NT[].
Compute the average remaining energy Ea in NT[]. In the conditions of Er≥ Ea, select the nearest mobile node from itself as the cluster-head, to KeyFixed_Msg.Cid.
KeyFixed_Msg.MidLIst[1..k]=NT[1..k];  /*store the previous k nodes in NT[] to ID list of mobile member nodes in KeyFixed_Msg message.*/
Broadcast(KeyFixed_Msg);  /*broadcast the messages to inner-cluster members, including its own ID and the ID list of inner-cluster mobile member nodes*/
  }
  if (Nd.state≠'Y')      /*only for mobile nodes*/
  { Receive(KeyFixed_Msg);
   if (Nd.ID∈KeyFixed_Msg.MidList[1..k])
{ if(Nd.ID= KeyFixed_Msg.Cid)  /*if it's cluster-head */
   { Nd.state='B';  /*store the KeyFixed_Msg message into NT[] of the cluster-head */
   }
 else Nd.state='G';  /*the k selected inner-cluster nodes*/
  }else sleep;  /*nodes not be selected in cluster */
 }
}/* selecting cluster-head TH stage end*/
while( the time (TD) is not expired)
          /*collecting data TD stage begin*/
{ for(i=1;i<Cycle;i++) /*the times of data collection in this round*/
{  if(Nd.state='B') /*cluster-head compute TDMA slots in Td/2*/
{  Td= TD /Cycle;
Clusterhead_Msg .Tbegin[1]=(i-1)* Td;
if(the addition of x,y coordinates of ID be odd)
   for(j=2;j<=k;j++)  Clusterhead_Msg .Tbegin[j]=
Clusterhead_Msg .Tbegin[j-1]+( Td/2)/k;
if(the addition of x,y coordinates of ID be even)
 for(j=2;j<=k; j ++)  Clusterhead_Msg .Tbegin[j]=
Clusterhead_Msg .Tbegin[j]+( Td/2)/k+ Td/2;
Broadcast(Clusterhead_Msg) to  The members in cluster;
/*Clusterhead_Msg containing ID and the slots of inner-cluster member*/
  }
  while(the own slot of inner-cluster mobile node is not time-out)
  { if(Nd.state='B')  Broadcast(Hello_Msg);
/*cluster-head broadcast Hello_Msg, and make sure if it is in this cluster*/
   if((Nd.state='G')
   { if(receive Hello_Msg)
Broadcast(Data_Msg);
    else
Delete this node from slots list of cluster-head;/* this node has left this cluster*/
   }
  }
 }/*end for */
}/* data collection TD stage end*/
}
```

V. SIMULATION RESULT DESIGN

We mainly compare the simulation performance between this EDG and LEACH-M protocol.

The experimental results are the even results of 100 times independent experiments, which use different random topologies. A specific set of experimental parameters in Table 2.

(1) The average delay at different number

The relation of the data average delay and number in surveillance region 1000*1000 m^2, the delay of this EDG protocol will slightly increase along with the number, but the LEACH-M will significantly increase along with it. Because along with the increase in the number of nodes, the latter increase the delay of inner-cluster data

TABLE II. EXPERIMENTAL DEFAULT PARAMETERS

Parameter	Value
Network area size（m）	1000×1000
FixedNode number, Init_Energy（J）	144,50
MobileNode number, Init_Energy(J)	2000,20
Size of square grid(m)	80
FixedNode Transmission Radiu(m)	85
MobileNode Transmission Radiu(m)	60
Speed of mobile node(m/s)	0～3
Cluster radius (rc)(m)	60
Sensing range (rs)(m)	20
E_{elec} (nJ/bit)	50
ε_{fs} (pJ/bit/㎡)	10
E_{fusion} (nJ/bit)	5
Data packet size(Byte)	512
Round	2000
satisfying expected coverage degree of application φ	0.99,0.95,0.98
Period of each round	5s

collection, the entire network delay also increases. The results shown in Figure 2.

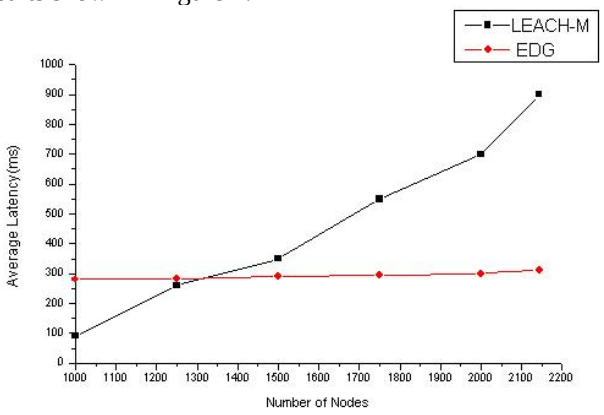


Figure2. the delay comparison at different number

(2) The impact of the number of nodes deployed in the region to performance

Monitoring area 1000*1000 m^2, the number of nodes deployed from 1000 to 3000, the node cluster radius and sensing radius, respectively, 60m and 20m, the coverage scale 99% of user's requirements. In experiment, EDG uses the data gathering protocol of inner-cluster node schedule; but EDG1 doesn't use the schedule. The results shown in Figure 3.

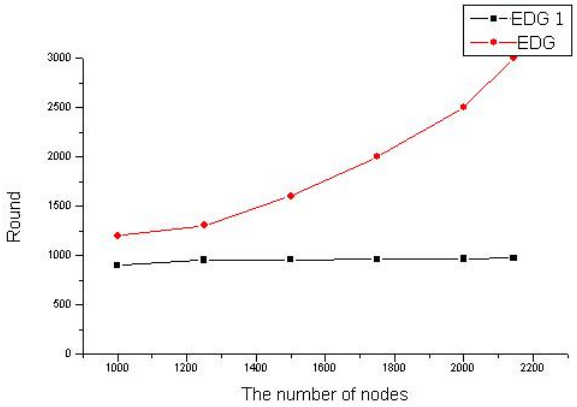


Figure3. the impact comparison of the number to the performance of EDG protocol

As can be seen from Figure 3, the EDG1 without inner-cluster schedule algorithm, along with the increase of the number of nodes, can't effectively prolong the lifetime of network. This is mainly due to nodes redundancy up to a certain extent. The number of cluster in EDG1 protocol has relation to the number of fixed nodes, no relation to the number of nodes. For EDG1, the number of inner-cluster member nodes also increase, when the node density increases, the overhead of cluster-head significantly increases, the round number of data collection will not increase, be up to certain, and the cluster-head is likely to die. After the EDG uses inner-cluster nodes schedule algorithm, the number of working nodes in each cluster will definitely decease, the higher degree of redundancy of nodes, the more sleeping nodes. Clearly, the longer lifetime, the more round number of data collection, so there is a better protocol scalability.

REFERENCES

[1] Ian F Akyildiz, Weilian Su, Yogesh Sankarasubramanian and Erdal Cayirci, A survey on Sensor Networks, IEEE Communications Magazine, August 2002

[2] Jennifer Yick, Biswanath Mukherjee and Dipak Ghosal, Wireless sensor network survey, Computer Networks, Vol. 52, No. 12, pp. 2292-2330, Aug. 2008.

[3] Sriram Chellappan， On Deployment and Security in Mobile Wireless Sensor Networks:Algorithms Design, Vulnerability Assessment and Analysis ， ISBN: 363918257X,pages 7–9.VDM Verlag,July 19,2009

[4] Do-Seong Kim,Yeong-Jee Chung.Self-Organization Routing Protocol Supporting Mobile Nodes for Wireless Sensor Network[C].Proceedings of the First International Multi-Symposiums on Computer and Computational Sciences (IMSCCS'06), 2006

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 153-156

A New Method to Classify Shots in Basketball Video

Yun Liu, Chao Huang, and Xueying Liu
College of Information Science and Technology of
Qingdao University of Science and Technology, Shandong province, P.R.China
Email:Lyun-1027@163.com
Email:{hchaopro, lxying2009}@qq.com

***Abstract*—As the foundation of the sports video annotation, shots classification is presented in this paper. Using non-supervised method the shots are clustered into defined classes (in-play, close-up and free-throw) based on the low-level features of the image (the main color and the histogram). After comparing the clusters None Euclidean Relational Fuzzy C-means (NERFCM) is applied to cluster the shots. Experiments prove its efficiency and sensitivity.**

***Index Terms*--shots classification; NERF C-mean; shots boundary detection**

I. INTRODUCTION

With the development of the computer and internet, multimedia is widely used on the internet and becoming the main carrier of the information, while people face a lot of tremendous problems for its enormous size which is how to use the data efficiently. Currently, most works focus on specific sports games in order to investigate the roles of different information sources or statistical learning algorithms in structure analysis and semantics extraction. The main challenge lies in the amount of variation in low-level visual and auditory features, and game-specific rules.

One of the popular approaches in sports video analysis is event detection based method [1,2,3]. These methods aim to find interesting events (e.g., shots in a basketball game) by combining multiple cues (visual, audio, and textual features). On the other hand, other efforts tried to find interesting points of a video by modeling users' excitement level[4,5] . Based on findings from psychophysiology, human excitements are mimicked using low–level features such as activity of motion, audio level, and characteristics from video editing.

Although these systems have been successful in their respective applications, shot type classification for sports video is still an unsolved problem.

In this paper we propose a shot classify system to cluster the shots based on the low-level features of the frames. The rest of this paper is organized as the follows. Section II is an introduction of the new method to find the boundary of the video stream, and then features of the frames are extracted in section III, in section IV 4 C-means algorithms are compared and the optimal method is chosen to cluster the frames in a shot. Some shots classes clustered are listed in Section **V.** In Section **VI,** we draw conclusions and discuss same future work.

II. SHOT BOUNDARY DETECTION

This section explains the algorithms for shot boundary detection, which is the detection of the different types of lenses in the video.

Shot boundary detection is usually the first step in generic video processing. Although it has a long research history, it is not a completely solved problem[6]. Sports video is arguably one of the most challenging domains for robust shot boundary detection due to following observations:

1) There is strong color correlation between sports video shots that usually does not occur in a generic video.

2) Sports video is characterized by large camera and object motions. Pans and zooms are extensively used to track and focus moving game objects, respectively.

3) A sports video clip almost always contains both cuts and gradual transitions, such as wipes and dissolves.

At present, the research on the boundary detection fall into two primary categories: detection based on the compressed video and detection based on the primary video. The later is faster in detection because it doesn't decompress the video, but for this very reason, the features are limited and inaccurate[7].

A. The histogram of the image

The algorithm basing on the histogram is developed in the basis of pixels. In a shot the adjacent frames' pixel difference is tiny and so is the diagram's standard deviation. If the scene cuts from one shot to another, the difference between them will be very large and so is the standard deviation, for that reason we select standard deviation to measure the shot transition.

In Fig.1 we have shown different types of the frames we classify in our paper and their histograms. (a) and (b) are the frames in play and their histogram, (c) and (d) one are the frames in close up and their histograms, (b) and (c) are the shot boundary frame. From the difference of the histogram (g) we can conclude that histogram can be expressed as the feature. At the boundary the standard is very large as a single peak as shown in Figure1.

Generally speaking, the algorithm basing on the histogram is simple to realize and has lower complexity, moreover it can get a preferable result on a detecting video abrupt change. However, compute pixel by pixel is

This paper is supported by Natural Science Fund of Shandong (Y2008G09).

AP-PROC-CS-10CN006

inefficient and computationally expensive, therefore we use the following method to calculate the critical region in section B.

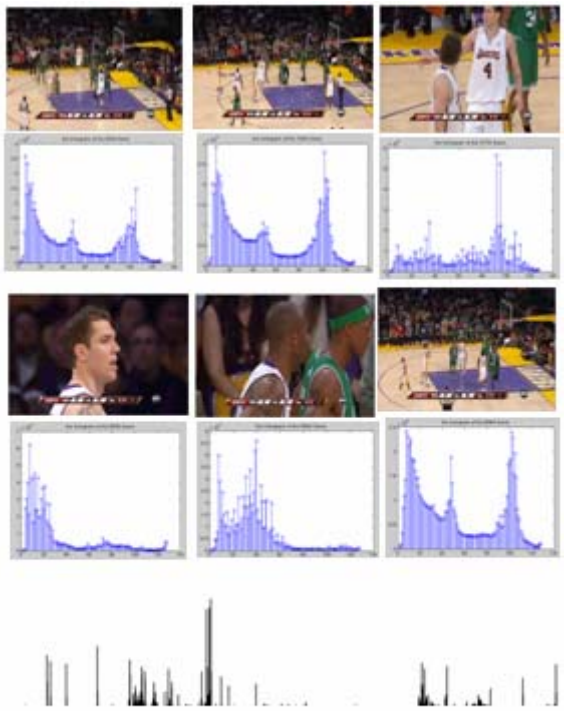


Fig. 1 The Frames and Their Histograms of the Video and the last figure is the Difference Between the Frames. From the top down is a, b, c, d, e, f, g.

B. The histogram of the key area

We propose a compute-easy, yet very efficient, cinematographic algorithm for the frames with a high value. We define regions by using Golden Section spatial composition rule [8,9], which suggests dividing up the screen in 3 : 5 : 3 proportion in both directions, and positioning the main subjects on the intersection points of these lines. We have revised this rule for basketball video, and divide the whole frame. In Figure 2, the examples of the regions obtained by Golden Section rule is displayed on several close up and in play shot.

Fig. 2. Shot Types in Basketball Match the Original Frame of in-play (a), close-up(c), in the Video and their Histogram (b) (d)(f);

III. FEATURES EXTRACTION

In the basketball video, the basic feature consists of the color feature and histogram, which is sufficient for cluster the frame into different classes. In this section the color feature and histogram will be extracted as the input of the classifier.

A. The court color detection

Based on the partition of the frames we propose the implementation to detect the main color of the court. We convert the RGB to HSL in region 2 defined in section 2 and calculate the main color as the frame's color feature.

Due to the sensitivity of the human eye than the right of the Hue and Saturation, we chose the Hue value of the basketball court as the feature and calculate the proportion of the court color.

B. The histogram of the R2 region

This section explains the histogram of the middlemost region. Due to its usage of the statistic of the pixel's light and color without regard to the location of the pixel, histogram is one of the most used methods to calculate the inter-frame difference. Suppose D is the difference of two frames, and D is defined as:

$$D = \sum_{k=0}^{k-1} | H[f(x,y,t),k] - H[f(x,y,t+1),k] | \qquad (1)$$

If the difference D is bigger than a threshold, the conversion occurred in the lens.

IV. CLUSTERING BY C-MEANS

Clustering is a mathematical tool that attempts to discover structures or certain patterns in a data set, where the objects inside each cluster show a certain degree of similarity. In the framework of fuzzy clustering, it allows each feature vector to belong to more than one cluster with different membership degrees (between 0 and 1) and vague or fuzzy boundaries between clusters[10] . And also an alternative way of NERF C-means is proposed recently for any relational data clustering [11].

Automatic recognition of frames is a challenging problem and is highly demanded in an intelligent video surveillance system. A video can be characterized by a sequence of shots, each shot contains a frame sequence which is difference from the others.

Our observation on Fig.2 has shown that similar frames share a set of similar histogram. The process involves feature extraction at the histogram to form a descriptor for each frame and clustering of the frames into a set.

A. Pros and cons of FCM

The popularity and usefulness of fuzzy C-means result from three facts. The algorithms are simple; they are very effective at finding minimizes of objective function J_m: give data set

$$X = \{x_1, \quad x_2, \quad \cdots \quad x_n\} \qquad (2)$$

Where n is the number of data points in X , $x_k \in R^p$, p is the number of features in each vector x_k ; in order to cluster X into C prototypes, J_m is sought as

$$\min_{(U,V)} \left\{ J_m(U,v) = \sum_{i=1}^{c} \sum_{k=1}^{n} u_{ik}^m D_{ik}^2 \right\} \quad (3)$$

Constrain $\sum_{i=1}^{c} u_{ik} = 1, \forall k$, and distance

$$D_{ik}^2 = \|x_k - v_i\|_A^2 \quad (4),$$

$$\text{norm } \|x_k\|_A = \sqrt{\langle x, x \rangle_A} = \sqrt{x^T A x} \quad (5),$$

Degree of fuzzy $m \geq 1$,

$$v = (v_1, \quad v_2, \quad \cdots \quad v_C)^T \quad (6)$$

And U is the membership functions, those minimizes usually represent the structure of X very well; test result is shown in Fig.3. Various theoretical properties of the algorithms are well understood, and are described in Refs [15]. Additionally, this method is unsupervised and always convergent.

Also this method does have some disadvantages, such as, long computational time, Sensitivity to the initial guess (speed, local minima), unable to handle noisy data and outliers, very large or very small values could skew the mean, not suitable to discover clusters with non-convex shapes.

B. Out-performance of RFCM compared to FCM

1) Broaden the application domains of FCM

The RFCM classifier is useful when a feature space has an extremely high dimensionality that exceeds the number of objects and many of the feature values are missing, or when only relational data are available instead of the object data. So RFCM can deal with more than the problems that FCM can do.

2) Efficiency on computations

Whenever relational data are available that corresponds to measures of pair wise distances between objects, RFCM can be used instead which rely on its computation efficiency. One of the advantages is that their driving criterion is "global".

3) Limitations of RFCM

RFCM has a strong restriction which restrains its applications. The relation matrix R must be Euclidean, i.e., there exists a set of N object data points in some p-space whose squared Euclidean distances match values in R. To ease the restrictions that RFCM imposes on the dissimilarity matrix, there are two improved versions of RFCM which are introduced in the following.

C. Analysis of two improved RFCM Algorithms

NERFCM can transform the Euclidean relational matrix into Euclidean ones by using the β-spread transformation introduced in [5]. This transformation consists of adding a positive number β to all off-diagonal elements of R. As proved in [4], there exists a positive number β_0 such that the β-spread transformed matrix R_β is Euclidean for all $\beta \geq \beta_0$, and is not Euclidean for all $\beta \leq \beta_0$. The parameter β, which determines the amount of spreading, should be chosen as small as possible to avoid unnecessary spreads of data with consequent loss of cluster information.

On the other hand, the exact computation of β_0 involves an expensive eigenvalue computation.

D. Frames Clustering Implementation

We propose to extract the histogram of every frame and We define the HOG similarity of two postures with histogram intersection method, which is:

$$S(p,q) = \sum_{u=1}^{B} \min\{p^{(u)}, q^{(u)}\} \quad (7)$$

Where p and q are 2 histograms with B bins, if they are the same, the similarity s is 1, so the dissimilarity can be defined as $d = 1 - s$. Consequently, the HOG dissimilarity of the total frame N can be calculated, and the whole HOG dissimilarity can form a dissimilarity matrix:

$$D = [d_{ij}]_N = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1N} \\ d_{21} & d_{22} & \cdots & d_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ d_{N1} & d_{N2} & \cdots & d_{NN} \end{bmatrix} \quad (8)$$

Where the value of diagonal elements d_{ii} is 0, the other elements value d_{ij} is the dissimilarity between i and j .The N frames are then clustered into M clusters by employing a pair-wise clustering algorithm which takes the dissimilarity matrix of every pair of samples to be clustered. It intuitively satisfies human senses that sample frames in Fig.5 are classified into the same cluster.

Fig. 3 The Example Frames from Cluster of 2

(a) Example frames of free-throw

(b) Example frames of in-play

Fig. 4 One Cluster Shared by 2 Different Semantic Classes Consisting of frames with Similar Histogram

The relation of HOG dissimilarity in our paper is fuzzy and Non-Euclidean, so we adopt the Non-Euclidean Relational Fuzzy (NERF) C-Means. Its classification

result is the probability value of each sample belonging to the whole posture cluster samples, and the posture cluster is decided by the maximum probability. In this paper N is 3314 and cluster number M is 3. Some clustering results are show in Fig.3 and Fig.4.

It intuitively satisfies human senses that sample frames in Fig.4 are classified into the same cluster, because they are indeed of similar histogram or main color.

V. EXPERIMENTAL RESULTS

In this section, we describe the experiments aimed at evaluating the proposed method, which integrated into a system that was tested using two basketball video sequences.

A. Shots boundary detection

Our proposed shot boundary detection approach has been tested on two matches of basketball from the 2009 All Star NBA matches. The first video is 20 minutes long and contains 127 shots. The second video is of 30 minutes and contains 233 shots. From table 1 the accuracy of our shot boundary detection algorithm is about 93.7%. The false detection may be caused by photo flash, for instance, if the difference of the histogram of the frames located at the photo flash will be tremendous, consequently, will result in incorrect detection. Table 1 lists the performance in terms of the precision and recall.

The dissimilarity matching on two frames are all based on the histogram of the frames shown in Fig.2. A promising performance, Recall 85%-93%, and precision 90%-95%, has been achieved.

TABLE1. THE RSULTOF OUR SHOT BOUNDARY DETECTION

	manually	auto	precision	recall
Video 1	127	106	96.4%	91%
Video 2	233	212	98%	96%

C. Shot classification in basketball video

The total length of the basketball video is about 60 minutes (153 shots) consisting of NBA All Star match (30 min) and CBA (30 min). Table 2 shows recall and precision of shot classification. The lower accuracy of the NBA sequence is caused by close up of players having skin colors similar to play court.

TABLE 2. THE RESULT OF OUR SHOT CLASSIFICATION
M: Manually A: auto R: recall P: precision

	In-play shot				Non-play shot			
	M	A	R	P	M	A	R	P
NBA	16	18	100%	88.9%	69	65	92.8%	98.5%
CBA	21	20	90.5%	95%	83	77	90.4%	97.4%

VI. CONCLUSIONS

We have presented an effective high-level semantic concept of "semantic shot classes", which frequently occurs in broadcast sports video, and compared four fuzzy c-means algorithms and found serious shortcomings for fuzzy c-means. In order to detect this concept, we have proposed a method for semantic shot classification, consequently we give two improved versions for FRCM, and we carried out our research on image clustering with NERFCM effectively. Experiments have shown that an appropriate construction of mid-level representations can improve the accuracy and flexibility of shot classification.

We have justified the proposed mid-level representations through the task of video shot classification. Our future work includes the evaluation of individual features for various tasks; extension of the proposed framework to different sports, such as football, basketball, and baseball, which require different event and object detection modules.

REFERENCES

[1] A. Ekin, A. Tekalp, and R. Mehrotra, "Automatic soccer video analysis and summarization," IEEE Transactions on Circuits & Systems for Video Technology, pp. 796-807, 2003.

[2] C. Liang, J. Kuo, W. Chu, and J. Wu, "Semantic units detection and summarization of baseball videos," 47th Midwest Symposium on Circuits and Systems, 2004.

[3] Y. Zhang, C. Xu, Y. Rui, J. Wang. And H. Lu, "Semantic Event Extraction from Basketball Games using Multi-Modal Analysis," IEEE International Conference on Multimedia and Expo, 2007

[4] A. Hanjalic, "Adaptive extraction of highlights from a sport video based on excitement modeling," IEEE Transactions on Multimedia, pp. 1114-1122, 2005.

[5] R. Ren, J. Jose, and H. Yin.: Affective sports highlight detection," European Signal Processing Conference, 2007

[6] A. Hanjalic, "Shot-boundary detection: Unraveled and resolved?," *IEEETrans. Circuits Syst. Video Technol.*, vol. 12, pp. 90–105, Feb. 2002.

[7] Y. H. Gong, L. T. Sin, C. H. Chuan, H. J. Zhang, and M. Sakauchi, "Automatic parsing of TV soccer programs," in *Proc. Int. Conf. Multimedia Computing and Systems*, Washington, DC, May 15, 1995, pp. 167–174

[8] Cheng Yong, Xu De. A method for shot boundary detection using adaptive threshold [J]. Acta Electronica Sinca, 2004,32(3): 508 511 (in Chinese)]

[9] G. Millerson, The Technique of Television Production, 12th ed. New York: Focal, March 1990

[10] John C. Russ (2005). Image Analysis Of Food Microstructure CRC Press. ISBN 0849322413.

[11] RICHARD J. HATHAWAY and JOHN W. DAVENPORT, Pattern Recognition, Vol. 22, No. 2, pp. 205 212, 1989.

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 157-160

Developing Mobile Input Method System Using An Improved Design Pattern Approach

Fei Zhu[1,2], Hongjun Diao[1,2*], and Wei Huang[1,2*]
1 School of Computer Science and Technology, Soochow University
2 Provincial Key Laboratory for Computer Information Processing Technology, Soochow University
Email: zhufei@suda.edu.cn
Suzhou, Jiangsu, China, 215006
Email: { hjdiao, huangwei}@suda.edu.cn

***Abstract*—Popularity of operation system on mobile phone makes it possible to develop a third party input method system for it. However variety in mobile models and systems makes it difficult to develop a general input method for different models. As design pattern usually add reusability and flexibility to software development process, we propose an approach to develop input method for mobile using state design pattern. With the idea of event-based finite state machine, we design a novel event-condition-state pattern framework. We use it develop a thirty party input method system which contains a Windows Mobile system version and a Symbian system version. We utilized common part of the systems with sub pattern model and then rewrote related interfaces to different systems, instead of carrying out two developing activities. With the state design pattern, the whole development process turns to simpler and clearer, saving us much work in design and developing. It also provides convenience for future maintenance and updating as well.**

***Index Terms*—software developing, design pattern, mobile developing, mobile input method**

I. INTRODUCTION

In software engineering, a design pattern is a general reusable solution to a commonly occurring problem in software design. It is widely used in software development. State pattern, a behavioral software design pattern, is often used to represent the states of an object and is applied to develop software that has multiple states. For example, documents in office automation system with workflow typically have multiple states, including editing, waiting for auditing, being audited, replying and forwarding. Microsoft Pinyin input method system can be in one of the states, including initial state before input, input state, and selecting state after input. State pattern can be applied in these two kinds of software.

However, design pattern is a description or framework for how to solve a problem that can be used in many different situations but not a determined design that can be transformed directly into code. The famous state design pattern proposed by Erich Gamma, Richard Helm, Ralph Johnson and John Vlissides is not a determined design that can be transformed directly into code either.

One of the common methods to implement the state design pattern is to use IF-ELSE-IF statement provided by most programming languages. However too many branches of IF-ELSE-IF will make the program structure and logic unclear, making it harder for future potential modification and updating. Meanwhile, using event-response mechanism provided by modern programming technology, it is of benefit to refine the similar functional structures and change IF-ELSE-IF statement to event-response ones so as to promote readability and reusability as well as relieving design efforts.

In this paper, we propose a novel event-condition-state pattern framework, which is an event-based finite sate machine, for software development. With the pattern framework, we developed a third party input method system which has a Microsoft Mobile version for Microsoft Mobile system and a Symbian version for Symbian system. We found it easy to fulfill development with the help of the framework. We did not carry out two developing activities for input method system. Instead, we utilized common part of the systems with sub pattern model and then rewrote related interfaces to different systems. By this, we spared lots of redundant efforts in developing, and gained convenience in later synchronous updating and maintenance.

II. IMPROVING STATE DESIGN PATTERN

Erich Gamma, Richard Helm, Ralph Johnson and John Vlissides, referred to as the GoF, described recurring solutions to common problems in software design. As we can see in figure 1, although GoF described a reasonable state pattern, their framework only gave relations among context objects and state objects in short, which is not enough for implementation. There normally are many response events, causing the structure of processing method very much complex, which will lead to inefficiency in software development.

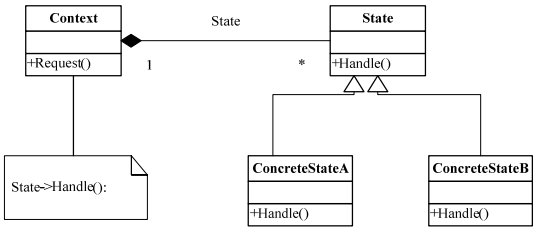


Figure 1. State pattern class diagram which merely shows relations among context object and state object without interpreting details of implementation

As GoF state pattern did not give sufficient details for implementation, developers can complete in various ways. Some prefer setting properties of Context in method Handle. Some use state transferring table in method Request of object Context to retrieve state table to fulfill state transferring, which separates concrete rules out of state pattern.

Context in GoF state pattern is used as user operation interface and also used for state class management. User operation part is usually related with dedicated application, while state management part is normally unchanged in each state pattern program. Therefore, we can separate it to two objects. One is for client object of user operation interface, changing with applications. The other is for state management, achieving state transferring by retrieving and interpreting state transferring table. Classes and their relations of the whole state pattern are shown as figure 2.

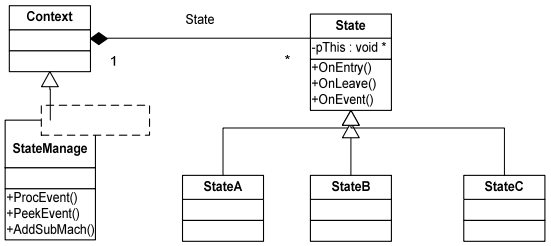


Figure 2. Novel state pattern class relation diagram which demonstrates relations among Context, State and StateManage as well as details of implementation

In figure 2, state management class StateManage inherits from client class Context, while StateManage has an abstract class State to represent state. Taking advantage of StateManage, we can separate class Context of dedicated application from the framework to maximal extent, including functionalities in class State that are related with state management and indexing state transferring table, and embrace user event mapping mechanism in class State. As these functionalities normally keep unchanged, we can place them in event-condition-state general framework as constant part.

IV. Design Mobile Input Method With Event-Condition-State Design Pattern

Demands for intelligence in mobile phone urge development of intelligent mobile phone. Popularity of operation system on mobile phone makes it possible to develop a third party input method system on it. Meanwhile there are various mobile models and systems presently, making it difficult to develop an input method software for them. Moreover, the future update and maintenance of input method will be in trouble as well.

As we can see, although a mobile phone may have different input methods, mobile phone will eventually be in one of the three states which are numeric number input state, English input state and non-English input state. These three states have similar state transferring rules to the ones in event-condition-state pattern. At the same time, each input state can be described by a sub state model. For example, English ABC, an English input method, will be in initial state before user input and in input state after user input. If there is no input within an interval, the system will return to initial state. The whole view of state transferring is shown in figure 3.

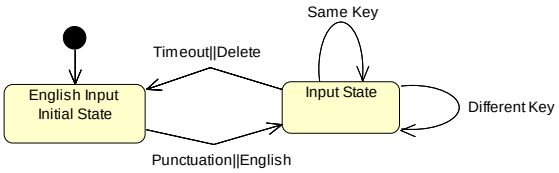


Figure 3. State transferring diagram of English input method

There are usually far less keyboards in mobile than in PC, usually consisting of numeric key 0 to numeric key 9, key shift and key star. To save as many keys as possible, one single key is normally given several alternatives in different cases, e.g. numeric key 0 to numeric key 9 represent numeric 0 to 9 separately in numeric number input state. However, they are used as candidate keys in other situations, e.g. they are used as English candidate keys in English input mode. There will be even more complicated combinations in non-English inputting methods. For instance, in Chinese character input method, they are treated as input candidate keys and they will be turned to selecting keys when a user is selecting from candidate Chinese character. So it is the same with other keys. Key shift is used as shifting and key star is used to input character star. Meanwhile they have other meanings in different input methods.

Although such one key with multiple interpretations is sophisticated, it conforms to the event-condition-state pattern. Thus, it is possible to develop input method systems in mobile with the framework proposed.

Here, we use client interface class Context to encapsulate interfaces to mobile operation system, internal data and message processing component, of which two main interfaces are PressKey which receives and processes information when the key is pressed, and CommitResult which delivers character that is selected by the user.

V. Developing Mobile Input Method Using The Pattern

A. For Windows Mobile System

Windows Mobile is an open and intelligent mobile system. It provides developer a set of interface functions that are similar to those of Windows system. These interface functions are dynamic libraries between system and application. They get user's pressing key message, and send one or more characters to application after processing. Their relations are shown in figure 4.

We can see from figure 3 that it is input manager that interacts with input model directly. User's pressing key message is passed to input method model through input method manager. Characters from input method are also delivered to application through input manager.

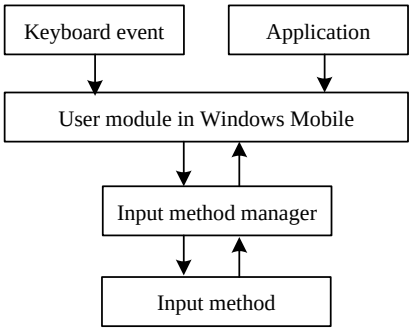


Figure 4. Relation between input method and system in Windows Mobile system

In fact input method manager can interact by input method user interface or by function interface of dynamic library. The difference is that input method user interface is the interface for input manager, while function interface is interface for input manager to maintain the whole input method modules. When applying the model in Windows Mobile, a Context object is necessary to implement these two kinds of interact interfaces requiring by input method manager. The following is part of key code of Context in Windows Mobile system.

```
class Context:public CWinIME{
PreRequest(int iEvent,short wParam,long lParam);
Request(int iEvent,short wParam,long lParam);
:
}
```

Here CWinIME is an UI class for input method, which encapsulates all UI message processing code for input method. Class Context provides UI by inheriting from class CWinIME. Method Prerequest and method Request are two main entries for users to manipulate state machine. Prerequest is used to query how state machine will deal with current message and Request is used to deliver current message to state machine for processing.

B. For Symbian System

Symbian OS is an operating system designed for mobile devices, with associated libraries, user interface frameworks and reference implementations of common tools. It provides an API for input method, called FEP (Front-End Processor), to developers. Developers can implement the interface by inheriting from class CcoeFep which exists in form of polymorphous and dynamic link library. It receives user's pressing key message, and sends characters to application for processing after input method system completes coping with the message.

The whole input method system in Symbian is composed of FEP and FEP server. FEP deals with internal logic of input method. FEP server acts as input method server and is in charge of data resource related processing. Symbian is a multitasking operation system, and there may be several input method instances running simultaneously. Different input method instances can share one thesaurus or corpus without adding more overhead, with which Symbian system achieves data share among input method instances by client server architecture. The relations among user, FEP and application are shown in figure 5.

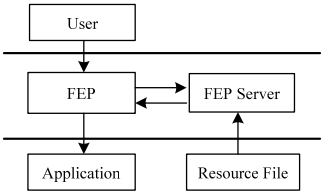


Figure 5. Relations among user, FEP, FEP server and application in Symbian system

We implement FEP API in Context to apply the input method model in Symbian system. The following is part of key code of Context in Symbian system.

```
class Context:public CSybFep{
TKeyResponse OfferKeyEventL (const TKeyEvent& aKeyEvent, TEventCode aEventCode);
int Request(int iEvent,short wParam,long lParam);
:
}
```

Here CSybFep is derived from base class CcoeFep, of which method ConstructL will generate a CSybFepControl object that has the highest priority to handle input event so that it can deal with input event ahead of application. Context implements API in Symbian system by inheriting from class CSybFep. Method OfferKeyEventL and method Request are designed to receive key pressing message and manipulate state machine. OfferKeyEventL overrides method OfferKeyEventL of CSybFep, gets key pressing message with CSybFepControl, and calls method OfferKeyEvent in Context to transform data. Then OfferKeyEvent transmits transformed data to method Request, which delivers its current event to concrete sub state model for further processing.

VI. Conclusion

Before we develop a thirty party input method system in Windows Mobile system and Symbian system, we analyzed different versions of an input method on both systems, finding all the internal logics are basically the same except interfaces to the two systems are different, which has the obvious feature to fit the event-condition-state framework. So we use idea of state pattern in design and other phases in development.

Subsequently we rewrote Context part related with system interfaces of different systems. To avoid redundant developing work on both systems, we encapsulated the part related with internal logic of input method by concrete subclasses. The approach proved to be of benefit for providing convenience for later synchronous updating and maintenance of two different versions of input methods.

The event-condition-state framework proposed in the paper is not only fit for input method development, but also fit for other software design and development.

Acknowledgment

Funding: This work was supported by School of Computer Science and Technology, Soochow University

and Provincial Key Laboratory for Computer Information Processing Technology, Soochow University.

Corresponding authors: Hongjun Diao and Wei Huang.

The work in this paper is part work of “Developing Software Using A Novel Event-Condition-State Pattern Framework: Taking Mobile Input Method System As A Case Study”.

REFERENCES

[1] Fei Zhu, Xiaoxu Zhu, Qiang Lv, Yiyong Shi, Yang Yang and Qiaoming Zhu, “ZFlow: Workflow for Cooperative Editing System”, Journal of Software, Vol.4, NO. 4, p. 339-347, June 2009.

[2] Fei Zhu, Qiang Lv, “ACEAC: A Novel Access Control Model for Cooperative Editing with Workflow”, 2008 International Symposium on Electronic Commerce and Security, pp.1010-1014. IEEE Press, New York, 2008.

[3] Microsoft Developer Network Simplified Chinese MSPY, http://msdn.microsoft.com/en-us/library.

[4] Erich Gamma, Richard Helm, Ralph Johnson and John Vlissides, “Design Patterns:Elements of Reusable Object-Oriented software,” Addison Wesley Pearson, June, 2000.

[5] http://en.wikipedia.org/wiki/.

[6] Hongjun Diao, Fei Zhu, “A novel general event-condition-state pattern framework,” ICIEA, pp. 1024-1027, Industrial Electronics and Applications, 2009.

[7] Andy Wigley, Daniel Moth and Peter Foot, “Microsoft Mobile Development Handbook,” Microsoft Press, March, 2008.

[8] Microsoft Mobile Developer Center, http://msdn.microsoft.com/en-us/windowsmobile/.

[9] Symbian Developer Network, http://developer.symbian.com/.

[10] Ben Morris, “The Symbian OS Architecture Sourcebook: Design and Evolution of a Mobile Phone OS ,” Wiley , July, 2008.

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 161-164

Reviewing some Cloud Computing Platforms[+]

Dong Xu, and Hui Liu
School of Computer Engineering and Science, Shanghai University
High Performance Computing Center, Shanghai University
Shanghai, China
dxu@shu.edu.cn

***Abstract*—Cloud Computing is an emerging technology in which the research community and industries have recently embarked. However, the infrastructures of most cloud computing systems today are invisible to the research community, or are not explicitly designed to the researchers interested in cloud computing systems. In this paper, we depict the Cloud computing paradigm from a variety of aspects, such as definitions, features, and technologies. Moreover, some Cloud computing Systems are illustrated as well as their architectural trade-offs. It would contribute to better comprehension of the Cloud computing technology so as to enable some communities to design more efficient clouds, and the adoption of this novel computing approach is facilitated in scientific areas.**

***Index Terms*—cloud computing, virtualization, distributed paradigm**

I. INTRODUCTION

Currently, the Cloud computing has been emerging as a hot topic since the late of 2007. Industry and academia are starting projects related to Cloud computing. For example, Microsoft has published its Cloud computing system- Windows Azure Platform [1]. Amazon Elastic Compute Cloud [2]; IBM's Blue Cloud [3]; HP, Intel Corporation and Yahoo! Inc. recently announced the creation of a global, multi-data center, open source Cloud computing test bed for industry, research and education [4].

In the last few years, virtualization has introduced some novel system techniques so that the cloud-provider can transparently satisfy its cloud customers' requirements without impacting their own system utilization. Cloud computing differs from grid computing in this regard, it can run in conjunction with the original business workloads. Moreover, novel virtualization technologies, e.g. live-migration and pause-resume give rapid and transparent solutions, interference may not occur between the original systems and the cloud workloads [6]. Consequently, several business cloud models rapidly evolved to harness the technology by providing computing infrastructure, data-storage, software applications and programming platforms as services. However, the inter-relations inside cloud computing system have been ambiguous and the enable feasibility of inter-operability has been debatable while referring to the core cloud computing services. Furthermore, every cloud computing service has an interface and employs a different access control protocol. A unified interface for each cloud computing service to support integrated access to cloud computing services is not existent, though portals and gateways can provide unified web-based user interface. So, the introduction of the cloud computing domain, its components and their inner relations are necessary to help the researchers achieve a better understanding of the novel technology.

The rest of this paper is structured as follows. The motivation for this study is introduced in section II. Section III addresses the three layers in cloud computing, while section IV illustrates some cloud computing platforms. We conclude our paper and give the further work in Section V.

II. MOTIVATION

For the Cloud computing, there aren't yet widely accepted definitions in spite of much attention has been given to the Cloud computing. Many researchers and practitioners have attempted to define Cloud computing in various ways. Here are some views from different perspectives.

A. Wikipedia's view

According to wikipedia's view, cloud computing is defined as that "Cloud computing is Internet- ('cloud-') based development and use of computer technology ('computing'). In concept, it is a paradigm shift whereby details are abstracted from the users who no longer need knowledge of, expertise in, or control over the technology infrastructure 'in the cloud' that supports them. Cloud computing describes a new supplement, consumption and delivery model for IT services based on Internet, and it typically involves the provision of dynamically scalable and often virtualized resources as a service over the Internet"[5].

B. A user's view

After using some services from a cloud platform, a common user maybe says, "No more servers, no more IT managers, no more licensing. All I need is a browser. I will find the service I need and pay it when I use it based how much I used it", i.e, Software is as a Service [17].

C. A Developer's view

In the opinion of a programmer with could computing, he concludes, "Programming an application will not be a 30 man/month job with a stupid and horrible project manager. It will be done by a couple of friends, in a long

[+]This work is supported by Shanghai Leading Academic Discipline Project, Project Number: J50103.

AP-PROC-CS-10CN006

hard working weekend with PHP on a machine God knows where it is'', i.e, Platform is as a Service [17].

D. A computer scientist's view

Scientists investigate cloud computing into "We are not programming a single machine, rather the World Wide Computer''. It means infrastructure is as a Service

E. An economist's view

Economically, "The most interesting thing about the cloud computing is not the technology, but the new evolving social standards and business models, and the ramifications of egalitarianism on a global scale''. Everything he thinks is as a Service.

Based on our understanding of the essence of what Clouds are promising to be, we propose that Cloud computing is a usage model in which resources are deliveries and, it means to provide resources, such as hardware, software and applications as scalable and "on-demand" services via public network in a multi-tenant environment. The providing resource network is called 'Cloud'. All resources in the 'Cloud' are used whenever as utility and scalable infinitely.

Ⅲ. THREE LAYERS OF CLOUD COMPUTING

Cloud computing is generally related to three different levels as follows.

A. Software as a Service: SAAS

Software as a Service delivers special-purpose software which is remotely accessible by consumers via the Internet with a usage-based pricing model. Salesforce Customer Relationships Management (CRM) system [7] and Google Apps [8] are two examples of SaaS. Moreover, Microsoft's Live Mesh allows files and folders to be shared, and synchronized across multiple devices.

B. Platform as a Service: PAAS

Platform as a Service provides a high-level integrated environment to design, build, test, deploy and update online custom applications. A typical example is Google's App Engine [9], which enables customers to build Web applications on the same scalable systems which power Google applications. Another example is SalesForce Apex language that allows the developers of the cloud applications to design, along with their applications' logic, their page layout, workflow, and customer reports. In this respect, a Hadoop [11] deployment on the cloud would be considered a cloud software environment, as it provides its applications' developers with a programming environment, i.e. map reduce framework for the cloud. Similarly, Yahoo's Pig [12], a high-level language to enable processing of very large files on the hadoop environment, may be viewed as an open-source implementation of the cloud platform layer.

C. Infrastructure as a Service: IAAS

Infrastructure as a Service offers hardware, software, and equipments which are mostly at the unified resource layer or part of the fabric layer, to deliver software application environments with a resource usage-based pricing model. Infrastructure can scale up and down dynamically based on application resource needs. Amazon's Elastic Compute Cloud (EC2 [13]), and Enomalism elastic computing infrastructure [14] are arguably the two most popular examples of commercial systems available in this cloud category. In this space, there are also several academic open-source cloud projects, such as Eucalyptus [15].

Some examples of data storage systems are: distributed file systems (e.g., GFS), replicated relational databases (RDBMS) and key-value stores. RDBMS, for example opt to present a stricter consistency model at the cost of the availability of the data, while key-value stores have placed more importance on the availability of the data while relaxing the consistency model for the storage.

Although user interface to the cloud infrastructure components varies substantially from one system to another, SOAP and REST are two examples of interface protocols used with some cloud computational resources.

Ⅳ. SOME CLOUD COMPUTING PLATFORMS

Industry analysts have made bullish projections on how Cloud computing will transform the entire computing industry.

A. Microsoft's Windows Azure platform

Microsoft's Windows Azure platform is a group of cloud technologies, each providing a specific set of services to application developers. This platform can be used both by applications running in the cloud and by applications running on local systems [1]. It contains the following components and is shown in Figure 1.

- **Windows Azure:** A Windows-based environment for running applications in Microsoft data centers**.**
- **SQL Azure:** SQL Server based data services.
- **.NET Services**: Distributed infrastructure services to cloud-based and local applications.

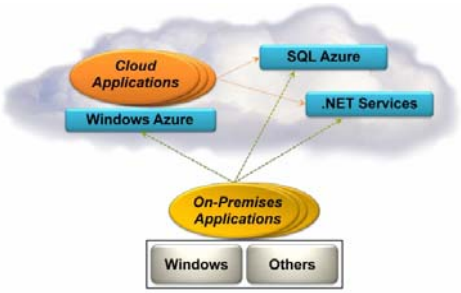


Figure 1. The Windows Azure platform

1) WINDOWS AZURE

Windows Azure runs on a large number of machines, all located in Microsoft data centers and accessible via the Internet. A common Windows Azure fabric knits this plethora of processing power into a unified whole. Windows Azure compute and storage services are built on top of this fabric.

2) SQL AZURE

SQL Azure will eventually include a range of data-oriented capabilities, including reporting, data analytics, and others, the first SQL Azure components to appear are SQL Azure Database and "Huron" Data Sync.

- SQL Azure Database: Unlike the Windows Azure storage service, SQL Azure Database is built on Microsoft SQL Server.
- "Huron" Data Sync: "Huron" Data Sync. technology synchronizes relational data across various on-premises DBMSs. The owners of that data can determine what should be synchronized, how conflicts should be handled, and more.

3) .NET SERVICES

.NET Services provides cloud-based infrastructure services that can be used by either on-premises applications or cloud applications. It includes Access Control and Service Bus. Service Bus provides a mechanism to manage Web services exposed by applications.

B. Amazon Web Services for Cloud Competing

Amazon Web Services (AWS) for Cloud Competing provides a highly scalable cloud computing platform which is high availability and dependability, and the flexibility to enable customers to build a wide range of applications [16]. Amazon Web Services for Cloud Competing contains the following components.

1) Amazon Elastic Compute Cloud (EC2)

Amazon Elastic Compute Cloud (Amazon EC2) shown in Figure 2 is a web service that enables customers to launch and manage Linux/UNIX/Windows server instances in the data centers of Amazon.

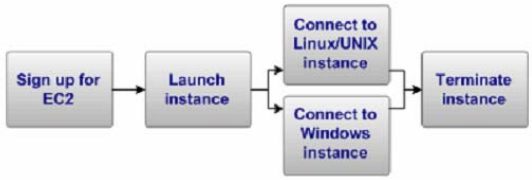


Figure 2. Amazon Elastic Compute Cloud (Amazon EC2)

Amazon EC2 provides the following major features:

a) Resource delivered as AMI (Amazon Machine Image).

b) Compute instance.

c) Explicit access control

2) Amazon Simple Storage Service (S3)

Amazon S3 enables people to upload, store, and download data via the Internet. The data can be used in conjunction with other AWS services, e.g. EC2, Amazon Import/Export, and Amazon Elastic MapReduce. The features of Amazon S3 are listed as follows.

a) Flat object storage model with key.

b) Bucket as object container.

c) Establish connection.

d) Create bucket.

e) Upload file.

f) Standards-based Interfaces: REST and SOAP with URL for each object.

3) Amazon Simple Queue Service (SQS)

SQS is a distributed queue that enables web service applications to queue messages quickly and reliably that one component in the application can be consumed by another component. The below describes the major features of Amazon SQS.

a) Redundant infrastructure.

b) Multiple writers and readers.

c) Configurable settings per queue.

d) Variable message size.

e) Unlimited queues and messages.

f) Access control.

4) Amazon SimpleDB (SDB)

Amazon SimpleDB is also a web service for running queries in real time on structured data. This service works in conjunction with Amazon S3 and Amazon EC2, providing collectively the ability to process, store and query data in the cloud. Amazon SimpleDB is designed to provide the following major features.

a) Simple to use.

b) Flexible.

c) Scalable.

d) Fast.

e) Reliable.

f) Works with other Amazon Web Services.

g) Inexpensive.

5) Amazon Flexible Payment Service

Amazon Flexible Payments Service is designed for developers in the form of the first payments service and it allows develop highly customized payment solutions for kinds of businesses. Amazon FPS service is extensive so that its functionality is divided into five Quick Start implementations below.

a) Amazon FPS Basic Quick Start

b) Amazon FPS Marketplace Quick Start

c) Amazon FPS Advanced Quick Start

d) Amazon FPS Aggregated Payments Quick Start

e) Amazon FPS Account Management Quick Start

6) Amazon DevPay

You may get paid for products that you build on either Amazon EC2 or the Amazon S3 through Amazon DevPay which is a bill and account management service. Amazon DevPay designs the order pipeline, billing and account system for you.

C. Google Cloud

1) Google Engine

Other than supporting the Python standard library, Google App Engine also supports Application Programming Interfaces (APIs) for the datastore, Google

Accounts, URL fetch, image manipulation, and email services [8].

2) Google Apps

Google Apps is one of the most sophisticated and comprehensive collaborative products available. The program includes applications for email, calendars, instant messaging, room reservations, document storage and editing and video sharing.

3) Google File system (GFS)

GFS provides a reliable distributed storage system that can grow to petabyte scale by keeping data in 64-megabyte "chunks" stored on disks spread across thousands of machines. Each chunk is replicated, usually 3 times, on different machines so GFS can recover seamlessly from disk or machine failure. Figure 3 shows its Architecture.

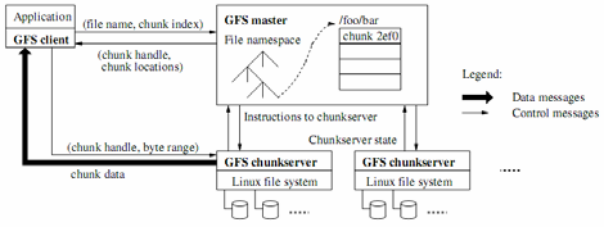


Figure 3. Google File System Architecture.

4) Work queue and MapReduce

Work queue provides a handling mechanism for scheduling a job to run on a cluster of machines. It schedules jobs, allocates resources, reports status, and collects the results. Its execution flow is demonstrated in Figure 4.

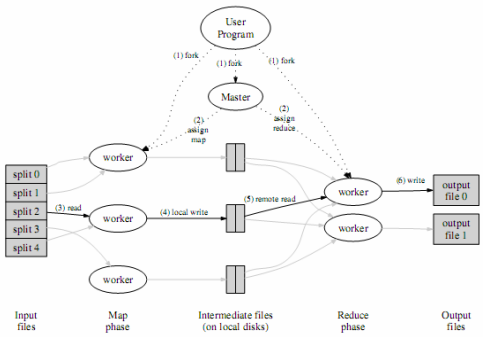


Figure 4. MapReduce Execution overview.

D. Eucalyptus

Eucalyptus is an open source software infrastructure for creating on-premise clouds on existing Enterprise. It is a service provider infrastructure and an elastic utility computing Architecture [14]. Eucalyptus supports the cloud interface of the popular Amazon Web Services, can work with kinds of hypervisors and virtualization technologies and allows on-premise clouds to interact with public clouds with the help of a common programming interface. Enterprise Eucalyptus can provide customized service level agreements (SLAs), metering, cloud monitoring, and supports for a highly scalable and available cloud platform.

Eucalyptus was designed modularly and its components were a set of web services interoperating using standard communication protocols. A description of the components within the Eucalyptus is briefly as follows.

1) Cloud Controller (CLC)
2) Cluster Controller (CC)
3) Node Controller (NC)
4) Storage Controller (SC)

Ⅴ. CONCLUSION AND TRENDS

In this paper, we have proposed Cloud computing paradigm from a variety of aspects, such as definitions, features, and technologies. Moreover, we have illustrated several representative platforms for the state-of-the-art Cloud computing.

Cloud computing needs to be extended to support negotiation of QoS based on Service Level Agreement (SLAs). The corresponding algorithms should be designed for allocation of VM resources to meet SLAs between service providers and end-users. The risks of the violation of SLAs must be managed effectively. Furthermore, we must extend some protocols to support interaction between different Cloud service providers.

REFERENCES

[1] Windows Azure Platform, v1.3—Chappell. http://www.microsoft.com/windowsazure/.
[2] Amazon Elastic Compute Cloud[URL]. http://aws.amazon.com/ec2/
[3] IBM Blue Cloud project. http://www-03.ibm.com/press/us/en/pressrelease/22613.wss/.
[4] Global Cloud computing test bed. http://www.hp.com/hpinfo/newsroom/press/2008/080729xa.html/.
[5] http://en.wikipedia.org/wiki/Cloud_computing#cite_note-0
[6] Lamia Youseff, Maria Butrico, Dilma Da Silva. Toward a Unified Ontology of Cloud Computing. GCE'08, 2008.
[7] Salesforce Customer Relationships Management (CRM) system, http://www.salesforce.com/.
[8] GOOGLE Apps, http://www.google.com/apps/business/index.
[9] GOOGLE App Engine, http://code.google.com/appengine.
[10] Apex: Salesforce on-demand programming language and framework, http://developer.force.com/.
[11] Hadoop, http://hadoop.apache.org/.
[12] C. Olston, B. Reed, et al., "Pig latin: a not-so-foreign language for data processing," Proceedings of the 2008 ACM SIGMOD international conference on Management of data. New York, NY, USA: ACM, 2008, pp. 1099-1110.
[13] Enomalism elastic computing infrastructure, http://www.enomaly.com.
[14] Eucalyptus systems, http://eucalyptus.cs.ucsb.edu/.
[15] S. Ghemawat, H. Gobioff, and S.-T. Leung, "The google file system," SIGOPS Oper. Syst. Rev., vol. 37, no. 5, pp. 29-43, 2003.
[16] "Amazon simple storage service," http://aws.amazon.com/s3/.
[17] Yike Guo. Introduction to Cloud Computing. TR. 11, 2009.

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 165-169

Weighted Suffix Tree Document Model for Web Documents Clustering

Ruilong Yang, Qingsheng Zhu, and Yunni Xia
College of Computer Science, Chongqing University, Chongqing, China
Email: ruilong.yangrl@gmail.com, qszhu@cqu.edu.cn, xiayunni@yahoo.com.cn

***Abstract*—A novel Weighted Suffix Tree Document (WSTD) model is proposed to construct Web document feature vector for computing the pairwise similarities of documents with weighted phrase. The weighted phrase-based document similarity is applied to the Group-average Hierarchical Agglomerative Clustering (GHAC) algorithm to develop a new Web document clustering approach. First, different document parts are assigned different levels of significance as structure weights. Second, the WSTD is built with sentences and their structure weights. Third, each internal node and its weights in WSTD model are mapped into a unique feature term in the Vector Space Document (VSD) model; the new weighted phrase-based document similarity extends the term TF-IDF weighting scheme in computing the document similarity with weighted phrases. Finally, the GHAC algorithm is employed to generate final clusters. The improvements of evaluation experiments indicate that the new clustering approach is more effective on clustering the documents than the approach which ignores the Web documents structures. In conclusion, the WSTD model much contributes to improving effectiveness of Web documents clustering.**

***Index Terms*—suffix tree, web document clustering, weight computing, phrase-based similarity, document structure**

I. INTRODUCTION

In an effort to keep up with the exponential growth of the WWW, many researches are targeted on how to organize such information in a way that will make it easier for the end users to find the information efficiently and accurately. One of these is document clustering [1-4] which attempt to segregate the documents into groups where each group represents a topic that is different than those topics represented by the other groups[4].

In general, the clustering techniques are based on four concepts [4, 5]: data representation model, similarity measure, clustering model, and clustering algorithm. Most of the current documents clustering methods are based on the Vector Space Document (VSD) model. The common framework of this data model starts with a representation of any document as a feature vector of the words that appear in the documents of a data set.

To achieve a more accurate document clustering, phrase has been considered as a more informative feature term in recent research work and literature [3-7]. A important phrase-based clustering, Suffix Tree Clustering (STC), is based on the Suffix Tree Document (STD) model which was proposed by Zamir and Etzioni[6]. The STC algorithm was used in their meta-searching engine to cluster the document snippets returned from other search engine in realtime. The algorithm is based on identifying the phrases that are common to groups of documents.

By studying the STD model, Chim and Deng [5] find that this model can provide a flexible n-gram method to identify and extract all overlap phrases in the documents. They combine the advantages of STD and VSD models to develop an approach for document clustering. By mapping each node of a suffix tree except the root node into a unique dimension of an M-dimensional term space (M is the total number of nodes except the root node). each document is represented by a feature vector of M nodes. With these feature vectors, the *cosine* similarity measure is employed to compute the pairwise similarities of documents which are applied to the Group-average Hierarchical Agglomerative Clustering (GHAC) algorithm [8].

However, when applying these clustering methods to Web documents, the characteristics of the Web documents are ignored. HTML tags are used to designate different parts of the document and identify key parts based on this structure[4].

A Weighted Suffix Tree Document (WSTD) model is proposed to develop a novel approach for clustering Web documents. This approach consists of four parts: (1) Analyze the HTML document and assign different levels of significance to different document parts as structure weights. (2) Represent different parts of the documents as some sentences with levels of significance. The sentence is represented as a sequence of words, not characters. These sentences instead of documents are used to build WSTD model. Except the root node, each node of the WSTD model contains document ID, sentence ID with its level of significance, and the traversed times of each sentence. (3) Map each node into a unique dimension of an M-dimensional term space. Each document is represented by a feature vector of M nodes. The statistical features and significance levels of all nodes are taken into account of the feature term weights and similarity measures. (4) Apply the documents similarities to GHAC algorithm to cluster the Web documents.

The rest of this paper is organized as follows: Web document structure analysis is presented in Section 2; Section 3 starts with the definition of the weighted suffix

Supported by National Key Technology R&D Program of China (No.2007BAH08B04), Chongqing Key Technology R&D Program of China (No.2008AC20084), CSTC Research Program of Chongqing of China (No.2009BB2203) and Post-doctorial Science Foundation of China (No. 20090450091).

AP-PROC-CS-10CN006

tree and WSTD model, and then presents the detailed design of the new weighted phrase-based document similarity; Section 4 illustrates the experimental results of evaluation for the new WSTD model, with the comparison of STD model; Section 5 summarizes the work.

II. Web Document Representation

For Web documents, one of three levels of significance is assigned to the different parts; HIGH (H), MEDIUM (M), and LOW (L)[4]. Examples of HIGH significance parts are the title, meta keywords, meta description, and section headings. Examples of MEDIUM significance parts are texts that appear in bold, italic, colored, hyper-linked text. LOW significance parts are usually comprised of the document body text that was not assigned any of the other levels. A formal model is presented that the document is represented as a vector of sentences, which in turn are composed of a vector of terms[4]:

$$d_i=\{s_{ij}: j=1,\ldots,p_i\} , \quad (1a)$$

$$s_{ij}=\{t_{ijk}:k=1,\ldots,l_{ij}; w_{ij}\} , \quad (1b)$$

where

d_i: is document i in documents set D,
s_{ij}: is sentence j in document i,
p_i: is the number of sentences in document i,
t_{ijk}: is term k of sentence s_{ij},
l_{ij}: is the length of sentence s_{ij}, and
w_{ij}: is the level of the significance associated with sentence s_{ij}.

III. Clustering Method Based on WSTD Model

The main steps of the new approach are constructing WSTD model, mapping each node into a unique dimension of an M-dimensional vector space, computing feature term weights, finally applying GHAC to clustering Web documents.

A. Definition of Weighted Suffix Tree

For a sentence s of n words, in order to build a suffix tree[6, 9], a nonempty character \$ is appended to the end of the sentence s with Ukkonen's algorithm[9]. The length of s is changed to n+1. When constructing the suffix tree, the sentences s is represented as a sequence of words, not characters. The kth suffix of a sentence s is the substring of s that starts with word k to the end of s.

Definition 1 The suffix tree T of the sentence s is a rooted, compact trie containing all the suffixes of s. Each internal node has at least 2 children. Each edge is labeled with a nonempty substring of s. The label (phrase from the node) of a node is defined to be the concatenation of the edge-labels on the path from the root to that node. No two edges out of the same node can have edge-labels that begin with the same word. For each suffix $s[k\ldots n+1]$ of s, there exists a suffix-node whose label equals it.

Definition 2 The weighted suffix tree T_s of the sentence s is based on a suffix tree of s, each node v of which records the sentence s, the level of significance w of s, and the times r that s traverses the node v. Its information is represented as a triple (s, r, w).

Definition 3 The weighted suffix tree T_d of document d_i is based on a suffix tree which p_i sentences are inserted into. Each node records a list of triple (s_{ij}, r_{ij}, w_{ij}).

Definition 4 The weighted suffix tree T_D of documents set D is based on a suffix tree which N documents are inserted into. Each document d_i contains p_i sentences. Thus, $\sum p_i$ sentences are inserted into T_D. Each node records a list of triple (s_{ij}, r_{ij}, w_{ij}).

Four properties are concluded from T_D.

Property 1 Any node v of the weighted suffix tree T_D has its depth level L_v. Each phrase P_v denoted by any internal node v contains at least two words at a higher level ($L_v \geq 2$). Phrases of various lengths can be extracted.

Property 2 Structure weights of any node appearing in different sentences or documents can be obtained from T_D.

Property 3 Summing up all times which a node is traversed by the sentences of a document, the node frequency, i.e., phrase frequency in a document, can be obtained. Phrase frequency is denoted by *tf*.

Property 4 Summing up all times which a node is traversed by different documents, the document frequency of a node can be obtained. Document frequency is denoted by *df*.

Fig.1 is an example of a suffix tree composed from two Web documents.

d_0: <title>cat ate cheese. </title><body>mouse ate cheese too. </body>

d_1: <bold>cat ate mouse too. </bold>

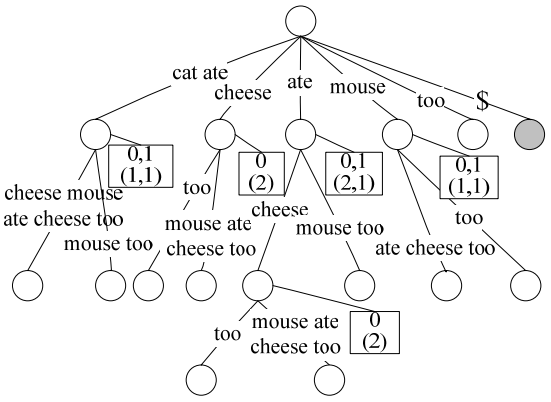


Figure 1. Suffix tree of documents set.

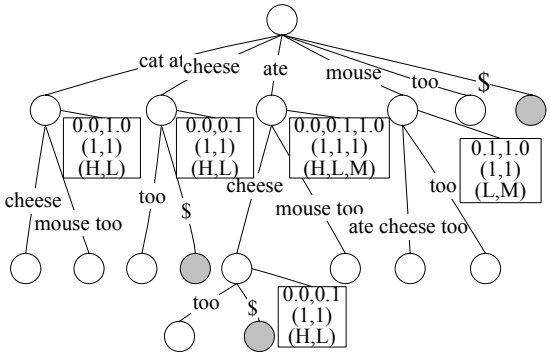


Figure 2. Weighted suffix tree of documents set.

The nodes in the suffix tree are drawn in circles. There are three kinds of nodes: the root node, internal node, and leaf node. Each internal node represents a common phrase shared by at least two suffix substrings. Each internal node is attached to an individual box. Each upper number in the box designate a document identifier, the number below designates the traversed times of the document.

Fig.2 is an example of a weighted suffix tree composed of three sentences with levels of significance which originate from the two documents d_0 and d_1 partitioned by HTML tags.

Each internal node is attached to an individual box. Each upper number designates a sentence identifier which contains document identifier in the left of the period mark, the middle number designates the traversed times of the corresponding sentence, the number below designates the level of significance of corresponding sentence.

The length of phrases extracted from weighted suffix tree is much less than from ordinary suffix tree. From the weighted suffix tree, structure weight *w*, *tf*, and *df* of each node can be obtained.

B. Similarity Based on the Weighted Phrases

In this paper, the symbols *N*, *M*, and *q* are used to denote the number of documents, the number of feature terms, and the number of clusters, respectively. The C_1, $C_2,\ldots,C_q$ are used to denote each one of the *q* clusters of result.

Each document *d* is considered to be a vector in the *M*-dimensional feature term space. In this paper, the TF-IDF weighting scheme is employed, in which each document *d* can be represented as[5]:

$$\vec{d} = \{w(1,d), w(2,d),..., w(M,d)\} \ , \qquad (2a)$$

$$w(i,d)=(0.1+\log tf(i,d))\times\log(0.1+N/df(i)) \ , \qquad (2b)$$

where $w(i,d)$ is *i*th feature term weight in document *d*, $tf(i,d)$ is the frequency of the *i*th term in the document *d*, and $df(i)$ is the number of documents containing the *i*th term.

In the VSD model, the *cosine* similarity is the most commonly used to measure and compute the pairwise similarity of two document d_i and d_j, which is defined as:

$$sim(d_i, d_j) = \frac{\vec{d}_i \bullet \vec{d}_j}{\left|\vec{d}_i\right| \times \left|\vec{d}_j\right|} \ . \qquad (3)$$

With respect to WSTD model, a node *v* can appear in different sentences with its level of significance. When mapping a node *v* of WSTD model to VSD model, besides $df(v)$ and $tf(v,d)$, the feature term weight depends on the level of significance of *v*.

The symbols $tf^H(v,d)$, $tf^M(v,d)$, and $tf^L(v,d)$ are used to denote the term frequency of the node *v* in HIGH, MEDIUM and LOW significance parts of a document respectively. The term frequency $tf(v,d)$ of node *v* is updated to:

$$tf(v,d)=w^H\times tf^H(v,d)+w^M\times tf^M(v,d)+w^L\times tf^L(v,d) \ , \qquad (4)$$

where w^H, w^M, w^L are used to denote different level of significance of a node *v*, w^L is assigned 1, and the others are greater than w^L. The feature term weight of node *v* is represented as:

$$w(v,d)=(0.1+\log tf(v,d))\times\log(0.1+N/df(v)) \ . \qquad (5)$$

C. Computation Complexity of the Novel Approach

Ukkonen[9] provides an building suffix algorithm for a document of *m* words with time complexity of $O(m)$. In WSTD model, the length of a document *m* is assumed to be bounded by a constant. Therefore, the time complexity of building WSTD model is $O(N)$.

The time complexity of using GHAC algorithm to cluster Web documents is $O(N^2\log N)$[8]. For the dimensionality *M* of features vector usually is large, similarity computations are time consuming.

As a result, the overall time complexity of the proposed approach is $O(N^2\log N)$.

IV. Experimental Results

A. Experimental Setup

Three data sets are used, two of which are Web document data sets, and the third is a collection of the search engine results called AMBIENT[10] in which each document includes a title and a snippet. Table 1 describes the data sets. Documents set DS1 and DS2 are described and used in [4].

The proposed approach is implemented in Java with JDK 6. For the weighted suffix tree construction, an open source library of the Carrot2[11] is used. But there exist some errors and defects for counting the times which documents traverse nodes. These errors are corrected in the work and Carrot2 is extended with levels of significance of nodes for WSTD model. The GHAC algorithm is implemented in Java by following the description in [8].

B. Evaluation Metrics

In order to evaluate the quality of the clustering, two quality measures are adopted, which are widely used in the text mining literature for the purpose of document clustering [2-5]. The first is the *F*-measure, which combines the *Precision* and *recall* ideas in information retrieval. The second measure is the *Entropy*, which tells how homogeneous a cluster is. The higher the homogeneity of a cluster, the lower the entropy is, and vice versa.

To sum up, the approach maximizes the *F*-measure and minimizes the *entropy* score of the clusters to achieve a high-quality clustering.

TABLE I. Data Sets Descriptions

Data Sets	Name	Type	#Docs	Categories	Avg. words
DS1	UW-CAN	HTML	314	10	469
DS2	Yahoo news	HTML	2340	20	289
DS3	AMBIENT	TEXT	4400	44	41

C. Comparison of STD and WSTD Model

In experiments, GHAC algorithm is chosen for some reasons in [5], which measures the similarity of two clusters with the average of pairwise similarities of the documents from each cluster. The effectiveness of two document models is evaluated with GHAC algorithm in the comparison experiments: STD model and WSTD model.

In the comparison experiments, two subsets of each original data set are selected to be six small data sets: DS1-1, DS1-2, DS2-1, DS2-2, DS3-1 and DS3-2, as illustrated in Table 2, where #Cates denotes the number of categories, #Docs denotes the number of documents. For WSTD model, coefficient w^H is 2 times w^L, w^M is 1.5 times w^L. The stop criterion of GHAC is adjusted so that the number of clusters is equal to the number of the original class set of each data set.

Table 3 and Table 4, respectively, illustrates the *F*-measure and *Entropy* scores computed from the clustering results of GHAC algorithm with STD and WSTD model. Table 3 illustrates the improvement on the *F*-measure score, and Table 4 illustrates the improvement on the *Entropy* score.

The improvements on the *F*-measure and *Entropy* scores indicate that the new weighted phrase-based document clustering approach based on WSTD model tends to be a highly accurate documents clustering approach.

D. The Performance Evaluation

The time cost and the nodes mapped into the vector space are shown in Table 2, where #Nodes denotes the number of nodes mapped into vector space, #All denotes the total time of the approach executed, and #GHAC denotes the time of GHAC algorithm except the time of building WSTD model. The unit of time is millisecond (ms). Comparing #All and #GHAC, it is concluded that the main time cost is caused by GHAC algorithm, and constructing the WSTD model needs a little time.

TABLE II. DATA SETS NODES AND TIME

Data Sets	#Cates	#Docs	#Nodes	#All (ms)	#GHAC (ms)
DS1-1	5	100	11162	18344	16219
DS1-2	5	142	20969	64671	60515
DS2-1	5	100	6422	10313	9172
DS2-2	5	200	13894	115656	112563
DS3-1	5	100	439	985	844
DS3-2	5	220	1109	11641	11454

TABLE III. IMPROVEMENT ON F-MEASURE

Data Sets	STD	WSTD	Improvement
DS1-1	0.899	0.96	6.7%
DS1-2	0.957	0.977	2.1%
DS2-1	0.542	0.702	29.4%
DS2-2	0.569	0.614	7.8%
DS3-1	0.817	0.909	11.3%
DS3-2	0.821	0.915	11.5%

TABLE IV. IMPROVEMENT ON ENTROPY

Data Sets	STD	WSTD	Improvement
DS1-1	0.112	0.065	42.2%
DS1-2	0.064	0.040	37.4%
DS2-1	0.419	0.318	24.2%
DS2-2	0.368	0.358	2.6%
DS3-1	0.254	0.178	29.8%
DS3-2	0.185	0.157	15.2%

E. The Variable Length of Phrases

In WSTD model, a document is partitioned into some sentences. The length of a sentence is assumed to be bounded by a constant. The Phrases extracted from WSTD model are also bounded a constant. As described in [4, 6], the length of phrases less than 6 is appropriate and efficient.

The total number of six lengths of phrases (from 1 to 6) accounts for about 99 percent of the amount of all nodes mapped into VSD model on DS2-1 and DS3-1.

The distributions of six lengths of phrases (from 1 to 6) indicate that the most efficient phrases can be extracted from the WSTD model.

V. CONCLUSIONS

By mapping all internal nodes in the STD model into VSD model, phrase-based similarity successfully connects the two models.

For Web documents, different parts are assigned a level of significance. A document is partitioned into some sentences with significance; a sentence is represented as a sequence of words. These sentences are used to build the weighted suffix tree instead of documents. The STD model is extended to the WSTD model with the weighted nodes. Thus, the weighted phrase-based document similarity is proposed. The significant improvement of the clustering quality in experiments clearly indicates that the new clustering approach with the weighted phrase-based document similarity is more effective on clustering the Web documents.

Using sentences to build WSTD model, most efficient phrases can be extracted from the WSTD model which is used as an n-gram technique to identify and extract weighted phrases in Web documents.

ACKNOWLEDGMENT

This work was partly supported by the National Key Technology R&D Program of China (No.2007BAH08B04), the Chongqing Key Technology R&D Program of China (No.2008AC20084), CSTC Research Program of Chongqing of China (No.2009BB2203) and Post-doctorial Science Foundation of China (No. 20090450091).

REFERENCES

[1] N. Oikonomakou, and M. Vazirgiannis, "A Review of Web Document Clustering Approaches," Data Mining and Knowledge Discovery Handbook, pp. 921-943: Springer US, 2005.

[2] L. Yanjun, “Text Clustering with Feature Selection by Using Statistical Data,” IEEE Transactions on Knowledge and Data Engineering, vol. 20, pp. 641-652, 2007.

[3] Y. Li, S. M. Chung, and J. D. Holt, “Text Document Clustering Based on Frequent Word Meaning Sequences,” Data & Knowledge Engineering, vol. 64, no. 1, pp. 381-404, 2008.
[4] K. M. Hammouda, and M. S. Kamel, “Efficient Phrase-Based Document Indexing for Web Document Clustering,” IEEE Transactions on Knowledge and Data Engineering, vol. 16, no. 10, pp. 1279-1296, 2004.
[5] H. Chim, and X. Deng, “Efficient Phrase-Based Document Similarity for Clustering,” IEEE Transactions on Knowledge and Data Engineering, vol. 20, no. 9, pp. 1217-1229, 2008.
[6] O. Zamir, and O. Etzioni, "Web Document Clustering: A Feasibility Demonstration," Proceedings of the 21st International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 46-54, 1998.
[7] S. Zu Eissen, B. Stein, and M. Potthast, “The Suffix Tree Document Model Revisited,” in Proceedings of the 5th International Conference on Knowledge Management (I-KNOW 05), Graz, Austria, 2005, pp. 596-603.
[8] C. Manning, P. Raghavan, and H. Schütze, "An introduction to information retrieval," p. 377~400, Cambridge,England: Cambridge University Press, 2009.
[9] E. Ukkonen, “On-Line Construction of Suffix Trees,” Algorithmica, vol. 14, no. 3, pp. 249-260, 1995.
[10] C. Carpineto, and G. Romano. "Ambient Dataset," 2008; http://credo.fub.it/ambient/.
[11] S. Osiński, and D. Weiss, “Carrot 2: Design of a Flexible and Efficient Web Information Retrieval Framework,” Advances in Web Intelligence, vol. 3528, pp. 439-444, 2005.

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 170-173

Comprehensive Evaluation of Drilling Risk Based on Unascertained Measurement

Jun'e Liu[1], and Rong Wang[2]
[1] Information school , Beijing Wuzi University, Beijing 101149, China ;
Email: zl-je@163.com
[2] School of Economy and Management, Hebei University of Engineering, Handan 056038, China
Email: wangrong246437@sohu.com

***Abstract*—Drilling engineering is not only a systematic engineering that involves several workcategories, work procedures,intersectant and continuous operation, but also is a concealed underground engineering. There are uncertain factors in drilling projects, which have the characteristics of high investment and risk. On the basis of the overall analysis of the two factors of geology and engineering, the risk evaluation model of drilling project is established by creating the hierarchy structure of drilling risk and introducing the uncertain measure comprehensive evaluation method. Taking Yanchang Oilfield 21 Well as an example, a comprehensive evaluation is studied and then draw a reliable conclusion. Therefore, unascertained measure evaluation method is applicable, efficient, and necessary to promote safety production, which provides the new way for risk assessment of drilling project..**

***Index Terms*—drilling project, risk evaluation, unascertained measure**

I. INTRODUCTION

Drilling is a covert and underground work, which is full of fuzzinesses, randomness and uncertainty, so it is truely a kind of high-risk work. The object of drilling is the stratum rock, and the goal is the oil and gas exploration. In drilling operation process, the lack of understanding to rock buried deeply in earth crust (objective factor) or technical factor (engineering factor) as well as decision-making fault (human factor) tend to bring about many complex mine situations and even serious downhole accidents that consume effort, money and time and may even lead to the waste of oil and gas resource as well as the abandonment of the mining site. According to analysis of drilling data in recent years, the time of solving complex mine situations and drilling accidents approximately accounts for 5%~8% of the total drilling time. Dealing with complex mine situations caused by geologic agent correctly and avoiding or reducing drilling accidents caused by decision fault have important significance to guarantee safety production, staff health and environmental protection[1]. Therefore, combined with engineering characteristics of drilling, comprehensive evaluation of drilling risks is made on the basis of analysis of single-factor,so as to provide a basis for scientific decision-making.

II. RISK RECOGNITION OF DRILLING

On the basis of the overall analysis of survey results of local drilling environment; drilling activities prone to accidents; routine management experience; people's behavior, physical condition and environmental factors etc., we carries on the distinction to various risks in a latent state in drilling implementation process. Only identificating drilling risks accurately, knowing types, characteristics and potential hazard of risk incident can we take the necessary steps to guard against.

There are a large number of factors that make drilling complex and dangerous, but in summary , of which mainly are two categories---geological factors (the objective conditions), and engineering factors (subjective decision-making) [2] .

A. Geological factors

In terms of oil and gas drilling projects, the geological factors include rock types and its nature of engineering mechanics , geological structure, formation fluid situation and stratigraphy circumstances and so on. Taking into account the complexity of geological factors and the extent of indicators quantified, geological factor is divided to formation factor and sulfur-containing gases.

*1) Formation factor :*Drilling is a process of drilling gradually by constantly breaking formation rock. Drilling process and its risk factors closely related to the mechanical properties of rock and formation pressures, which directly influence drilling efficiency and technological measures, particularly the balance between formation pressure and bottom-hole pressure. In oil and gas drilling, we should maintain balance between formation pressure and bottom-hole pressure (static liquid column pressure). Otherwise, many complex situations or accidents such as well-kick, blowout, lost circulation, fracturing formation are likely to happen, so as to make drilling difficult to carry out, and even make borehole abandoned.

2)Sulfur-containing gases： Sulfur-containing gases mainly refers to the hydrogen sulfide gases encountered in the process of drilling. The sources of hydrogen sulfide can be attributed to the following kinds:： ①sulfate in the stratum produces hydrogen sulfide under high-temperature reduction; ②decomposition of the sulfur-containing in oil compounds and resultes in hydrogen sulfide; ③hydrogen sulfide in deep crust migrate and accumulate upwards through cracks; ④decomposition of

some drilling fluid produces hydrogen sulfide under the action of high temperature.

Hydrogen sulfide is a gas of colorless, toxic, strong acid. Hydrogen sulfide gas in low volume fraction has a smell of rotten eggs, the relative density of which is 1.19. Compared with air, it is heavy, and it can dissolve in water. When burning, it accompanied by a blue flame and produces sulfur dioxide gas that is very harmful to the person. The mixture of hydrogen sulfide and air, while mixture gas volume fraction lie the range of 4.3% ~ 45% , form a kind of explosive mixture. Hydrogen sulfide is easy to gather, not easy to drift, often gathering at the bottom of drilling platform or the lower well site. It can react with many metals and can corrode of metal severely. When volume fraction of hydrogen sulfide gas in the air is greater than 10 × 10-6 , people feel discomfort, and when that is more than 20 × 10-6, may cause poisoning.

In general, volume fraction of hydrogen sulfide increases with increasing formation depth. The presence of hydrogen sulfide is in two forms, either in the form of hydrogen sulfide gas alone or coexisting with oil and gas; or in the form of dissolving in water and oil. In drilling process, the hazards of hydrogen sulfide can be summed up in four aspects: ① harm on the human body; ② corrosion of drilling equipment; ③ contamination of drilling fluid; ④ pollution of the environment[3].

B. Engineering factors

Geological factor is the objective reality and can not be changed. If we have better understanding about it and take corresponding countermeasures, we can turn danger into safety, reduce the complexity of drilling, and avoid the accident. Engineering factors is a specific manifestation of subjective factors, that is construction design, technical operations, craft measures and so forth and is suitable for geological factors. It is human factors, modifiable and adjustable. For engineering factors, see Table1.

TABLE 1 MAIN ENGINEERING FACTOR

No.	Project	Main technical requirements	Function
1	Well profile	Tube sealing different pressures formation and instability stratum	Anti-collapse,anti-card,blowout prevention, leak-proof
2	Drilling equipment	Drill-pump displacement adjustable, having sufficient power.	Clean bottom, purify shaft, anti-card
		turntable speed adjustable.	Anti-limp, anti-broken
		Top drive drilling system	Timely process of downhole complex situations and reduce sticking
		Solid control intact, and the handling capacity meet the requirements.	Reduce the solid content
3	Well-control equipment	Pressurelevels and formation pressure match , and pressure test pass.	Blowout prevention, throttling
4	Drilling fluid	Select the appropriate type and performance parameters according to lithology and pressure.	Anti-collapse, anti-card, improve drilling speed.
5	Drilling tool structure	Select drilling tool structure and downhole tools according to lithology, dip, and drilling technology	Anti-oblique,anti-broken,anti-card anti-vibration, anti-out.
6	Drilling instrument	Reflect the drilling parameters comprehensively and accurately.	Provide real underground dynamic data, information and judge underground situation timely
7	Bit selection	Select bit type and drilling parameters according to stratigraphic drillability	Improve the penetration rate, anti-lost, anti-jump
8	Operating techniques	Technical operation procedures and standards in drilling.	Prevent operator errors and irregularities, and aviod the underground situation complicated or cause more accidents.

III. COMPREHENSIVE EVALUATION OF DRILLING RISK AND MODEL BUILDING

Comprehensive evaluation of drilling risk is that the organization of professionals, based on existing professional experience, evaluation criteria and guidelines , identify possible risks or impacts at pre-drilling activities, evaluate and analyze scientifically, calculate the probability ,value and consequences of the risk by quantifying risk factors, and thus determine whether the project risk may be accepted. Comprehensive evaluation methods mainly are analytic hierarchy process, decision tree method, monte-carlo method, fuzzy comprehensive evaluation etc. In this paper, unascertained measure evaluation method is adopted for analyzing and processing, and its evaluation results can be quantified accurately. The specific process is as follows:

A. Establish hierarchy structure of drilling risk

Establish hierarchy structure of risks by risk identification, like Figure1. Target layer denotes the object evaluated. Criterion layer is factor categories that evaluation object contains, that is two elements of drilling risk assessment. Each element is a relatively independent entity, which can be used as an object for study. Index layer is indicators that various factors included.

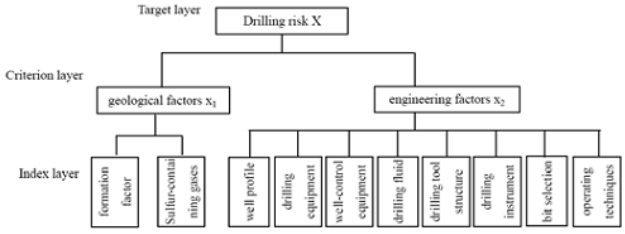


Figure 1. Drilling risk hierarchy structure diagram

B. Determine the weight of each factor

In drilling project, the importance of each evaluation factor is different, therefore we should determine their weighted coefficient according to relationship between primary and secondary to risk effect. In this paper, AHP method is adopted to determine the weight set of evaluation factors, that is to have experienced experts to make a comparison with the importance of two indexes to construct a consensus judgment matrix, then to get the weight value using the group decision-making coordinate method.

C. Establishment of unascertained measure evaluation model

In comprehensive evaluation of drilling engineering, analyzing and evaluating risk factors of system quantitatively and intuitively is an important part in theory circle. As drilling process involves a lot of uncertain information, use unascertained mathematics dealing with this issue.

1) Single-index measure modle of drilling risk assessment： In Figure1, X indicates object space,then: X =(x_1, x_2), thereinto x_1, x_2 indicate geological factors and engineering factors respectively. To evaluate the object need to identify m indexes, if x_{ij} indicates measured value of object x_i with regard to index I_j , then x_i can be expressed as m dimensional vector: x_i = $(x_{i1}, x_{i2}, \cdots, x_{im})$;then $x_1 = (x_{11}, x_{12}, x_{13})$, $x_2 = (x_{21}, x_{22}, \cdots, x_{28})$.

For x_{ij} ,if it has p evaluate grades and evaluation grade space denoted by U, then U ={c_1,c_2,⋯,c_p}, if c_k is "higher" than c_{k+1} , denoted by $c_k > c_{k+1}$ (k=1,2,⋯, p−1) . In the drilling risk evaluation, the evaluation grade can be divided into 5 grades, i.e. low risk, lower risk, medium risk, higher risk, high risk. The evaluation set is as follows: {c_1,c_2,c_3,c_4,c_5}={ low risk, lower risk, medium risk, higher risk, high risk }. Suppose $\mu_{ijk} = \mu\left(x_{ij} \in c_k\right)$ indicates measured value x_{ij} has the degree of evaluation level c_k ,Then μ is called to be unascertained measure function, measure for short. Calling the matrix:

$$\left(\mu_{ijk}\right)_{m\times p} = \begin{bmatrix} \mu_{i11} & \mu_{i12} & \cdots & \mu_{i1p} \\ \mu_{i21} & \mu_{i22} & \cdots & \mu_{i2p} \\ \vdots & \vdots & \vdots & \vdots \\ \mu_{im1} & \mu_{im2} & \cdots & \mu_{imp} \end{bmatrix}$$

is single-index evaluation matrix.

2) Structure unascertained measure function : Geological factor includes two second-level indexes. In order to evaluate facilitately and make evaluation results more accurate, take representative data of each indicator as a classification criteria of evaluation level. See Table 2.

TABLE 2 GRADE CLASSIFICATION OF INDEXES ON GEOLOGICAL FACTORS

Index \ Category	c_1	c_2	c_3	c_4	c_5
Formation factor	1	3	5	7	9
Sulfur-containing gases V	<$14*10^{-6}$	$14*10^{-6}$-$3*10^{-3}$	$3*10^{-3}$-$10*10^{-3}$	$10*10^{-3}$-$50*10^{-3}$	>$50*10^{-3}$

Note: In the table, data of formation factor is the relative risk index.

Single-index measure functions of two indexes of geological factor like Figure 2 and Figure3 respectively.

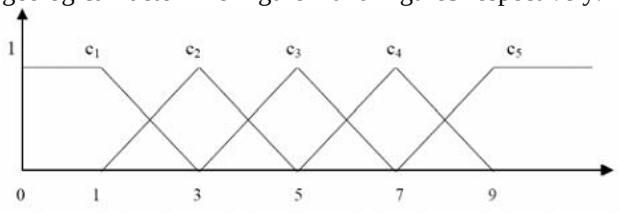


Figure 2. Single-index measure function of formation factor

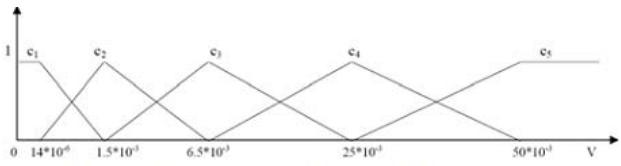


Figure 3. Single-index measure function of sulfur-containing gases

In addition to geological factors, the indexes contained in engineering factors are qualitative factors, using experts scoring method to determine risk value of each index. Approach is to get indexes assignment interval [0,100] and numbers in interval indicate security status of indexes, moreover the higher score standing for higher level of security. Inserting four points in the interval [0,100], thus the interval is divided into five risk levels, i.e. low risk, lower risk, medium risk, higher risk, high risk. See Table 3. Single-index measure function of engineering factors like Figure 4.

TABLE 3 GRADE CLASSIFICATION OF INDEXES ON ENGINEERING FACTORS

Classification Level	C_1	C_2	C_3	C_4	C_5
Index value	>90	90-80	80-70	70-60	<60

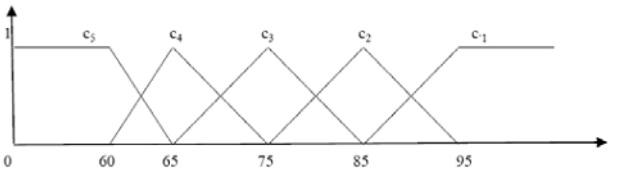


Figure 4. Single-index measure function of engineering factors

According to the characteristics of specific issues as well as variation degree of state ,we can assume different curves ,for example linear, quadratic curve, sine curve or exponential etc. ,to construct the specific expression of measure function .

3) Multi-index measure: Suppose that w_j indicates important degree of I_j relative to other index, that is to be weight of index I_j .If μ_{ik} satisfies: $0 \le \mu_{ik} \le 1$, $\mu_{ik} = \sum_{j=1}^{m} w_j \mu_{ijk}$, μ_{ik} is unascertained measure,

$$\left(\mu_{ik}\right)_{n\times p} = \begin{bmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1p} \\ \mu_{21} & \mu_{22} & \cdots & \mu_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ \mu_{n1} & \mu_{n2} & \cdots & \mu_{np} \end{bmatrix}$$

is multi-index measure matrix of comprehensive evaluation.

4) Confidence criteria for identification: In order to make final evaluation results of object evaluated, introduce the evaluation criteria of "confidence". Suppose λ indicate confidence（take λ =0.6 or 0.7）, $k_0 = \min\left(k : \sum_{l=1}^{k} \mu_l > \lambda, k = 1,2,\cdots, p\right)$

then judge drilling safety belonging to kth evaluation grade[4-6].

IV. PRACTICAL APPLICATION

Taking Yanchang Oilfield 21 Well as an example, unascertained measure model is adopted to evaluate comprehensively. The type of this well is straight, designedd well depth is 2776m.

Engineering factors x2 contains eight indicators. Through expert analysis of engineering factor of this well, measured values of eight indicators is gained. See Table 4.

TABLE. 4 ESTIMATES DATA TABLE OF EVALUATION INDEXES

Index	well profile	drilling equipment	well-control equipment	drilling fluid	drilling tool structure	drilling instrument	bit selection	operating techniques
Value	82	75	83	68	79	95	88	85

Substituting index values into single-index measure function of Figure 4. For example, x21=82 is substituted into the equation $y = \frac{1}{75-85}(x-85)$ and gain that index measure value is 0.3, that is, safety degree of well profile belonging to c3 is 0.3, then according to

"normalization", whose membership belongs to c4 being 0.7. Consequently, single index measure vector of well profile is （0 0 0.3 0.7 0）, and gain multi-index measure matrix of engineering factors by similarly algorithm:

$$(\mu_{ik})_{5\times 8} = \begin{bmatrix} 0 & 0.7 & 0.3 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0.8 & 0.2 & 0 & 0 \\ 0 & 0 & 0.3 & 0.7 & 0 \\ 0 & 0.4 & 0.6 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0.3 & 0.7 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

The weight vector of eight indicators by calculating is w_{x2} = (0.08, 0.15, 0.15, 0.2,0.12, 0.11,0.1,0.09), and there is comprehensive evaluation vector of engineering factors: $\mu_2 = \mu_{x2} \times w_{x2}$ =（0.14, 0.384, 0.336, 0.14, 0）. C Similarly, index weight vector and comprehensive evaluation vector of geological factor can be calculated, and there is :

$$\mu_1 = w_{x1} \times \mu_{x1} = (0.5, 0.5) \times \begin{bmatrix} 0 & 0.72 & 0.28 & 0 & 0 \\ 0.8 & 0.2 & 0 & 0 & 0 \end{bmatrix}$$

$$=(0.4,0.46,0.14,0,0)$$

In target layer, the weight vector of geological factor and engineering factor is w= （0.3, 0.7）, and comprehensive evaluation vector of target layer is:

$$\mu_x = w \times \mu = (0.3, 0.7) \times \begin{bmatrix} 0.4 & 0.46 & 0.14 & 0 & 0 \\ 0.14 & 0.384 & 0.336 & 0.14 & 0 \end{bmatrix}$$

=(0.218, 0.4068, 0.2772, 0.098, 0)

Taking confidence $\lambda = 0.6$, identify evaluation levels and there is 0.218+0.4068.>0.6, so that we can determine risk rating of this drilling project is c_2 , i.e. lower risk.

Ⅴ. Conclusion

There is a variety of risks in drilling project, not only needing to analyze each factor, but also needing comprehensive evaluation on the basis of various factors. Risk comprehensive evaluation based on unascertained measure method is feasible, which not only refer to the subjective experience of experts, but realize the quantitative evaluation according to mathematical models, and evaluation results can provide decision support for drilling projects. Taking various risk factors into account before decision-making of drilling program, and modifying the program constantly according to the actual situation, we can controll the possibility of risk at minimum level. The example show that this method is scientific, rational, efficient and applicable with wide application prospect[7].

Acknowledgment

Funding project for Academic Human Resources Development in Institutions of Higher Learning under the Juristiction of Beijing Municipality. PHR(IHLB)).

Logistics and Supply Chain Management Research. PHR 200907134.

References

[1] LI Qi, YU Linlin. Comprehensive Evaluation Method of Drilling Risk Factors and Model Building [J]. Processing, Utilization, and Environmental Protection，2008(5): 120-122 .

[2] DONG Guoyong. HSE Risk Management of Drilling Operations [M] .Beijing: Petroleum Industry Press，2001.

[3] ZHAO Junping. Risk Analysis and Management Research of Oil and Gas Drilling Project [D] . Heilongjiang: Daqing Petroleum Institute，2007.

[4] GAO Sanqiang. Risk Analysis Methods of Oil and Gas Field Development Project [J]. Oil Drilling Technology，2007(2):87-91.

[5] THOMASL,SAATY.Making and Validating Complex Decisions with the AHP/ANP[J] Journal of Systems Science and Systems Engineering，2005(1).

[6] KANG Jian, HOU Yunbing, SUN Guangyi. Stability Analysis of the Coal Seam Roof Based on Unascertained mathematics [J]. Journal of Heilongjiang Institute of Technology，2007，17（2）157-159.

[7] GONG Fengqiang，LI Xibing. Risk Evaluation of Gob Area Based on the theory of unascertained measure [J] . Journal of Rock Mechanics and Engineering，2008(2): 323-329.

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 174-177

Spatial Information Based Medical Image Registration Using Mutual Information

Benzheng Wei[1,2*], Zaitao Liu[1], and Xin Peng[1]
[1] College of Science and Engineering, Shandong University of Traditional Chinese Medicine, Jinan, 250355,China
Email: wbz99@sina.com
[2] College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing , 210016,China
Email: 6and5@sdutcm.edu.cn, pengxin0531@163.com

***Abstract*—Image registration is a problem that arises in many image processing applications whenever information from two or more scenes has to be aligned. A new approach to the problem of multimodality medical image registration is proposed, using a basic concept from spatial information, mutual information, and relative entropy, as a new matching criterion. Firstly, the feature characteristics like location, edge strength and orientation are taken into account to compute a joint probability distribution of corresponding edge points in two images. Then the mutual information based on this function is minimized to find the best alignment parameters. Finally, the translation parameters are calculated by using Powell algorithm. The experiment results showed that the method have achieved a good performance.**

***Index Terms*—medical image, image registration, Spatial information, Feature characteristics, Mutual information**

I. INTRODUCTION

Image registration is a valuable technique for medical diagnosis and treatment. In image registration the use of an adequate measure of alignment is a crucial issue. Current techniques are classified in two broad categories: area based and feature based. All methods include some similarity measure. Mutual information (MI) is currently a popular registration method to sca1e the similarities between two image sets and for convenience of calculation and analysis [1] . It was first proposed as a registration measure in medical image registration in 1995 . independently by Viola and Wells [2]and by Collignon [3] . Now MI has been accepted by many researchers as one of the most accurate and robust retrospective registration methods. However, mutual information is not a panacea method. Despite the general promising results, mutual information-based matching can result in misregistration when the images are of low resolution or the images contain little information or there is only a small region of overlap [5]. The mutual information registration function can be ill-defined, containing local maxima. A possible solution to reduce the failure of MI registration can be by using the spatial information and optimization algorithm . At present, many researchers have proposed some hybrid mutual information algorithms to enhance the performance of image registration, such as Pluim et al [1] combined mutual information with gradient method, Yao Yucui et al [6,7,8] employed mutual information with morphological gradient approach to improve the registration accuracy and robustness.

In this paper, a new approach to the problem of multimodality medical image registration is proposed, using a basic concept from spatial information, mutual information, and relative entropy, as a new matching criterion. In addition, this paper focus on the registration approach of multimodality image and a modified particle swarm optimization algorithm is proposed to restrain local maxima of mutual information function and improve the registration accuracy.

This paper is structured as follows: In Section 2, we present the mathematical model of spatial information detection approach. Section 3 describes the coarse registration based on the principal axes algorithm. The feature point set extraction approach and energy function of mutual information based the spatial information is presented in Section 4. Then in Section 5, the feasibility of our approach is illustrated where we show registration results for medical 2D examples. Finally, this paper is closed in section 6 with a brief conclusion.

II. CONTOUR FEATURE DETECTION

Contour feature is a significant part of information extracted from an image which provides more detailed understanding of the image. Common features include corresponding points, edges, contours or surfaces. In order to get a more accurate contour of the medical images, it's necessary to eliminate the interference of image noise and get the actual size of the tilt image as much as possible. According to those edge detection operators, such as Sobel, Log, Candy and other conventional edge detection operator, usually get a low positioning accuracy and a poor performance of anti-noise for medical image. Mathematical morphology is a new image processing method based on geometry algebraic and topology, which is proposed by French scholar G. Matherom and J. Seera [8]. It set the object-oriented for research, and the objects can be contacted through the structural elements. Expansion and corrosion is the basic operations of the morphology, and the opening and closing operators can be further expanded by the expansion and corrosion operators.

The morphology gradient is a nonlinear difference operator determinate by the difference value of maximum

*Corresponding author: Benzheng WEI.

AP-PROC-CS-10CN006

and minimum structural elements in definition domain, used to achieve the edge detection of the image or signals. Let $f(x,y)$ be original image and $g(x,y)$ be structural element, then the formula of morphology gradient operator is:

$$Grad(f) = (f \oplus g) - (f \Theta g) \quad (1)$$

Where $\oplus$ denotes the expansion operator, Θ denotes the corrosion operator. From (1), we can see that the operator make the input image gray-scale step to become even more dramatic. According to the principle of generalized morphological filter, as far as possible to improve the integrity and richness of edge feature detail and obtain the smoothness edge curves, the improved morphology gradient operator is designed based on the expansion, erosion, opening and closing four operators for image anti-noise characteristics. The improved morphology gradient operator is described as

$$IGrad(f) = (f \circ g) \oplus g - (f \bullet g) \Theta g \quad (2)$$

We call $IGrad(f)$ as the improved morphology gradient operator. In order to extract the direction information of the edge pixels and easy to construct vector group at the same time, we select two cross and cross-shaped 3×3 structural elements, and make them symmetrical and complement each other. Because the random noise may be smaller than the width of the structural elements of the width, which are mixed with the processed images, therefore, it's very difficult to obtain the best filtering effect by using the separate structural element, so they can be used as the linear combination form due to complementary characteristic of the two structural elements. The optimum weighted coefficients can be determined by the least mean square adaptive method, and the improved morphology gradient filter operator can be written as:

$$FGrad(f) = c_1 IGrad1(f) + c_2 IGrad2(f) \quad (3)$$

Where c_1, c_2 are the weighted coefficients, the $IGrad1$ and $IGrad2$ are the improved morphology gradients with the two structural elements.

The main purpose that the improved morphological gradient operator is introduced and applied to quickly identify the probable location of the edge and edge points, while the image is rapidly de-noised in the image processing. After the medical image is processed by the proposed new operator, it can be achieved an available gradient image with the direction information. Then the data set of the edge point vector group $\{E_1, E_2, \cdots, E_n\}$ can be stored as the data list, where $E_i =< e_i, \theta_i >$, e_i is the edge point i in the two-dimensional plane coordinates, and θ_i is the edge gradient direction. In addition, the improved morphological gradient filter can overcome the shortcoming of tend to produce the image edge blur which is often caused by the linear filtering method in image smoothing, and it's easy-to-parallel computing to better meet the real-time image processing requirements. In the experiment, the selected structural elements are[0 10;1 1 1;0 1 0] and [1 0 1;0 -1 0;1 0 1].

III. Principal Axes Algorithm Based Coarse Registration

A Principal Axes Algorithm

Suppose that image F is of $M \times N$ pixels with its upper left pixel being $(1,1)$, and $f(x,y)$ is the gray value at point (x,y). Since some moments of an image region are invariant to geometric transformation such as translation, rotation, and scale change, they are widely applied to object classification and identification. If $f(x,y)$ is a 2-D discrete function, then the moment of order $(p+q)$ is defined as

$$M_{p,q} = \sum_{x=1}^{M} \sum_{1}^{N} x^p y^q f(x,y) \quad (4)$$

Then the central moment is expressed as

$$M'_{p,q} = \sum_{x=1}^{M} \sum_{1}^{N} (x - \overline{x})^p (y - \overline{y})^q f(x,y) \quad (5)$$

Where $\overline{x} = M_{1,0} / M_{0,0}, \overline{y} = M_{0,1} / M_{0,0}$ the $(\overline{x}, \overline{y})$ is called centroid coordinates of the object. The moment of the object having rotated by θ, or computed with reference to axes X' and Y', is invariant to rotation transformation. The rotation angle θ is as follows:

$$\tan 2\theta = 2M'_{1,1} / (M'_{2,0} - M'_{0,2}) \quad (6)$$

Application of the above formula to calculate the difference of two images about the centroid coordinates and principal axis angle separately, and then initial registration parameters ($\Delta x, \Delta y, \Delta\theta$)can be obtained and used as the initial registration parameters for coarse registration.

B Model estimation

As in [9], we assume that the type of transformation is rigid and not deformable; the model that describes the geometric transformation has the following expression:

$$P_d = RP_s + T \quad (7)$$

Where, $P_s(x,y,z)$ is a source point; $P_d(x',y',z')$ is transformed corresponding point;R=[R_{11} R_{12} R_{13};R_{21} R_{22} R_{23};R_{31} R_{32} R_{33}] is a rotation matrix; $T = [T_x, T_y, T_z]^T$ is a translation vector.

Assumed that the α, β, γ are the angles of the transformed image rotated by x,y,z axes separately, then R can be rewritten as: $R = R_\gamma R_\beta R_\alpha$.

C Coarse registration method

The principal axes algorithm can be described as follows:

Firstly, calculate the centroid, and eigenvectors of the source and target images via an eigenvalue decomposition of the covariance matrices.

Secondly, align the centers of mass via a translation.

Finally, for each image determine the angle $\Delta\theta$, the maximal eigenvector forms with the horizontal axis, and

rotate the test image about its center by the difference in angles. The images are now aligned.

IV. THE HYBRID MUTUAL INFORMATION ALGORITHM

An important characteristic of an image is the high degree of correlation among the neighboring pixels. In other words, these neighboring pixels possess similar feature values, and the probability that they belong to the same cluster is great. This spatial relationship is important in clustering.

Recently, a new clustering algorithm called K-harmonic means(KHM) has been introduced by Zhang et al.[10,11] in 2000, which is arose from an optimization criterion based on the harmonic mean. This algorithm shows promised in finding good clustering solutions quickly, and outperforms k-means and Gaussian expectation-maximization in many tests. The KHM algorithm also has a novel feature that gives more influence to data points that are not well-modeled by the clustering solution.

In this paper, the KHM algorithm is used to extract the feature point sets of the medical images based the edge point vector group $\{E_1, E_2, \cdots, E_n\}$, and we select the number of clustering centre K=150 that means the feature point sets has 150 points. The more detail bout KHM algorithm, reference to[10,11].

Mutual Information is an error metric (or similarity metric) used in image registration based on ideas from information theory [1]. The strategy is this: minimize the information content of the difference image. The mutual information between two unlabeled point-sets is a function of the chosen spatial mapping (for example, rigid, similarity, affine). The energy function of the feature points matching will be presented based on mutual information following.

Denote the feature point-sets of pre-registration medical images by $X = \{X_i, i = 1,2,...,N_1\}$ and $Y = \{Y_i, i = 1,2,...,N_2\}$ respectively. The point sets are assumed to be in $\Re^2$ or $\Re^3$. Then, a suitable choice for the distance measure between X and Y is:

$$D(T) = \sum_{i=1}^{N} \| X - TY \|^q . \quad (8)$$

In (8), following the KHM clustering algorithm, we choose $q = 3.5$, T is the rigid spatial mapping linking the two point-sets in this paper. With the spatial mapping T, the mutual information between the point-sets is a function of the joint probability P:

$$MI(P) = M(X,Y) = \sum_{i=1}^{N_1}\sum_{j=1}^{N_2} P_{ij} \log \frac{P_{i,j}}{\sum_{k=1}^{N_1} P_{kj} \sum_{l=1}^{N_2} P_{il}}, \quad (9)$$

where the joint probability

$P_{ij} = P\{I = i, i \in (1,2,...,N_1), J = j, j \in (1,2,...,N_2)\}$

the above joint probability characterizes the likelihood of obtaining a pair of features (i, j), i form X and j form Y. Intuitively, the likelihood should be large if i and j are homologies and small if they are not. The solution for the joint probability is :

$$P_{ij}(T) = \exp(-\alpha D_{ij}(T) - \lambda) . \quad (10)$$

In (10), α and λ are two Lagrange parameters, the constraints on the expected value of the point matching distance measure and the probability sum respectively. The more detailed discussion about α and λ, reference to see [14].

$$E_{MI}(P,T,\lambda,\alpha) = \alpha(\sum_{ij} P_{ij} D_{ij}(T) - d) + \lambda(\sum_{ij} P_{ij} - 1) + \sum_{ij} P_{ij} \log P_{ij} - \kappa MI(P) \quad (11)$$

In (11) The noise level is,

$$d = \sum_{ij} D(T) \frac{\exp(-\alpha D(T))}{\sum_{ij} \exp(-\alpha D(T))} .$$

$\kappa > 0$ is a new parameter which acts as a weight on the mutual information vis-a-vis the entropy and the distance measure. If $\kappa = 1$, the separate entropy term and the joint entropy in the mutual information perfectly match one another. In this paper, we choose $\kappa \in [0.3,1]$ in the optimization processing. Our approaches to minimizing the energy function in (12):

$$\partial^* = \arg \min E_{MI}(P,T,\lambda,\alpha) . \quad (12)$$

The best registration would be achieved if the objective function (metric function) reaches the global minimum. However, the function generally contains a set of local minimums. To solve multiple local minimum problems, a Powell algorithm is developed to solve the registration problem in this paper.

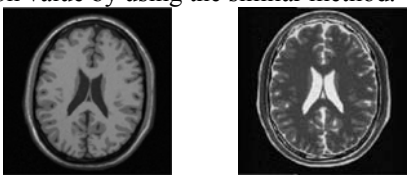
V. THE SIMULATION AND EXPERIMENT

We implement registration of brain image to confirm performance of proposed algorithm in last section. The algorithms are implemented by using Matlab 7.1 on a Pentium IV, 2.8 GHz, 1GB RAM computer, and image file is from database (www.bic.mni.mcgill.ca/brainweb). In the following experiment, the parameters in translate procedure is initialized as search space in x direction is x∈[-20, 20] and y direction is y∈[-20, 20], the search space in rotation is θ∈[-15° ,15°].

A Experiment I.

In Fig. 1, the image (a) and image (b) are illustrated. The size of each image is 512×512 pixels under DICOM format and the whole images have good quality and don't require a pre-treatment. In our experiment, image Fig.1(b) as source image and image Fig. 1(a) serve as target image. The two images are alignments by using MI and the presented hybrid MI (PHMI) method in this paper separately. And our experiment results are shown in Tab.1 and Tab. 2. The Tab. 1 shows the comparison of experiment result for image registration transform parameters (Δx, Δy, Δ θ), and the Tab. 2 shows the

results of the spatial transform parameters and mutual information value by using the similar method.



(a) Target image (b) Source image

Figure 1. Brain images for registration

TABLE 1. COMPARISON OF REGISTRATION PARAMETERS

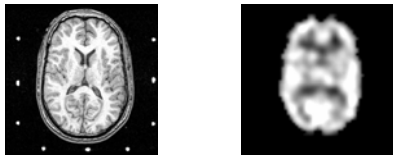
		Δx (pixel)	Δy (pixel)	Δθ (°)
Standard Value of Transformed		12	10	-10
Average Searching Value	MI	12.86	9.42	-10.38
	PHMI	12.16	10.02	-10.17

TABLE 2. THE SPATIAL TRANSFORM PARAMETERS AND MUTUAL INFORMATION

Algorithm	**MI**	**PHMI**
Solving times	100	100
Average Value	167.64	130.09
Standard deviation	7.567	3.12
Mutual information	2.157	2.541
Average times (s)	386.2	315.6

B Experiment II.

In this experiment, the two brain images (Shown in Fig. 2.) of the same patient are registered to test the performance of the presented new method in separately 100 times. The size of each image is 256×256 pixels, and the MR images is set as source image. The registration results are shown in table 3. After this simulation, we can see that the PHMI algorithm in accuracy and speed with the more obvious advantages than the other two algorithms.



(a) MR image (b) SPECT image

Figure 2. the two registration images

TABLE.3 THE COMPARISON REGISTRATION RESULTS

Algorithm	MR-SPECT	
	Time(s)	Success rate
MI	2168	72%
PHMI	1275	98%

VI. CONCLUSIONS

In the paper, we have presented an innovative PHMI method for the automatic registration of medical image, and it has achieved a good performance. This registration is actually a problem of searching optimum parameters in multi-dimension space. The existence of local maxima in object function brings much trouble in optimizing process. In this paper, we use Powell to extend the searching ability, the method can guarantees registration convergence to the global optimum. Though combining spatial information and mutual information as criterion for multimodality medical image registration still remains problems unsolved, the experiments results prove that the proposed method has robustness and effectiveness.

ACKNOWLEDGMENT

The work was supported by the Natural Science Foundation of Shandong Province (ZR2009CM085) and A Project of Shandong Province Higher Educational Science and Technology Program (NO.J09LF23).

REFERENCES

[1] Josien P W Pluim,Maintz J B Antoine ,Max AViergever. Image registration by maximization of combined mutual information and gradient information. IEEE Transactions on Medical Imaging , Vol. 19, No. 8, 2002, pp. 809-814.

[2] W. M. Wells III, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis, "Multi-modal volume registration by maximization of mutual information," Med. Image Anal., vol. 1, no. 1, pp. 35–51, 1996.

[3] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," IEEE Trans. Med. Imag., vol. 16, no. 2, pp. 187–198, 1997.

[4] C. Studholme, D. L. G. Hill, and D. J. Hawkes, "Automated three-dimensional registration of magnetic resonance and positron emission tomography brain images by multiresolution optimization of voxel similarity measures," Med. Phys., vol. 24, no. 1, pp. 25–35, 1997.

[5] G. P. Penney, J. Weese, J. A. Little, P. Desmedt, D. L. G. Hill, and D. J. Hawkes, "A comparison of similarity measures for use in 2D-3D medical image registration," IEEE Trans. Med. Imag., vol. 17, no. 4, pp.586–595, 1999.

[6] YAO YuCui ,YANG LiCai , LI jinLiang. Medical image registration by combined morphology gradient and mutual information[J].Journal of Biomedical Engineering Research, Vol. 25, No. 2, 2006, pp. 75-78.

[7] TANG Min. Image registration based on improved mutual information with hybrid optimizer[J] . Chinese Journal of Biomedical Engineering , Vol. 17, No. 1, 2008, pp. 18-25.

[8] WEI Benzheng, ZHAO Zhimin, SONG Yizhong. Anovel algorithmfor multimodality medical image registration based on IPSOalgorithm and hybrid information[J], Journal of Optoelectronics • Laser, Vol. 20, No. 9, 2009, pp. 1271-1274.

[9] C. R. Maurer, R. J. Maciunas, and J. M. Fitzpatrick 'Registration of head CT images to physical space using a weighted combination of points and surfaces', IEEE Transactions on Medical Imaging, 17(5): pp 753–761, 1998.

[10] B.zhang.Generalized K-harmonic means-boosting in unsupervised learning. Technicl Report Hpl-2000-137, Hewlett-Packard Labs,2000.

[11] G.Hamerly, C. Elkan, "Alternatives to the k-means algorithm that find better clusterings", Proc. of the ACM Conference on Information and Knowledge Management, CIKM-2002, 2002, 600-607.

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 178-181

Study of Awareness-based Knowledge Communication

Li Wang
College of Computer and Software, Taiyuan University of Technology,Taiyuan, Shanxi, P. R. China, 030024
E-mail: l_lwang@126.com

***Abstract*—With the development of computer network, the smart of communication agents was significantly increased, more intelligent process can be done by computers. To realize this, agents need more detailed concrete message to be transferred. But what should be transferred? How to transfer necessary message to improve the receiver's understanding capability and not add too much communication load? In this paper, We put forward a knowledge communication definition and knowledge communication protocol to solve this problems. We build knowledge communication protocol stack and KC^2A^2P protocol ontology for content and context awareness. Based on such model, we utilize SRC algorithm to obtain the semantic relativity value and realize mutual awareness and understanding in communication. At last, some experiments are given to show our knowledge communication would improve the meaning delivery and optimize the communication quality.**

***Index Terms*—content awareness, context awareness, knowledge communication, ontology, semantic relativity**

I. INTRODUCTION

With the development and widely application of computer network technology, more and more new requirements for computer communication are put forward. Modern Internet technology has decreased the bit data transportation error and guaranteed the accuracy of data transmission to some extend. But communication goal is to know the opposite's intension and deal with things collaboratively, only getting right data is the most important ,but just the basic process for communication. When interaction happens between entities that have different profiles and contexts, some misunderstanding or information losing maybe happened. For example, in electronic commercial, providers maybe confuse buyer's true requirements because they always do not know each other and they have different backgrounds; in ubiquitous environment the dynamic situation makes that the same transported information always has different meanings. So, some intelligent protocols are required to improve content exchange, not just data transportation on computer communication. That is, agents need more detailed concrete message to be transferred. But what should be transferred? How to transfer necessary message to improve the receiver's understanding capability and not add too much communication load?

In this paper we will answer these questions. Our contributions are:

1) We proposed a kind of of knowledge communication definition and put forward related definitions. In knowledge communication model ,we list the content that should help receiver to understand sender's truth meaning and intension.

2) We build knowledge communication protocol stack and KC^2A^2P protocol ontology for content and context awareness. We utilize ontology to describe the detailed concepts,attributions,relationships in content and context awareness protocol.

3) We utilize semantic relativity algorithm SRC to realize "awareness" ability in knowledge communication.

The outline is: we will analysis related work in the second part and point out the main problems that knowledge communication is facing. The definition of knowledge communication is put forward in the third part. We build KC^2A^2P, the common shared ontology framework for knowledge communication protocol in the fourth part. To realize mutual awareness and understanding, a semantic relativity algorithm is chosen in the fifth part. Some experiments are done to verify the effect of KC^2A^2P and at last some conclusions are given.

II. RELATED WORK

Related researches are mainly divided into two categories:

A. Research on communication protocol content

Many works are proposed to establish communication ontology, add content domain ontology reference in protocol and conducting message content interpretation by endpoint. DARPA's AGENT communication language KQML[1] divides application layer protocol into three layers: communication layer, message layer and content layer. Reference information about communication content ontology is added in message layer so as to achieve communication content understanding. SHADE project [2] exploited KQML as protocol, put forward a scalable known-knowledge-based cooperation framework to achieve information share between multiple participants. Some scholars make use of RDF [3], OWL (Ontology Web Language) to improve the syntax of agent communication protocols. Liu Dayou etc [4] proposed methods for computing ontology deviation and dealing with logical exception information transformation to resolve the deficiency and abnormal contents in communication process. Juan Ignacio Vazquez [5] designed RDF Triples and variables supported PLANT two resources description methods, and use SPARQL and simple pattern to provide endpoint semantic process.

AP-PROC-CS-10CN006

B. Research on protocol actions

Communication action logic and knowledge-based action express research are always based on language and behavior theory in general. Researchers used modal logic, description logic, empirical semantics [6] and so on to set up communication behavior ontology for resolving the action semantic deviation problem.

Darina [7] proposed a communication schema based on the InFFrA society reasoning framework that used optimized decision-making theory methods to realize understanding communication meaning. Jesu' s Bermu' dez [8] et. al. based on event calculus to construct 3-tier communication acts ontology: super layer, standard layer and application later. Wang Li [9], Wang Xiuling [10], Michal Laclavik[11]et. al conducted related works targeting realization of knowledge-driven communication protocol methods.

As can be seen, current research mainly takes agent-based system for application background. This is partly due to that agent is important in intelligent information system application, another reason is as what WWW founder Tim berners-lee[12] mentioned, agent technology will be one of the main instruments in Semantic Web. Therefore, the knowledge-driven communication protocol in this article also takes agent-based information system as application background.

In the current research, researchers mainly focus on building communication ontology to achieve the translation between different terms and concepts and communication between heterogeneous systems. But as the accurate data transmission does not guarantee the correct understanding and process, different protocols vocabulary mapping and transformation can only realize semantic transmission, it cannot guarantee peers to exactly understand the communication information meaning and make right decision based on protocol content. To realize transfer meaning, we not only need a public sharing Knowledge base, but also the necessary auxiliary information to be added. In this paper, we put forward content and context awareness-based knowledge communication and its protocol mechanism.

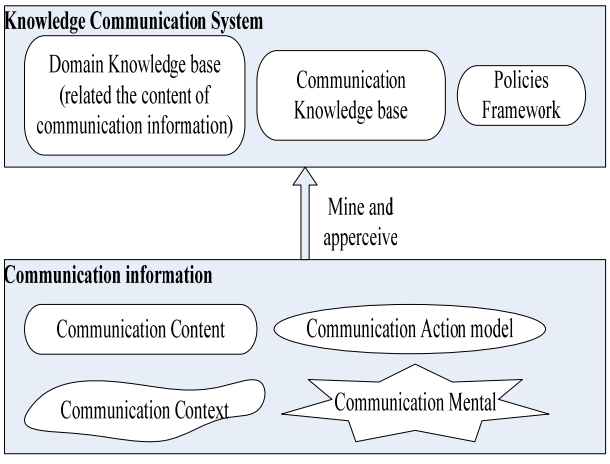


Fig 1 knowledge communication model

III. KNOWLEDGE COMMUNICATION

Definition1 (Knowledge Communication): KC is a 4-tuple:KC=(CI,KD,KC,PF), CI is Communication Information, DK is Domain Knowledge that is message domain ontology, CK is Communication Knowledge that describes the control policy, rule, information and knowledge, CP is Cognize Protocol that shows the framework of communication policy. (Fig1)

Definition 2（CI ,knowledge Communication Information）CI is 4-tuple: CI=（CC,CX,CA,CM）.CC is communication content, CX is communication context, CA is communication action model, CM is communication mental model.

Definition3 (The mechanism of knowledge communication). The mechanism of knowledge communication includes 6 steps (as fig.2) :①detecting communication information② apperceiving ,analyzing message and coding message ③delivering message④ receiving message and decoding ⑤understanding transferred message and discovering implied information ⑥processing information and making decision.

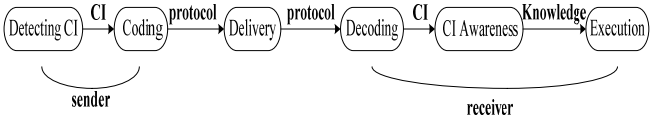


Fig2. The mechanism of knowledge communication

IV. KNOWLEDGE COMMUNICATION PROTOCOL

Knowledge communication protocol, KC^2A^2P (Knowledge-driven Content and Context-Awareness Adaptive Protocol) is a hierarchical structure and it adds context and act layers, extend the content of "content layer" as Fig3. The act layer has two parts, *performative* part defines the basic atomic communication actions, *interaction* part defines the communication state sequence classes that are composed by atomic communication actions based on some patterns. KC^2A^2P takes illocutionary of "speech act theory" as logic base, which is an act performed in saying something. So from the protocol syntax view, content,context and act of KC^2A^2P are the "content" part of protocol ,same as the "body" part of some protocols, which is un-control information; context part is used to assistant understanding content and some action implication will be obtained. At the same time, the "act" definition and logic will offer some semantic verification for content understanding. So in fact the different three parts are related closely.

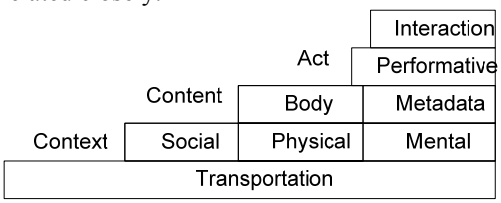


Fig3 The protocol stack structure of KC2A2P

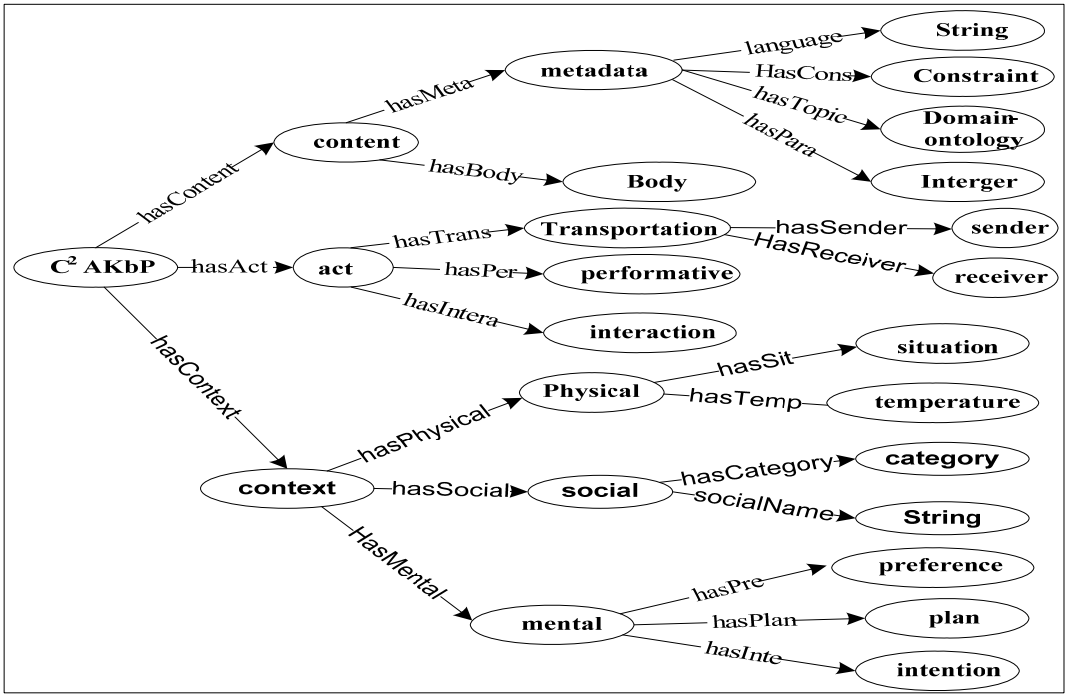


Fig 4. Content and context awareness protocol ontology

To understand each other correctly, accurate information semantic transportation is the first part. The core ontology of KC^2A^2P defines three concepts of content, context, act, which not only support the semantic interconnection between different protocols, but also offer framework for the protocol information formation. (Fig4)

V. ONTOLOGY-BASED AWARENESS COMPUTING

Based on KC^2A^2P ontology, senders can decide what it need to encode and receiver can know what the truth meaning of transferred message is and what is related the requirement.

Ontology-based semantic relatedness computing is important for realizing this mutual awareness. Semantic relatedness is a general concept that could include semantic similarity computing. And it is more useful than just semantic similarity computing. Because similar entities are always semantically related by virtue of their similarity (female-woman), but dissimilar entities may also be semantically related by lexical relationships such as meronymy (car–wheel) and antonymy (hot–cold), or just by any kind of functional relationship or frequent association (pencil–paper, penguin–Antarctica, rain–flood). Computational applications typically require relatedness rather than just similarity; for example, when asking for traveling information, some resource about hotel and weather is also required. SRC (Algorithm of Semantic Relatedness of Concepts) [13] is a semantic relativity algorithm ,which just bases on ontology and does not need any former knowledge. It has three main steps : reading OWL files and building proper graph; according to its subsume axioms producing sub graph(that is a tree structure) consisting only subsumption relationship and counting the semantic relatedness by myObjectMatching method; according to the relation path functions ,calculating the whole classes semantic similarity based on the graph filling with six kinds of relations. Some formulas are:

$$Sem\_Related\ [C_1][C_2]\ \ is\ \ the\ \ semantic\ \ relativity\ \ value\ of\ C_1\ \ and\ \ C_2$$

$$Sem\_Weight\ [C_2][C_3]\ \ is\ \ the\ \ direct\ \ semantic\ \ weight\ \ of\ C_2\ \ and\ \ C_3$$

$$Sem\_Road\ [C_1][C_3] = Sem\_Related\ [C_1][C_2] \times Sem\_Weight\ [C_2][C_3] \quad (1)$$

$$Sem\_Related\ [C_1][C_3] = \max_s \left(Sem\_Road\ [S][C_1][C_3] \right) \quad (2)$$

VI. EXPERIMENT

We take Education Information Platform as experiment background, Eclipse+JADE (Java Agent Development framework) as exploitation tool, research on the awareness-based knowledge communication process. One client inquires a book named Semantic Web. There are four different providers. Provider1 holds one book on Ontology, provider2 holds Semantic Web Journal and provider3 hasn't any resource about Semantic Web. When the buyer sends a request for looking for a book on Semantic Web, provider3 refuses the request because it doesnot have this resource. Provider1 and provider2 analyze the request by content awareness algorithm and send decision to the client. Client analyses these replies and chooses propose which is close to its own request. The interaction process of this example is illustrated by Figure 5.

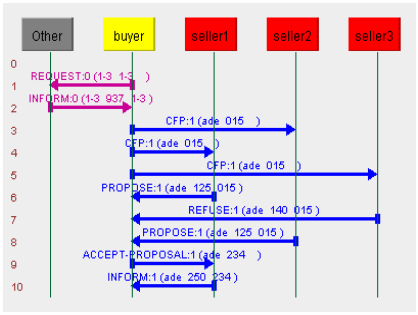


Fig5 one successful interaction process

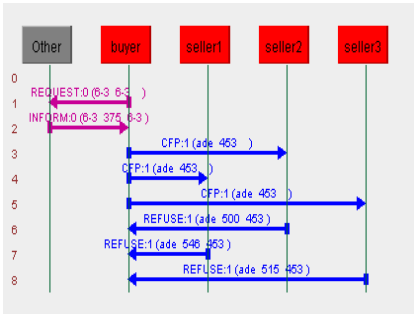


Fig 6 one failing interaction process

The communication without extra content and context information failed (Fig6). By the comparison of interaction process in Figure 5 and Figure 6, we can find out that adding content and context-awareness mechanism to knowledge communication can greatly improve the quality of communication.

VII. CONCLUSIONS

This paper proposed awareness-based knowledge communication, build knowledge communication protocol stack and KC^2A^2P protocol ontology for content and context awareness. Based on such model,we utilze SRC algorithm to obtain the semantic relativity value and realize mutual awareness and understanding in communication. At last, some experiments are given to show awareness-based knowledge communication would improve the meaning delivery and optimize the communication quality.

This article is part of our ongoing project. In the future, we will improve and fulfill the adaptive coding rule and carry out more research on intelligent software mechanism to enable the protocol architecture widely applied to web applications, agent-based systems, ubiquitous computing environment, and mobile communications and so on.

ACKNOWLEDGMENT

This work is funded by Shanxi Province under Grand No. 2006021015, Project of science and technology of ShanXi education department under Grant No.2007111.

REFERENCES

[1] T. Finin, Y. Labrou, J. Mayfield, KQML as an agent communication language, in: J. Bradshaw (Ed.), Software Agents, MIT Press, Cambridge, MA, 1997.

[2] Daniel R. Kuokka ,Jim McGuire . SHADE: Knowledge-Based Technology for the Re-Engineering Problem. http://www-ksl.stanford.edu/knowledge-sharing/papers/shade-overview.html \#ref_11.2009

[3] Kit-ying Hui, Alun Preece. An Infrastructure for Knowledge Communication in AKT. http://www.csd.abdn.ac.uk/~apreece/akt/aktbus/aktbus-1.0/index.html 2003

[4] Bai Yan,Liu Dayou. Logical exception handling method in agent communication.[J] Computer Research and Development. Vol. 44, no. 3, pp. 427-433. Mar. 2007

[5] Juan Ignacio Vazquez and Diego López de Ipi?a. mRDP: An HTTP-based Lightweight Semantic Discovery Protocol. The International Journal of Computer and Telecommunications Networking. 2007.

[6] Felix Fischer and Michael Rovatsos. Reasoning about communication: A practical approach based on empirical semantics. In Proceedings of the 8th International Workshop on Cooperative Information Agents (CIA-04), volume 3191 of Lecture Notes in Artificial Intelligence. Springer-Verlag, 2004.

[7] Darina Dicheva, Lora Aroyo. A Minimalist Approach to Support Ontology-driven Educational Information Systems Communication. International Workshop on Applications of Semantic Web Technologies for E-Learning (SW-EL). 2004

[8] Jesu′s Bermu′dez etc. Interoperation among agent-based information systems through a communication acts otology. Information systems 32 (2007)1121-1144

[9] Li wang,Liqing liu.Reserch on Content-awareness Protocol of Knowledge Communication Based on Ontology.Knowledge Discovery and Data Mining, WKDD2009.

[10] Wang xiuling,Wang li,Feng xiufang,Yu xueli. Research on semantic web service protocol Soap based on ontology Chinese Artificial intelligence annual academic conference 2005 : 1 482-487.

[11] Michal Laclavík. Ontology and Agent based Approach for Knowledge Management. Doctor Thesis. Institute of Informatics, Slovak Academy of Sciences.June 2005

[12] T. Berners-Lee, J. Hendler, O. Lassila, The semantic web ,Sci. Am. 284 (5) (2001) 34-43.

[13] Wang, L. and X.L. Yu. A method measuring semantic relatedness for concepts in ontology base. in 4th International Conference on Information Technology and Applications, ICITA 2007, January 15, 2007

-

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 182-185

Research on Tag-based Collaborative Filtering Strategy

Shuo Wan[1], and Huizhong Qiu[2]
[1] School of Software, University of
Electronic Science and Technology of China, Chengdu, P.R.China
Email: wanshuo168@gmail.com
[2] School of Computer Science and Engineering, University of
Electronic Science and Technology of China, Chengdu, P.R.China
Email: hzqiu@uestc.edu.cn

***Abstract*—Recommendation technology is designed to take the initiative to recommend using the user's history behavior information, without requiring users to explicitly specify the query case information. Collaborative filtering is the most widely recommended technique. However, some problems of the traditional collaborative filtering recommendation system still exist, and these problems significantly affect the recommended results. Tag system as the essential functions of Web2.0 websites in recent years has been very widely used. This article will combine the tag information with the collaborative filtering recommendation, and recommend resources by recommending tags. By analyzing a problem of the traditional collaborative filtering strategy, this experiment proves tag-based recommendation strategy can effectively solve these problems and improve the accuracy of recommendation.**

***Index Terms*—tag, collaborative filtering, recommendation, similarity, preference**

I. TRADITIONAL USER-BASED SIMILARITY MODEL ALGORITHM

Collaborative filtering assumes that a user in the system and his similar groups both have the similar preferences on the system resources, and the users have the similar preference on the similar resources. Thus mining the collective wisdom embedded in massive data, and predicting the individuals with the similar groups.

The traditional user-based similarity collaborative filtering algorithm is mainly in two steps [6]:

1. To calculate the user's similarity with other users, and receive the top N preference similar users.
2. Based on the similar user's preferences to predict the user's preferences, and recommend resources.

Formal description is as follows:

$U = \{u_1, u_2, ..., u_m\}$ stands for m users collection, $I = \{i_1, i_2, ..., i_n\}$ stands for n resources collection, the preferences of the user on the resources can make a matrix R (shown as follows), $r_{u,i}$ in matrix R stands for the preferences of user u on resource i, $r_{u,i} = 0$ stands for the user u have not evaluated preferences on resource i.

$$R = \begin{pmatrix} r_{11} & \cdots & r_{1n} \\ \vdots & \ddots & \vdots \\ r_{m1} & \cdots & r_{mn} \end{pmatrix}$$

1. The user's similarity calculation is based on the row vector of the matrix R. Calculate cosine similarity of the vector.

$$s_{u,v} = \frac{\sum_{i \in I_{u,v}} r_{u,i} \cdot r_{v,i}}{\sqrt{\sum_{i \in I_{u,v}} r_{u,i}^2 \sum_{i \in I_{u,v}} r_{v,i}^2}} \quad (1)$$

$I_{u,v}$ is the intersection of user u and v's preference resources.

2. Select the top N preference similar users. Predict the user's preference.

$$r_{u,i} = \overline{r}_u + \partial \sum_{v \in S_u \cap U_i} s_{u,v} \cdot (r_{v,i} - \overline{r}_v) \quad (2)$$

S_u is the similarity of the top-N users. U_i is the users' evaluate preference on resource i, $\partial = 1 / \sum_v s_{u,v} \quad v \in S_u \cap U_i$.

II. THE PROBLEM EXISTED IN USER-BASED SIMILARITY MODEL

The traditional user-based similarity model is based on two underlying assumptions:

1. There is only one user's preferences similar model, and using the only value generated by this similarity model will be able to determine the similarity of the user's preferences.
2. The preference similarity on one type of the resources produced by this model can also be applied to other types.

But there are some problems in these two assumptions. The article will illustrate the unreasonableness with the following examples.

A. Problem Description

In a small book recommendation system, the user collection is $\{u_0, u_1, u_2, ..., u_5\}$, the resource collection is $\{i_0, i_1, i_2, ..., i_5\}$, i_0, i_1, i_2 are classified literature books, i_3, i_4, i_5 are classified mathematics books. The system use the evaluate-preference way to score the books, the rating interval is [1, 5]. The system needs to predict u_0's score to book i_5 based on the other users' score. And the system decides whether to recommend i_5 to u_0.

The score matrix table is shown as follows:

TABLE I.
SCORE MATRIX TABLE

Res. / User	i_0	i_1	i_2	i_3	i_4	i_5
u_0	4	4	5	5	4	?
u_1	5	3	4	1	2	1
u_2	3	4	5	2	2	1
u_3	4	3	3	3	1	1
u_4	1	2	1	4	5	5
u_5	1	1	2	5	4	5

Using the user-based similarity model collaborative filtering, each row of the score matrix is every user's preference vector, for example, $preference(u_0) = \{4,4,5,5,4\}$.

Using Eq. (1), calculate the preference vectors similarity between user u_0 and user $u_1 ... u_5$, and the results are sorted from high to low.

$s(u_0, u_2) = 0.94;$

$s(u_0, u_3) = 0.94;$

$s(u_0, u_1) = 0.89;$

$s(u_0, u_5) = 0.87;$

$s(u_0, u_4) = 0.84;$

Suppose N=3, and select the top 3 preference similar users u_1, u_2, u_3 to predict u_0's score to book i_5. Using Eq. (2):

$r(u_0, i_5) = 2.73$

Therefore u_0 doesn't like book i_5.

This conclusion is clearly at odds with reality.

Seeing from u_0's score to book i_3, i_4, u_0 is also very favorite to mathematics books. Therefore, u_0 is likely to also show a preference for book i_5.

B. Problem Analysis

The reason is that this method ignores the individuals' preference differences in various aspects of interest, and the obtained similarity is vague and not accurate. both like literature books and mathematics books, but under the user-based similarity model, the users that the system recommended are only interested in literature books. Therefore, using their preference degree of mathematics books to predict the preference degree of is not accurate. In other words, using traditional user-based similarity model to obtain the similar users is incomplete and does not cover the user's preferences.

III. USER-BASED SIMILARITY MODEL

Tags are a special kind of meta-data (metadata), and they are one of the necessary functions of Web2.0 sites [5]. They are from the tag-maker's subjective experience on resources, while they are also used to describe resources and classify resources by user. Tag information directly stands for the user's interest and preference, and it describes the relationship between users and resources and has a high potential value. The use of tag information can provides a useful supplement for collaborative filtering system, and it can improve the recommendation accuracy.

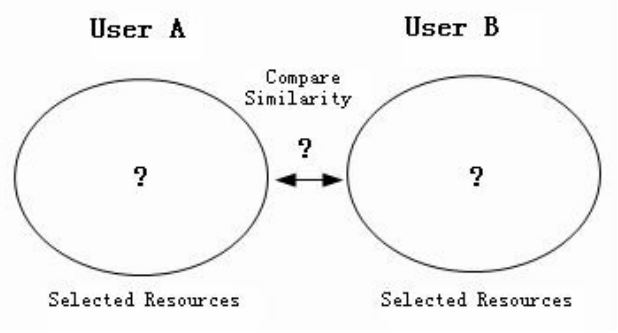


Figure 1. Traditional user-based similarity model

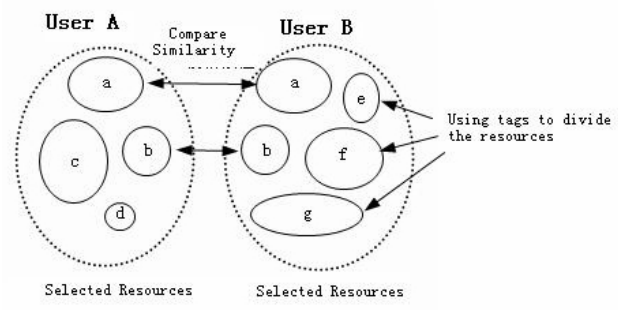


Figure 2. User-based similarity model

Tags classify the resources, and reflect the user's preferences. Through the user's preference tag to recommend resources, the users can get a more accurate recommendation. The methods are as follows:

The tags collections for user u and user v are T_u and T_v. The intersection $T_u \cap T_v$ is the two users' common

tags. Union $T_u \cup T_v$ is the two users' all tags. $s_{u,v}$ is the preference similarity of u and v.

$$s_{u,v} = \frac{|T_u \cap T_v|}{|T_u \cup T_v|} \quad (3)$$

Compare the preference similarity between the users. When $s_{u,v}$ =1, the two users use the same tags. Although the similarity is high, there is no value to recommend between the two users. So remove the users of this similarity firstly.

Then, get the top N users of preference similarity for target user, and the top N users are the tag-recommenders for target user. $score(u,t)$ is the possibility of the tag t recommended to target user. Sort the $score(u,t)$ from highest to lowest, and select the top K tags recommended to target user. The collection of top K tags is marked as $recommendTag(u)$.

$$score(u,t) = \frac{1}{|S_u \cap U_t|} \sum_{v \in S_u \cap U_t} s_{u,v} \quad (4)$$

S_u is the collection of top N users, and U_t is the collection of the users who used the tag t.

Then, calculate the relevance of resources and tag $t \in recommendTag(u)$. Through the frequency that all users marked the resource using this tag, the relevance between the resource and different tags can be measured.

$$relate(i,t) = \frac{countTagging(t,i)}{\sum_{k \in T_i} countTagging(k,i)} \quad (5)$$

T_i is the tag collection of resource i, and $countTagging(t,i)$ is the frequency that all users marked the resource i using tag t. The more $countTagging(t,i)$ is, the more relevant between resource i and tag t.

Remove the resources with a low relevance, and obtain the resource collection $I_{u,t}$ which is used to predict. In $I_{u,t}$, use Eq. (1) and Eq. (2) to recommend resources under each tag $t \in recommendTag(u)$.

IV. EXPERIMENT AND RESULT ANALYSIS

A. The experiment based on Movielens data set

The experiment is based on Movielens 10M100K data set, and select the one-tenth of the data as the experimental data sets. Randomly select 20% user data as test data, and the remaining 80% as training data. Repeat the experiment five times.

B. Evaluation Standard-MAE

Mean Absolute Error (MAE) is the absolute of average difference between predicting preferences and actual preferences [6]. It reflects the accuracy degree of the recommendation. If a recommended method obtained a lower MAE value, the average prediction error is lower. The method of preference prediction is more accurate and has better performance.

$$MAE = \frac{\sum_{i=1}^{N} |pRate_i - rRate_i|}{N} \quad (6)$$

$rRate_i$ is the user's actual preference of selected resources in testing data, and $pRate_i$ is the predicted preference.

C. Experiment results and analysis

The MAE results of two kinds of algorithm in Movielens data set in Figure 3:

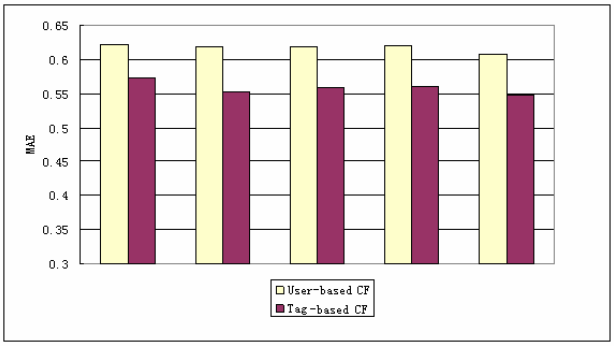


Figure 3. MAE result of two kinds of algorithm

If the MAE is a lower value, the effect of recommendation method is better. As the chart shows, the tag-based collaborative filtering method proposed in this paper is better than the traditional user-based collaborative filtering method in aspect of the recommendation accuracy.

V. SUMMARY

This paper presents an improvement of traditional collaborative filtering method, and introduces the tagging system for similarity analysis. The new method reduces the scarcity of score matrix and is effective to classify the resources for recommendation. Use Movielens data sets and MAE indicators to prove the effectiveness of the recommendation with tags. In addition, the user-tag relevance and trust degree between users can also be introduced, and the more accurate recommendation can be obtained.

REFERENCES

[1] JiaweiHan，MichelineKamber. Data Mining Concepts and Techniques, Second Edition. 2007

[2] Goldberg, D., Nichols, D., Oki, B., & Terry, D. (1992). Using collaborative filtering to weave an information tapestry. Communications of the ACM, 35(12), 61–70.

[3] P. Resnick, N. Iakovou, M. Sushak, P. Bergstrom, and J. Riedl, "GroupLens: An Open Architecture for Collaborative Filtering of Netnews," Proc. 1994 Computer Supported Cooperative Work Conf, 1994.

[4] Huang, Hung, and J. Y. jen Hsu. You are what you tag. In Proceeding of AAAI Spring Symposium on Social Information (AAAI-SIP). 2008. 36–41.
[5] Li, Guo, Zhao. Tag-based social interest discovery. In Proceeding of the 17th International Conference World Wide Web (WWW '08). 2008. 675–684.
[6] Sarwar B.M. et al., "Item-Based Collaborative Filtering Recommendation Algorithms," 10th Int'l World Wide Web Conference, ACM Press, 2001, pp. 285-295.
[7] G. Linden, B. Smith, and J. York. Amazon.com Recommendations: Item-to-Item Collaborative Filtering. IEEE Internet Computing 2003, VOL 7; NUMB 1, pages 76-80
[8] Lang. Newsweeder: Learning to filter netnews. In Proceedings of the 12th International Conference on Machine Learning. 1995. 331-339
[9] Mirza, B.J. (2001). Jumping connections: A graph-theoretic model for recommender systems. Retrieved February 1, 2003,
[10] Schmitz, Hotho, Jaschke, et al. Mining association rules in folksonomies. In Proceedings of Data Science and Classification. Springer, 2006. 261–270,
[11] Nathan N.Liu, Qiang Yang, EigenRank: A Ranking-Oriented Apporach to Collaborative. ACM SIGIR 2008, pages 83-90.
[12] Herlocker, Konstan, Riedl. Explaining Collaborative Filtering Recommendations. In Proceedings of Computer Supported Cooperative Work. ACM, 2000. 241-250
[13] Massa, P., B. Bhattacharjee. 2004. Using Trust in Recommender Systems: an Experimental Analysis. In Proceedings of iTrust2004 International Conference.
[14] Trant, Jennife. Studying Social Tagging and Folksonomy: A Review and Framework. Journal of Digital Information, 2009, 10(1).
[15] Begelman, Keller, Smadja. Automated tag clustering: Improving search and exploration in the tag space. In Proceedings of Collaborative Web Tagging Workshop at WWW2006. 2006.

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 186-189

An SVM model for Water Quality Monitoring Using Remote Sensing Image

Wei Huang[1], Fengchen Huang[2], and Jing Song [2]
[1] College of Hydrology and Water Resources, Hohai University, Nanjing, China
Email: david_morika@hhu.edu.cn
[2] College of Computer and Information, Hohai University, Nanjing, China
Email: hhice@126.com, songjing5996@yahoo.com.cn

***Abstract*—The accuracy of traditional monitoring methods using remote sensing was lower, because of the limited number of monitoring points on the Tai lake. This paper proposed to use the Least Squares Support Vector Machine (LS-SVM) theory to improve the accuracy of water quality retrieval, which is suitable for the small-sample fitting. The LS-SVM model was used to monitor concentration of suspended matter. In this paper, the Radial Basic Function (RBF) was chosen as the kernel function of the retrieval model, and the grid searching and *k*-cross validation were used to choose and optimize the parameters. From the results of experiment, it showed that the proposed method had good performance and at the same time, the complexity is lower and the speed of the modeling was rapid.**

***Index Terms*—LS-SVM, water retrieval, grid searching, remote sensing**

I. INTRODUCTION

The use of remote sensing for water quality monitoring is an uncertain problem [1]. How to establish an appropriate model for the retrieval of the water quality parameters is a difficult problem, when it referred to either different water regions or different water characteristic.

Because of using linear regression to estimate the water quality parameters could not get the accurate retrieval results, so the nonlinear algorithms, for example, BP neural network and SVM were used in the filed of the water monitoring. Unfortunately, BP algorithm is easy to fall into the local minimum; the hidden number is hard to determine and the training speed is slower. So the BP neural network method affected the accuracy of retrieved result. The drawback of the SVM is that the optimum kernel transfer function and its corresponding parameters for respective data sets are difficult to set [2]. It reduces the generalization ability, and the complexity of algorithm is increasing.

LS-SVM is a reformulation of the standard SVM [3]. LS-SVM uses the least-squares linear system instead of the quadratic programming method to estimating the function. So it has good generalization ability. According to the limited sample points of water monitoring, and the requirement of good generalization ability, the LS-SVM algorithm and theory was proposed to retrieve the water quality of Tai Lake.

II. THE MODEL OF LS-SVM

The LS-SVM uses the square of deviance cost function instead of ε-insensitive loss function, so the restriction of the inequality is transformed into the restriction of the equation. Therefore, compared with the SVM, LS-SVM algorithm greatly reduces the complexity of the algorithm, making the training speed very fast, so the retrieval precision is increased.

For the training sample sets $D=(x_k, y_k)$, $x_k \in R^n$, $y_k \in R^{n_h}$, $k=1,2,...N$ .Where, x_k is the input data, y_k is the output data. According to the theorem Mercer, the kernel function $K(\cdot,\cdot)$ and the mapping function $\varphi(\cdot)$ allow:

$$K(x_k, y_k) = \varphi(x_i)^T \varphi(x_j) \tag{1}$$

LS-SVM is solved in the original space, so the following optimization problem is obtained:

$$\begin{aligned} &\min J(\omega, b, \xi) = \frac{1}{2}\omega^T\omega + \frac{1}{2}\gamma\sum_{k=1}^{N}\xi_k^2 \\ &s.t. y_k = \omega^T\varphi(x_k) + b + \xi_k (k=1,2...N) \end{aligned} \tag{2}$$

Where γ is an adjustable regularization parameter, γ is the compromise between the training error and model complexity, so that the desired function has good generalization ability. The value of γ is greater, the regression error of the model is smaller [4].The variable ω reflects the complexity of the function, which is a linear combination of non-linear mapping function $\varphi(\cdot)$.

LS-SVM defines a different loss function compared with the standard SVM, and it transfers the restriction of the inequality into the restriction of the equation. By adding Lagrange function:

$$L(\omega, b, \xi, a) = J(\omega, b, \xi) - \sum_{k=1}^{N} a_k[\omega^T\varphi(x_k) + b + \xi_k - y_k] \tag{3}$$

Lagrange multiplies $a_k \in R$, the optimal a, b can be obtained through the KKT conditions:

$$\frac{\partial_L}{\partial_\omega} = 0, \frac{\partial_L}{\partial_b} = 0, \frac{\partial_L}{\partial_\xi} = 0, \frac{\partial_L}{\partial_a} = 0 \tag{4}$$

Manuscript received January 30, 2010; revised March 12, 2010; Copyright credit, project number, corresponding author, etc.

AP-PROC-CS-10CN006

Further by calculating, the following equations will be obtained:

$$\begin{cases} \omega = \sum_{k=1}^{N} a_k \varphi(x_k) \\ \sum_{k=1}^{N} a_k = 0 \\ a_k = \gamma \xi_k \\ \omega^T \varphi(x_k) + b + \xi_k - y_k = 0 \end{cases} \quad (5)$$

The matrix equation (6) will be given by the equation 8and equation (10), by eliminating the variables ω and ξ:

$$\begin{bmatrix} 0 & \Theta^T \\ \Theta & K + \gamma^{-1} I_n \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \quad (6)$$

Where $y = [y_1, \cdots, y_n]^T$, $\Theta = [1, \cdots, l]$, $a = [a_1, \cdots, a_N]^T$ and

$$K_{i,j} = \varphi(x_i)^T \varphi(x_j) \quad (i, j = 1, \cdots, N) \quad (7)$$

The values a, b will be obtained by solving the equation (6). So the following equation shows LS-SVM model function that is used for water quality monitoring.

$$f(x) = \sum_{i=1}^{N} a_i K(x, x_i) + b \quad (8)$$

The selection and construction of kernel functions is a key issue which greatly influences the performance of LS-SVM, and provides an important approach to expand LS-SVM from linear field to nonlinear field. At present, there are several commonly used kernel functions [5]:

1) Linear kernel:

$$K(x, x_i) = x^T x_i \quad (9)$$

2) Q-order polynomial kernel function:

$$K(x_i, x_j) = (\gamma x_i^T x_j + r)^q, \gamma > 0 \quad (10)$$

3) RBF function:

$$K(x_i, x_j) = \exp(-\frac{\left\| x_i - x_j \right\|^2}{\sigma^2}) \quad (11)$$

4) Sigmoid kernel function:

$$K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r) \quad (12)$$

III. STRUCTURE OF THE WATER QUALITY RETRIEVAL

For a long time, quantitative retrieval of the inland water quality has been the difficulty for water research using remote sensing. The lack of an effective means of monitoring and evaluation system make the existing measurement data can not be well analyzed and mined [1].In this paper, LS-SVM model was applied to the field of water quality monitoring. Tai Lake was as an example accepted by this study. The original information included Landsat5 TM remote sensing image data and the synchronous ground survey data on May 4, 1997.Twelve monitoring points were set up; the measured data were acquired from these monitoring points. The process of the retrieval model based on LS-SVM was shown in Fig.1.

The association information between the remote sensing and ground-monitoring were used to build the information processing model, which was fused by the ground survey data and remote sensed data. The model can provide good generalization ability and reduce the complexity of the system. As the Fig.1 shows, the sensing data, which is corresponding to the ground monitoring points on the remote sensing image and the ground survey data, were composed to the training sets, the input vector is sensing data, and the target vector is the ground survey data. The pre-prepared training sets was used to train LS-SVM, and then the required retrieval region on the remote sensing image will be input into the trained LS-SVM pixel-by-pixel to retrieve. Finally, the water quality retrieval results of the part of or the entire water field will be obtained.

IV. THE THE EXPERIMENT AND ANALYSIS

In this paper, the experimental platform was Matlab 7.1; the concentration of the suspended matter of Tai Lake was retrieved on this platform. The LS-SVM lab contains Matlab/C implementation for a number of LS-SVM algorithms related to classification, regression and time-series prediction. The LS-SVM lab is an open-source code package. It provides a large number of interfaces for classification and regression. In this paper, the LS-SVMlab1.5 code package was added into the Matlab toolbox, which used the interface provided by LS-SVM to realize the regression function.

A. The Selection Parameters of the LS-SVM

The kernel function and its parameters γ and σ^2 of the LS-SVM have great effect to the performance of the algorithm [4], and there is no simple way to determine the parameters. In this paper, the grid-searching and cross-validation were used to select the parameters of the LS-SVM automatically. Among such kernel functions, RBF kernel function is a local kernel function with a stronger learning ability [10], and it can achieve a good balance between the computing time and the effectiveness of the prediction. So the RBF function was accepted by this paper. Because there were only two parameters, so the search space of parameters was reduced to two-dimensional from tree-dimensional of the SVM, and the speed of modeling was accelerated.

After the repeated experiment, the search scope of γ was $[2^{-2}, 2^{-1} \cdots 2^{10}]$, and σ^2 was $[2^{-2}, 2^{-1} \cdots 2^{5}]$ .For each γ and σ^2 group, the cross-validation evaluation algorithm can be used, γ and σ^2 as a set of parameters would be searched, whose algorithm performance was the best.

So a two-dimensional network was constituted in the

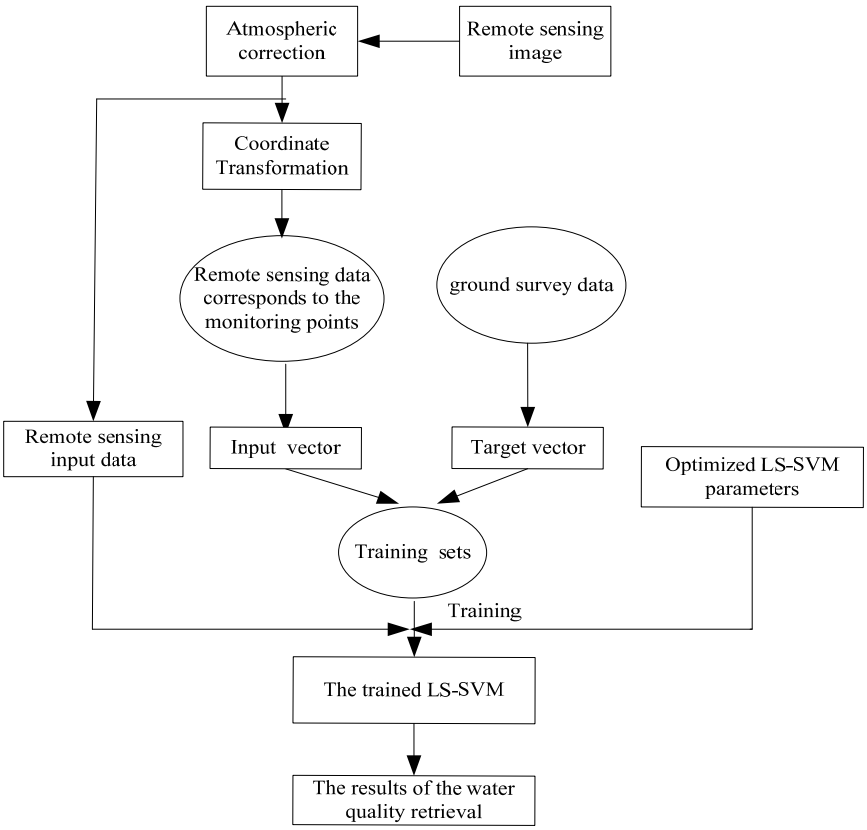


Figure 1.The structure of water quality retrieval model based on LS-SVM

coordinate system. A small scope search area was chosen to search the two parameters, which were set at the center of the area, consequently a set of optimal parameters was searched. In the process of conducting cross-training, the value of k was taken 4. Every time, one of the three sets of the nine monitoring points as the training sample, the rest of the three data as the testing sample. After the final trials, the value of $\gamma = 100$ and $\sigma^2 = 0.5$.

B. The Implication of the Water Quality Retrieval

The ground survey and remote sensing data was fused by this paper, LS-SVM model was built to complete the retrieval of the concentration of suspended matter. The ground survey data is the concentration of suspended matter value, which is also the output of the model, was measured in Tai Lake. After the atmospheric correction, the remote sensing data was used by this model.

By analyzing the reflection spectra of the Tai Lake water, the reflectance values in the vicinity of 580nm and 810nm of the concentration of suspended matter is more sensitive than others [8]. TheTM1toTM4 are just within the scope of this spectrum. The reflectivity of 12 monitoring points and the corresponding ground survey data of the concentration of suspended matter were listed in the Table I. The content of the Table I is the data for training. From the Table I, it can be seen that ρ_1 is the reflectivity of the TM1, so are the others.

The interface of the trainlssvm() input the training data of the Table I and the two optimal parameters into LS-SVM to train, a new model will be produced to predict the results, as a result the interface of simlssvm() will be called. After that, a new file called *output _ file* will be produced, in which is the retrieval concentration values of the suspended matter.

The Table II showed the measured concentration and the retrieval concentration. According to the experiments, the retrieval concentration of suspended matter map is shown in the Fig.2.

From the Table II, it can be seen that the retrieval value of the monitoring point was close to measured values. From the Fig. 2 it can be seen that in the central

TABLE I.
THE DATE OF TRAINING

Monitoring Points	Measured concentration	ρ_1	ρ_2	ρ_3	ρ_4
1	46.0	0.029	0.039	0.052	0.037
2	107.0	0.040	0.052	0.065	0.053
3	14.0	0.038	0.047	0.041	0.037
4	18.0	0.038	0.047	0.047	0.048
5	41.0	0.036	0.049	0.054	0.053
6	35.0	0.038	0.049	0.058	0.045
7	15.0	0.039	0.051	0.050	0.043
8	15.0	0.039	0.051	0.065	0.056
9	22.0	0.020	0.021	0.027	0.045
10	16.0	0.019	0.021	0.032	0.054
11	41.0	0.037	0.049	0.055	0.069
12	25.0	0.015	0.015	0.017	0.048

TABLE II.
THE CONCENTRATION BETWEEN THE MEASURED AND THE RETRIEVAL RESULTS (MG/L)

Monitoring Points	Measured concentration	Retrieval concentration
1	46.0	45.9968
2	107.0	107.997
3	14.0	13.9974
4	18.0	18.0034
5	41.0	40.9966
6	35.0	35.0029
7	15.0	15.0029
8	15.0	14.8045
9	22.0	22.0033
10	16.0	16.0029
11	41.0	40.9969
12	25.0	24.9969

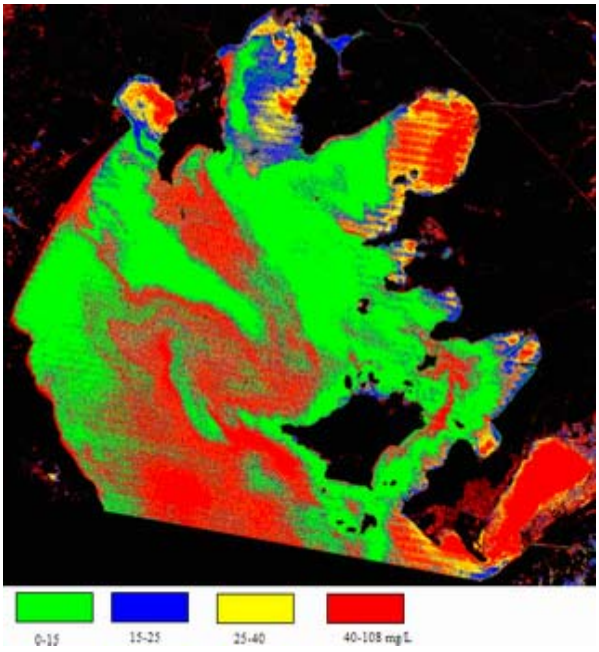


Figure 2. Retrieval of the suspended solids concentration map

and south-east of the Tai Lake the concentration of the suspended matter was high. In the southeastern area where was the active exchange of the lake water and the area where other rivers go into the lake, the suspended matter was also high. The Fig.2 also told that, the pollution of the lake bank district was high too. These places were regions that the eutrophication was serious, this was due to human activities such as the sewage discharge, and paddock breeding would accelerate the development of the eutrophication. In western and eastern areas of cardiac side, the concentration of the suspended matter is lower, eutrophication was not very serious, and other area was between in the two areas.

V. CONCLUSION

Referred to analysis of relationship between the remote sensing data and the ground survey data of Tai lake, this paper established the water quality monitoring model based on LS-SVM theory and algorithm.

The concentration of the suspended matter in Tai lake was monitored by the proposed method. The results indicated that, the proposed modeling method was simple, the adjustments of the parameters were convenient and the speed of the learning was fast. The non-linear retrieve system, which was established by the LS-SVM method, can give a high precision, so it was able to satisfy the demand of the water quality monitoring.

ACKNOWLEDGMENT

This paper was supported by National Natural Science Foundation of China (No. 60774092 and 60901003).

REFERENCES

[1] SHI Ai-ye, and XU Li-zhong, “Remote sensed images fusion and lake water quality identification based on neural networks and evidence theory,” *Journal of Image and Graphics*, vol.10, no.3, pp. 18-20, Mar., 2005.

[2] LIU Dong, and ZHENG Ning, “Feature selection and model parameters optimization for SVM based on genetic algorithm,” *Computer Application and Software*, vol.26, no.1, pp. 30–37, Jan., 2009.

[3] Zhang Youjing, and Chen Liang, “Estimating urban impervious surfaces using LS-SVM with multi-scale texture,” *2009 Urban Remote Sensing Joint Event*, 2009.

[4] WU De-hui, and YANG Shi-yuan, “A dynamic system identification method for sensors based on LS-SVM,” *Journal of Electronic Measurement and Instrument*, vol.20, no.6, pp.1653-1662, Dec., 2006.

[5] LIAO Shi-Zhong, and DING Li-zhong, “Simulataneous tuning of mutiple parameters for support vector regression,” *JOURNAL OF NANJING UNIVERSITY, NATURAL SCIENCES*, vol.45, no.5, pp. 405-410, Sept., 2009.

[6] Karasuyama M, and Nakano R, “Optimizing SVR hyper parameters via fast Cross-Validation using AOSVR,” *Proceedings of the International Joint Conference on Neural Networks*, New York, pp. 1186-1191, 2007.

[7] Hu B H, and Zhang H J, “A cloud amount forecasting method based on local linear estimation,” *Natural Sciences*, vol. 45, no.1, pp.120-128, 2009.

[8] Gin K Y-H, and KOH S T, “Spectral irradiance profiles of suspended marine clay for the estimation of suspended sediment concentration in tropical waters,” *Int J Remote Sensing*, vol.24, no.6, pp. 3235~3245, 2003.

[9] K.Iwashita, and K.Kudoh, “Satellite analysis for water flow of Lake Inbanuma,” *Advances in Space Research*, 2004, pp.93-98.

[10] I. Steinwart, and D.Hush, “An explicit of the reproducing kernel hibert spaces of Gaussian RBF kernels,” *IEEE Transaction on Information Theory* , vol.52, no.10, pp. 4635-4643, Oct., 2006.

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 190-192

A Study on High-Strength Communication Scheme Based on Signed Digital Envelope

Wenping Guo, Ying Chen, and Xiaoming Zhao
School of Mathematics and Information Engineering, Taizhou University, Linhai, China
Email: guo_wp@126.com , ychen222@163.com , tzxyzxm@yahoo.com.cn

***Abstract*—The digital signature and digital envelope, two specific applications of computer technology, have been widely used in data communication, but the application of either one may cause some deficiencies in practice. Based on the current cryptographic techniques and the study on digital signature and digital envelope, this paper proposed a new scheme, signed digital envelope. It can overcome the shortcomings of the two technologies that bring along, and the security of data communication is thereby greatly strengthened.**

***Index Terms*—Digital signature, Digital envelope, Signed digital envelope**

I. INTRODUCTION

With widespread use of the Internet, people are more aware of the network security. How to efficiently block intruders' access to personal computer and their alteration of unauthorized data is one of the burning problems at present. In the field of data communication, data security is mainly concerning three indicators, confidentiality, integrity and undeniability, but the conventional digital signature and digital envelope still need some improvements before achieving the above- mentioned three indicators. In this paper, we took advantages of digital signature and digital envelope's role in data protection and proposed a high-secure scheme, signed digital envelope, which ensures data communication against being intruded.

II. SIGNED DIGITAL ENVELOPE

A. Digital Signature

Digital signature (DS) [1] is a mathematical scheme for demonstrating the authenticity of a digital message or document. A valid digital signature gives a recipient reason to believe that the message was created by a known sender, and that it was not altered in transit. It typically consists of three algorithms: (1) A key generation algorithm that selects a private key uniformly at random from a set of possible private keys. The algorithm outputs the private key and a corresponding public key. (2)A signing algorithm which, given a message and a private key, produces a signature. And (3) A signature verifying algorithm which given a message, public key and a signature, either accepts or rejects.

Two main properties are required. First, a signature generated from a fixed message and fixed private key should verify on that message and the corresponding public key. Secondly, it should be computationally infeasible to generate a valid signature for a party who does not possess the private key.

Digital signature is used to ensure that the message in transit is integrated, and that the source of the message is authenticated and undeniable. Digital signature uses what is known as public key algorithm. The use of public key cryptography for digital signature is similar with public key encryption. It has two keys, one is signing key which must be kept private, referred to as private key, the other is verification key which is open, referred to as public key. The big computing workload slows down the computing speed of public key algorithm, so we can use secure one-way hash function to summarize the message so as to lessen the computing workload. Figure 1 illustrates the formation of digital signature. Figure 2 illustrates the verification of digital signature.

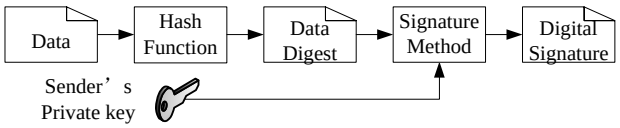


Figure 1. Formation of Digital Signature

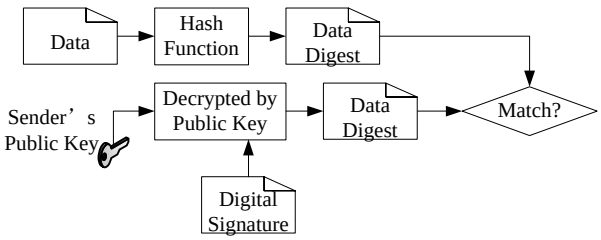


Figure 2. Verification of Digital Signature

B. Digital Envelope

Digital Envelope [2] is a type of security that uses two layers of encryption to protect a message. First, the message itself is encoded using symmetric encryption, and then the key to decode the message is encrypted using public-key encryption. This technique overcomes one of the problems of public-key encryption, which is that it is slower than symmetric encryption. Because only the key is protected with public-key encryption, there is very little overhead.

With the advantages of private key encryption and public key encryption, digital envelope can overcome private key's distribution difficulty in private key encryption and public key's long encryption. The flexibility of the public key and the high efficiency of the

AP-PROC-CS-10CN006

private key that obtained by using two layers of encryption can guarantee the security of messages. Figure 3 illustrates the encryption process of the digital envelope. First, the plaintext is encrypted by symmetric key, and then symmetric key is encrypted by receiver's public key so as to obtain digital envelop, that is, the data ciphertext and key ciphertext. [1]

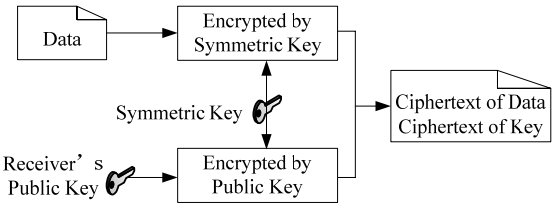


Figure 3. Encryption Process of the Digital Envelope

C. Signed Digital Envelope

The conventional digital envelope can ensure the confidentiality of the data, but its adoption of the public key encryption makes it possible for malicious user to destroy the data, so the integrity and undeniability of the data can not be guaranteed. However, signed digital envelope can make it up for conventional digital envelope due to its adoption of private key encryption. And a new and improved encryption scheme is born by combining digital envelope with digital signature

Figure 4 illustrates the improved scheme. First, the digital signature is obtained by signing plaintext data with sender's private key, and then plaintext data are encrypted by symmetric key, and after that, both the signature and the symmetric key are encrypted by sender's public key. In this way, only the authorized receiver has access to the initial data that can not be altered by malicious user. Figure 5 illustrates the process that the authorized user receives the data and has them verified.

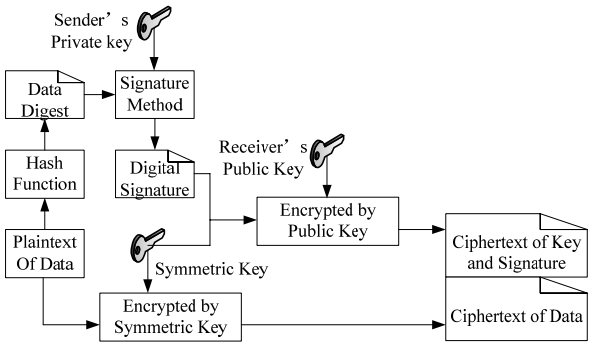


Figure 4. Improved Encryption Scheme

III. SECURE COMMUNICATION BASED ON SIGNED DIGITAL ENVELOPE

The .Net provides all the developers with first-class encryption model which has great scalability and doesn't have to resort to that obscure function from the unmanaged class. Therefore, .Net was chosen as the development environment to bring about a new scheme in this paper.

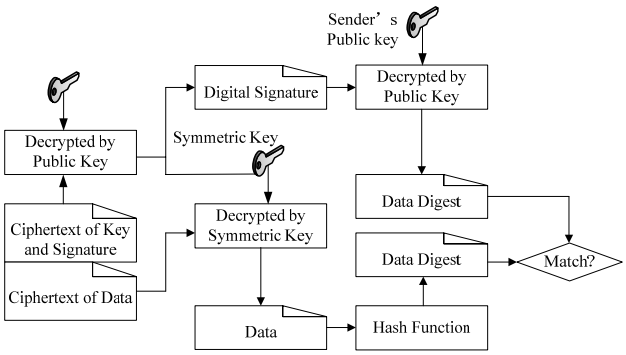


Figure 5. Improved Decryption Scheme

A. Methods for Symmetric Encryption and Decryption

The entire document should be in Times New Roman or Times font. Type 3 fonts must not be used. Other font types may be used if needed for special purposes.

The steps of symmetric encryption are shown as follows:

（1）Creating encryption key

TripleDESCryptoServiceProvider crypt = new TripleDESCryptoServiceProvider();

……

……

crypt.GenerateKey();//creating symmetric key

（2）Creating encrypted object and memory stream

MemoryStream ms = new MemoryStream（）;

……

（3）Packing the memory stream with CrytoStream and encrypting it.

CryptoStream cs = new CrytoStream（ms,crypt.CreateEncryptor（）, CryptoStreamMode. Write）; // packing the memory stream with CrytoStream

StreamWriter sw = new StreamWriter（cs）;

sw.WriteLine ("plaintext"); // importing plaintext data

sw.Flush（）;

cs.FlushFinalBlock（）;

……

Both decryption and encryption need to pack memory with CryoStream, but in this regard, CreateDecrytor proves itself better than CreateEncyptor in practice.

B. Methods for Public Key Encryption and Decryption

Being similar with digital signature, here we can use RSACryptoServiceProvider class and DSACryptoServiceProvider. But digital signature is to use one's own private key to encrypt the data, and what is known as asymmetric encryption, in most cases, refers to the use of counterpart's public key to encrypt the data. Different encryption keys can be applied to different circumstances.

The steps of public key encryption are shown as follows:

（1）Creating key pair

RSACryptoServiceProvider crypt = new RSACryptoServiceProvider();

……

string strPublicKey = crypt.ToXmlString(false)；

// public key

string strPrivateKey = crypt.ToXmlString(true)；

// private key

（2）Code-conversing the data

……

byte[] bytes = Encoding.GetEncoding("GB2312").GetBytes(“plaintext”)；//importing plaintext data and symmetric key

（3）Encrypting with Encrypt ( ) of RSA-CryptoServiceProvider

……

byte[] cipherBytes = crypt.Encrypt(bytes, false)；

When decrypting with this method, we only use Decrypt () of RSACryptoServiceProvider.

C. Data Signature and the Verification

.NET Framework provides signed data and documents with RSACryptoServiceProvider and DSACrypto-ServiceProvider class.

That is, RSACryptoServiceProvider crypt = new RSACryptoServiceProvider() importing one’s own private key for signing.

The theory of singing with crypt.FromXmlString-(strPrivateKey), as a matter of fact, is using plaintext to get hash code and then encrypting the hash code with one’s own private key. Based on the plaintext, we can use SignData () to get hash code and signed data.

byte[] signBytes = crypt.SignData(bytData, offset, count, SHA1.Create())； //SHA1.Create() enable us to obtain hash code of the data.

The data verification is accomplished by comparing the re-generated hash with the original hash to determine whether the data have been altered or not. After importing counterpart’s public key, the verification results prove that the new scheme is up to expectation by using VerifyData(bytData, SHA1.Create(), bytSignature).

Ⅳ. CONCLUSIONS

This paper studied the two conventional computer security technologies, digital signature and digital envelope, based on which, a more secure scheme was constructed. It is proved that the communication data can meet the three indicators, confidentiality, integrity and undeniablility. Our next step is to put the new scheme into practical application; however, it is necessary to point out that the .Net, as the development environment, provides first-class encryption model and encryption class, making the implementation of the scheme become much easier.

REFERENCES

[1] [Wikipedia’s digital signature，]Wikipedia page: Digital signature, last modified on 1 September 2009, http://en.wikipedia.org/wiki/Digital_signature

[2] PKCS #7 Version 1.6 Bulletin:Cryptographic message syntax, RSA Laboratories[S].

[3] ZHAO Yan-bo, ZHANG Xue-jie, JIANG Yong-ling. Application researches on high strength file encryption based on digital envelope. *Computer Engineering and Design*[J].2007.9:4357-4359

[4] Gu C. X., Zhu Y. F., Zhang Y. J. A Formal Model for the Security of Proxy Signature Schemes. *Wuhan University Journal of Natural Science*, 10(1), 2005: 275-278.

[5] Ureche O, Plamondon R. Document transport, transfer and exchange:Security and commercial aspects [C].ICDAR.Washington,DC,USA:*IEEE Computer Society*,1999:585-588.

[6] Okamoto T., Inomata A., Okamoto E. A Proposal of Short Proxy Signature Using Pairing. In: *International Conference on Information Technology (ITCC 2005)*. USA: IEEE press, 2005: 631－635.

[7] Huang X.Y., Mu Y., Susilo W., et al. A Short Proxy Signature Scheme: Efficient Authentication in the Ubiquitous World. In: *EUCWorkshops 2005*. Berlin: Springer-Verlag, 2005: 480－489.

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 193-196

A Framework for Credibility Evaluation of Web-Based Competitive Intelligence

Jie Zhao[1, 2], and Peiquan Jin[3]
[1]School of Business Administration, Anhui University, Hefei, China
[2]School of Management, University of Science and Technology of China, Hefei, China
zjpq@mail.hf.ah.cn
[3]School of Computer Science and Technology
University of Science and Technology of China, Hefei, China
jpq@ustc.edu.cn

***Abstract*—With the rapid increasing of Web data volume, it has been a hot issue to find competitive intelligence from Web. However, previous studies in this direction focused on the extracting algorithms for Web-based competitive intelligence, and little work has been done in the credibility evaluation of Web-based competitive intelligence. The credibility of Web-based competitive intelligence plays critical roles in the application of competitive intelligence and can determine the effectiveness of competitive intelligence. In this paper, we focus on the credibility evaluation issue of Web-based competitive intelligence, and present a framework to model and evaluate the credibility of Web-based competitive intelligence. A networked credibility model and some new methods for credibility evaluation are proposed in the paper.**

I. INTRODUCTION

With the rapid development of Web technologies, the amount of Web data has reached a very huge value. A recent report said that the volume of Web pages has been 200, 000 TB. The huge volume of Web data brings new challenges in many areas, especially in the extraction and application of competitive intelligence. Researches on Competitive intelligence mainly focus on extracting useful competitive intelligence from a large set of data, and thus provide support for enterprise management and decisions. A survey in 2007 shown that about ninety percentage of competitive intelligence could be acquired from the Web (Lamar, 2007). Therefore, many researchers devoted themselves in the extraction of Web-based competitive intelligence.

However, previous work in Web-based competitive intelligence mostly concentrated on how to extract competitive intelligence from Web pages (Deng and Luo, 2007), and little work has been done in the credibility evaluation of Web-based competitive intelligence. The credibility of Web-based competitive intelligence is a critical issue in the application of competitive intelligence, because incredible intelligence will bring wrong developing strategies management mechanisms to enterprises.

In this paper, we will focus on the credibility evaluation issue of Web-based competitive intelligence. Our goal is to develop a system to automatically collect and evaluate Web-based competitive intelligence. This paper is not a system demonstration, but a description of our designing framework. We will discuss the system architecture of the credibility evaluation of Web-based competitive intelligence, and pay more attention to the details of concrete modules in the system. The main contributions of the paper can be summarized as follows:

(1) We present a Web-based framework for the extraction and credibility evaluation of competitive intelligence. The major components of such a system are analyzed (see Section 3).

(2) We present a new technique to perform the credibility evaluation of Web-based competitive intelligence, which is the spatiotemporal-evolution-based approach (see Section 4).

The following of the paper is structured as follows. In Section 2 we discuss the related work. Section 3 discusses the framework of Web-based competitive intelligence extraction and credibility evaluation. Section 4 gives the discussion about spatiotemporal-evolution-based approach to evaluating the credibility of Web-based competitive intelligence. And conclusions and future work are in the Section 5.

II. RELATED WORK

According to our knowledge, there are few works focused on the credibility evaluation of competitive intelligence. Most of previous related works concentrated on information credibility. Basically, competitive intelligence stems from information. But competitive intelligence credibility is different from information credibility. There is some relationship between those two types of credibility, which is still an unrevealed issue in the research on competitive intelligence.

Information credibility refers to the believability of some information and/or its source (Metzger, 2007). It not only refers to the objective evaluation on information quality and precise, but also refers to the measurement on information source. Recently, Web information credibility has been a hot topic and some works have been conducted. The earliest research on this area can be found in (Alfarez and Hailes, 1999), in which the authors present a new method considering some trust mechanism in society to measure the information credibility.

However, most works in Web information credibility were published after 2005.

There are some prototypes in Web information credibility evaluation, among which the most famous ones are WISDOM (Akamine et al., 2009) and Honto?Search (Yamamoto and Tanaka, 2009). WISDOM extracts information from Web pages and clusters them according to senders and opinions. It is designed as a computer-aided tool to help users to determine the credibility of querying topics. Honto?Search is a Web Q/A system. It allows users to input a query about some fact and delivers the clustered analysis on the given fact. It is also a computer-aided system to help users evaluate information credibility. Besides, HONCode (Fritch, 2003) and MedPICS (Eysenbach, 2000) are two prototypes in the medical domain which also support Web information credibility evaluation. HONCode is built by the NGO Health on the Net Foundation. It can help users to find the credible medical websites, which are trusted by some third-party authoritative organization. The third-party-based evaluation method is very common is some specific areas, such as electronic commerce. MedPICS allows website owner to add some trust tags in the Web pages. And then users can filter Web information based on the trust tags in Web pages. For example, they can require that only the information whose trust tags are higher that a certain value be returned to them.

Previous works usually focus on different contents in the Web. Most researchers paid attention to the Web news credibility, searched results credibility, and products information credibility. Those works are generally based on Web pages and try to compute the credibility of Web pages. For example, Google News (Google, 2005) uses the trustiness of news posters to evaluate the news credibility. Some news websites adopt a vote-based approach to measure the news credibility, such as www.126.com and www.sohu.com. There are also some works concerning Web information quality and the credibility of Web information sources. A lot of people also make investigation on Web information credibility. For example, a survey in 2004, which is focused in the electronic commerce area, shown that about 26% American posted comments on products in the Web (Rainie and Hitlin, 2007), which indicates that users' comments is a key factor in the information credibility evaluation.

The basic methods used in Web information credibility evaluation can be divided into four types, which are the Checklist method, the cognitive authority method, the iterative model, and the credibility seal programs. The Checklist method uses a checklist to perform a user survey and then to determine the information credibility. This method is usually not practical in real applications. For example, some checklists contain too many questions that will consume too much time of users (Metzger, 2007). The cognitive authority method pays much attention to the authority of information. It is similar with the checklist method, except that it usually utilizes some automatic tools. For example, it suggests users use the Whois, Traceoute, and other tools to evaluate the authority of the information senders and websites. The iterative model evaluates information credibility through three steps. First, it checks the appearance of the website. Second, it measures some detailed factors, including the professional level, the trustable level, the freshness, precise, and relevance to users' needs. Finally, users are required to mark the evaluated results. The similarity between the iterative model and the Checklist method is that both of them provide some criteria for users to mark the information credibility. The difference between them is that the iterative model pays more attention to the importance of the information receiver in the evaluation process. The credibility seal program is much different from other three methods. It provides some credibility seal program to help users to find credible sources in the Web. For example, the HONCode can help users to find trusted medical websites (Fritch, 2003). However, this method is usually restricted in certain areas due to the huge amount of Web information.

III. WEB-BASED COMPETITIVE INTELLIGENCE EXTRACTION AND CREDIBILITY EVALUATION

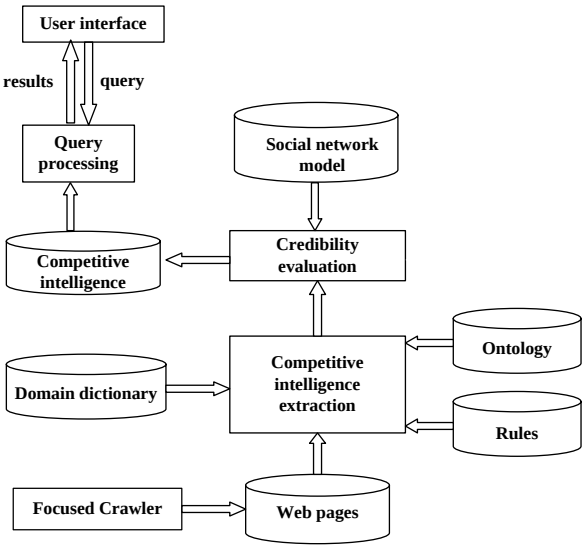


Figure 1. A Framework of Web-based Competitive Intelligence Extraction and Credibility Evaluation

According to practical applications, competitive intelligence must be connected to some specific domain. For example, many competitive intelligence softwares provide intelligence analysis towards a given domain. Based on this assumption, we propose a domain-constrained system to extract and evaluate Web competitive intelligence (as shown in Fig.1). The system consists of four modules, which are the competitive intelligence extraction module, the credibility evaluation module, the query processing module, and the user interface. The basic running process of the system is as follows. First we use a focused crawler to collect Web pages about a specific application domain. Then we extract competitive intelligence from Web pages according to some rules, the competitive intelligence

ontology, and the domain dictionary. After that, the credibility of the extracted competitive intelligence is evaluated based on a defined social-network-based credibility model. When users submit queries about competitive intelligence through the user interface, the query processing model will retrieve appropriate results from the competitive intelligence database.

The main function of the competitive intelligence extraction module is to extract domain-constrained competitive intelligence from Web pages and further to deliver them to the credibility evaluation module. We use an entity-based approach in this module to extract competitive intelligence, which will be discussed in the next section.

The credibility evaluation module adopts the social-network-based method to evaluate competitive intelligence credibility. The credibility of competitive intelligence is influenced by a lot of factors. These factors are classified into two types in our paper, which are the inner-site factors and inter-site factors. Then we use different algorithms to evaluate the competitive intelligence credibility according each type of factors. Finally we will integrate the both results and make a comprehensive evaluation on the competitive intelligence credibility.

The user interface supports keyword-based queries on competitive intelligence. Users are allowed to input topics, time, or locations as query conditions.

The query processing module aims at returning competitive intelligence related with given topics or other conditions. Competitive intelligence workers can further process the returned results and produce integrated competitive intelligence. This module contains two procedures. The first one is a database retrieval procedure, and the second is clustered visualization of the results. The system provides several ways of clustered visualization, including time-based clustering, location-based clustering, and topic-based clustering.

IV. A Spatiotemporal-Evolution-Based Approach to Credibility Evaluation

Most Web pages contain time and location information. That information is useful to evaluate the credibility of Web-based competitive intelligence. Our approach consists of two ideas. The first idea is to construct a timeline for Web-based competitive intelligence, and further to develop algorithms to determine the credibility of Web-based competitive intelligence. The second idea is to construct a location distribution of related Web pages, which contribute the extraction of Web-based competitive intelligence. And then we use the location distribution to measure the credibility of Web-based competitive intelligence.

Fig.3 and Fig.4 shows the two ideas. In Fig.3, we see that the number of Web pages related to a certain competitive intelligence element increased sharply in July, 2009, based on which we can infer that this element is very possible to be true from July, 2009. While in Fig.4, we find that in July, 2009, the IP locations of related Web pages are mostly in China. Suppose that the competitive intelligence is about China, we can draw the conclusion that in July, 2009, the competitive intelligence gained the most attention in China, so it is very possible that the competitive intelligence is credible.

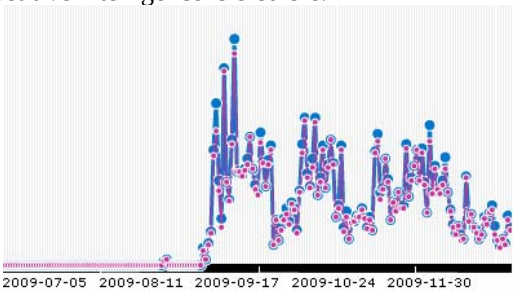


Figure 2. The timeline evaluation of Web-based competitive intelligence

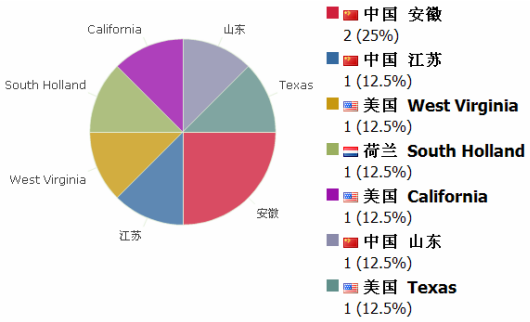


Figure 3. The location distribution of Web-based competitive intelligence

V. Conclusions and Future Work

In this paper we present a framework of extracting and evaluating Web-based competitive intelligence. Our future work will concentrate on the design and implementation of the algorithms mentioned in the paper and aim at implementing a prototype and conducting experimental study.

VI. Acknowledgement

This work is supported by the Open Projects Program of National Laboratory of Pattern Recognition, the Key Laboratory of Advanced Information Science and Network Technology of Beijing, and the National Science Foundation of China under the grant no. 60776801 and 70803001.

References

[1] Akamine, S., Kawahara, D., Kato, Y., et al., (2009) WISDOM: A Web Information Credibility Analysis System, *Proc. of the ACL-IJCNLP 2009 Software Demonstrations*, Suntec, Singapore, pp. 1-4

[2] Alfarez, A., Hailes S., (1999) Relying On Trust to Find Reliable Information, *Proc. of International Symposium on Database, Web and Cooperative Systems (DWA-COS)*

[3] Deng, D., Luo, L., (2007) An Exploratory Discuss of New Ways for Competitive Intelligence on WEB 2.0, In W. Wang (ed.), *Proc. of IFIP International Conference on e-*

Business, e-Services, and e-Society, Volume 2, Springer, Boston, pp. 597-604

[4] Eysenbach, G., (2000) Consumer Health Informatics, *British Medical Journal*, (320), pp. 1713-1716

[5] Fritch, J.W., (2003) Heuristics, Tools, and Systems for Evaluating Internet Information: Helping Users Assess a Tangled Web, *Online Information Review*, Vol.27(5), pp. 321-327

[6] Google, (2005) Google News Patent Application, In: http://www.webpronews.com/news/ebusinessnews/wpn-45-20050503GoogleNewsPatentApplicationFullText.html. Accessed in 2010.3

[7] LaMar, J., (2007) Competitive Intelligence Survey Report, In: http://joshlamar.com/documents/CIT Survey Report.pdf

[8] Metzger, M., (2007) Making Sense of Credibility on the Web: Models for Evaluating Online Information and Recommendations for Future Research, *Journal of the American Society of Information Science and Technology*, Vol. 58(13), pp. 2078-2091

[9] Mikroyannidis, A., Theodoulidis, B., Persidis, A., (2006) PARMENIDES: Towards Business Intelligence Discovery from Web Data, In *Proc. Of WI*, pp.1057-1060

[10] Rainie, L., Hitlin, P., (2007) The Use of Online Reputation and Rating Systems, In: http://www.pewinternet.org/ PPF/r/140/report_display.asp, Accessed in 2010.3

[11] Yamamoto, Y., Tanaka, K., (2009) Finding Comparative Facts and Aspects for Judging the Credibility of Uncertain Facts, *Proc. Of WISE*, , LNCS 5802, pp. 291-305

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp.197-200

Power Aware Job Scheduling with Quality of Service Guarantees: A Preliminary Study

Congfeng Jiang[1], Xianghua Xu[1], Jian Wan[1], Jilin Zhang[1], and Yinghui Zhao[2]
[1] Grid and Service Computing Technology Lab, Hangzhou Dianzi University, Hangzhou, 310037, China
Email :{cjiang,xuxh,wanjian,zhangjilin}@hdu.edu.cn
[2] Department of Hydraulic Engineering and Agriculture, Zhejiang Tongji Vocational College of Science and Technology, Hangzhou, 311231, China
Email: zhaoyinghuihust@gmail.com

***Abstract*—In large scale server systems, Quality of Service (QoS) performance cannot be guaranteed if only the objective is to minimize the total power consumptions separately. In this paper a power aware feedback control based power aware job scheduling algorithm (PFM) is proposed to minimize power consumption in computing system and to provide QoS guarantees. Simulations and experiments show that the proposed algorithm can reduce power consumptions significantly while still providing QoS guarantees and the performance degradation is acceptable. The results also show that conventional hardware based per-component and system-wide power management methods can save more power consumptions if they are in assistance with job-level adaptation.**

***Index Terms*—power aware computing system; job scheduling; Quality of Service (QoS); feedback control**

I. INTRODUCTION

Power consumption has been a major concern for not only large scale computing system but also mobile and embedded systems powered by batteries. With the scale of computing system increases, power consumption has become the major challenge to system performance and reliability [1, 2, 3, and 4].

With DFS/DVS support, processors are simply switched to a sleep mode while transitioning between frequencies and voltages and the system performance will be heavily deteriorated by the transition delay between various frequency and voltage levels in uniprocessors. This situation becomes even worse in multi core processor which does not support per-core DFS/DVS [5]. Conventional hardware based per-component and system-wide power management methods cannot save considerable power consumptions because they are coarse-grained and not adaptive to various fluctuating workloads in real scenarios. Moreover, the system performances, for example, availability, responsiveness, and throughput, do not scale with the number of processors but the power consumption does. Most unfortunately, the performance of the whole system can be deteriorated greatly if the objective is only to minimize the total power consumptions separately, despite of the violations of Quality of Service (QoS) requirements.

In this paper, we investigate the feasibility of hardware-software joint regulation for a feedback control based power aware job scheduling algorithm and test the algorithm by simulations and real workload. In the proposed algorithm, the feedback based power-aware job scheduler regulates the job dispatching and system performance dynamically for different workload characteristics. A testbed for investigating this scheme is implemented on a web server using Intel quad core processor. The performance and overhead of the algorithm are assessed under different workload. The results show the potential of power consumption reductions for hardware-software joint adaptations. We use a multimeter with USB connection cable to measure the real time power consumption of the system and the measurement results show that fine-grained job-level power aware scheduling can reduce considerable power consumptions. The results also show that the best algorithm varies for different experiment settings and real workload scenarios. In particular, feedback control based power aware job scheduling algorithm is not always better than other algorithms in all the performance dimensions. The performance depends on many factors such as the accuracy of workload characterization, the arrival rate of jobs, frequency-voltage transition delays in specific processors, the frequency/voltage levels available, etc.

The remainder of this paper is organized as follows: In section 2 we propose the controlling framework of the power aware job scheduling scheme. In section 3 we present the feedback control based power aware job scheduling algorithm model with QoS guarantees. Then, in Section 4, we present simulation results and real platform experiment results of the proposed scheduling algorithm. We also compare the performance data with conventional power-unaware job scheduling algorithms or job scheduling algorithms without QoS guarantees. Finally, we summarize the work in Section 5.

II. FEEDBACK CONTROL MODEL

There has been increasing research effort in applying control-theoretic approaches to power and performance management for computer systems such as internet web servers, databases and storage systems. Since today's large scale servers and applications are highly dynamical and change load conditions frequently, feedback control designs may provide desired performance guarantees.

In the power aware feedback mechanism, at the beginning of a job scheduling round, the scheduler

AP-PROC-CS-10CN006

chooses the average value of previous performance data as the controlled variable for our simple feedback mechanism. The average value is only the initial value for the future scheduling and it is a heuristic setting which can provide a minimum guarantee for QoS satisfactions. In the following scheduling period, each time when a job completes, the real time performance is sensed and collected by the feedback scheme and fed to the controller. And the following performance values are computed based on the initial value of performance threshold using moving average computation.

Here are the pseudo codes of the single feedback mechanism.

1. When scheduling event occurs {
2. for each task in the task set
3. Compute estimated performance data of each task and its execution time
4. End for
5. Compute estimated power consumption of each job on specific processor through code profiling
6. Compute QoS gains of each job
7. for each task in the task set
8. Schedule the tasks with minimum QoS gains and power consumptions
9. Delete the task from the task set
10. Update job table
11. If cpu_queue of targeted processor is exceeded the maximum length
12. Insert the task into next scheduling tasks set
13. Update job table
14. End if
15. Else
16. Insert the task into next scheduling tasks set
17. Update job table
18. End if
19. End for
20. }

Figure 1. The pseudo codes of SFM

III. POWER AWARE JOB SCHEDULING WITH QOS GUARANTEES

We use a periodic and independent task model in our framework. Let $J = \{J_i \mid i = 1,2,3,...,n\}$ denote jobs set, $J_i = (a_i, b_i, e_i, c_i, s_i, Q_i)$, M is set of processors, $M = \{M_j \mid j = 1,2,3,...,m\}$, $M_j = (p_j, f_j, d_j, BW_j, PW_j)$.

Where:

a_i is arrival time of job J_i, b_i is starting time of job J_i, e_i is the average execution time of job J_i on all the processors, i.e. expected executed time on processor M_j where there is no other running jobs except for J_i, $e_i = (e_i^1, e_i^2, e_i^3, ..., e_i^m)$, c_{ij} is the expected completion time of job J_i on processor M_j, s_i is the size of data needed by job J_i (MB), Q_i is the QoS of job J_i.

As for hosts set M, p_j is the speed of processor M_j (MHz), f_j is the available memory capacity of host M_j (MB), d_j is the available disk space on host M_j (MB), BW_j is the bandwidth of host M_j (Mb/s), PW_j is the level of power consumption of host M_j, $PW_j = (PW_j^1, PW_j^2, PW_j^3, ..., PW_j^u)$, $\forall v \in [1,u], 0 < PW_j^v < 1$, where 0 stands for the lowest and 1 the highest.

Let Q denote a set of Quality of Service constraints, $Q = \{Q_i \mid i = 1,2,...,n\}$, $Q_i = (T_i, R_i, S_i, A_i, P_i)$, where:

T_i is timeliness requirement, R_i is reliability requirement, S_i is security requirement, A_i is accuracy requirement, P_i (Priority) is priority requirement.

For simplicity, we use discrete values to modeling the Quality of Service constraints, i.e., the QoS constraint is presented by several levels like very low, low, medium, high, and very high, not a specific number like 10% or 90% because in real computing system with user interaction a user only cares the interactive experience, not the specific performance numbers.

Let $ec^J = (ec_1, ec_2, ..., ec_m)$ denote the power consumption of m threads, and the matrix of n performance counters in m threads is $C = [c_{i,j}](1 \le i \le m, 1 \le j \le n)$.

We define G_i^j is the gains of QoS of job J_i on host M_j, i.e.

$$G_i^j = \sum_{k=1}^{q} w_i^k \cdot g(Q_i^k, V_j^k) \qquad \text{(Eq.1)}$$

Where w_i^k is the weights of different QoS requirements of job J_i, and $\sum_{k=1}^{q} w_i^k = 1$; $g(Q_i^k, V_j^k)$ is the kth gains of QoS requirements of job J_i:

$$g(Q_i^k, V_j^k) = \begin{cases} Q_j^k - V_i^k & , when\ V_j^k > Q_i^k \\ 0 & , when\ V_j^k < Q_i^k \end{cases} \qquad \text{(Eq.2)}$$

Where V_j^k is the available QoS capacity of the corresponding host.

We define D_i as the available theoretical scheduling set of job J_i with QoS satisfactions, $\exists D_i = (D_i^j \mid G_i^j > 0)$

In order to avoid the QoS contention, the gains of QoS satisfactions must be minimized while still guarantying the QoS requirements. Assume that

$$OP_i = (D_i^j \mid G_i^j > 0) \cap \min\left\{\sum_{k=1}^{q} w_i^k \cdot g(Q_i^k, V_j^k)\right\} \qquad \text{(Eq.3)}$$

Then the objective function for power and QoS constrained scheduling for job set J_i is

$$S = \min\left\{\sum_{J_i \in J} OP_i\right\} \cap \min(\max\{_{T_i \in T}(c_i)\}) \qquad \text{(Eq.4)}$$

Eq.4 is NP-hard and can be solved by heuristics scheduling. With the above task model, we use heuristics scheduling algorithm to solve this problem of power aware job scheduling with QoS constraints.

IV. Simulation Results and Performance Analysis

A. Simulation Setup and Parameters Settings

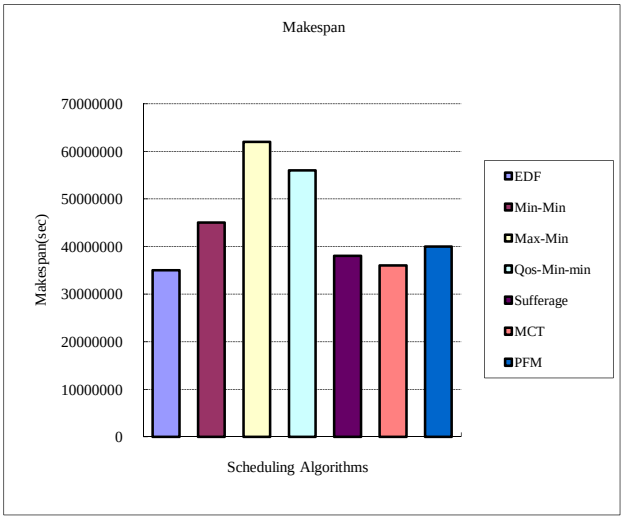
Feedback control based scheduling algorithm is capable of modifying its own scheduling decision and program behavior through time, depending on the execution characteristics. Here, the objective of the feedback control is to schedule jobs to processors while guarantying the QoS requirements and minimizing the total power consumptions of the computing system, preserving its simplicity and low overhead. We test this algorithm and describe its behavior under a number of workloads. Simulations include an analysis of the performance sensibility with the variation of the control parameters and its application in a multi processor computing system. Although the processors of a Massively Parallel Processing system such as a computing cluster or a supercomputer may slightly differ in clock frequency and available memory, these differences usually have little influence on most applications in practice [6]. Hence, in the simulation we use an MPP with multiple machines as the testbed and it is feasible for job migration.

B. Simulation Results and Analysis

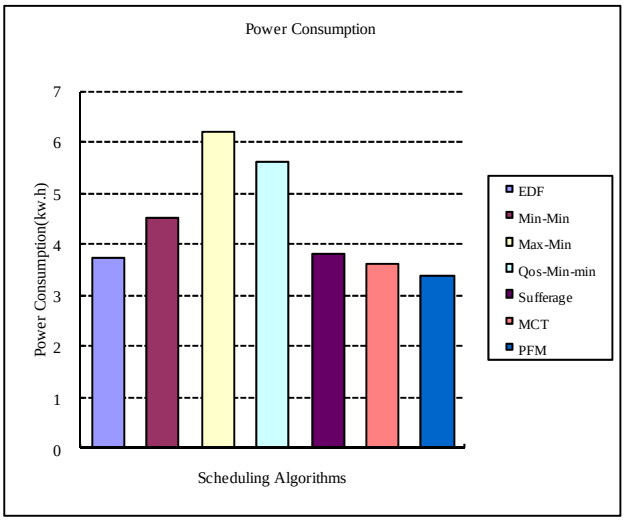
We use simulations and real workload experiments to study the performance of the proposed scheduling algorithm. Extensive sets of tests are executed to simulate the performance of our feedback scheduling power-aware framework under overload and under loaded conditions. In simulations, we use various matrix and vectors to modeling the controlling system, including task sets (i.e., arrival rate, task's period, deadline, actual execution time, worst case execution time, estimated execution time, etc), QoS requirements, power consumptions, and other system parameters. In a simulation setting, all tasks being scheduled are assumed to be periodic and each task's actual execution time is assumed to be known in advance and is indicated in matrix declared in the same .m program. Each task indicates its QoS needs quantitively to the scheduler, which in turn is able to know the global QoS requirements of the system to meet all task requirements. At the initial round when scheduling events occur (the threshold is reached and the scheduling is triggered), the scheduler make the task scheduling decision and the processor voltage/frequency scaling. After several scheduling periods, the scheduler makes the scheduling decision and adjusts the processor frequency and voltage level according to the feedback information collected from the sensor units.

The common approach to study the performance of the scheduling algorithm is to compare it with a non-power-aware or non-QoS-aware scheduling algorithm. Thus we studied and compared the performance of the simple and frequently used heuristics such as EDF, Min-min [7], Max-min [7], QoS guided Min-Min [8], Sufferage [9], and MCT (Minimum Completion Time [10], with PFM by testing various scenarios.

To evaluate the scheduling algorithm, we use the following metrics: Makespan, Average waiting time, Scheduling success rate, Power consumption, Average violating rate of QoS, Average migration rate, and Overall utilization. Due to space limitation, we only provide the makespan and the Power consumption results here.



(a) Makespan



(b) Power consumption

Figure 2. Relative performance

In Figure 2(a), the makespan order of the scheduling algorithms from maximum to minimum is: (1) Max-Min, (2) QoS-Min-min, (3) Min-Min, (4) PFM, (5) Sufferage, (6) MCT, and (7) EDF. The makespan of EDF is the smallest because of its smallest computation consumption. PFM dynamically schedules jobs to computing sites according to the real time power consumption and QoS constraints. Thus the makespan of PFM is relatively large.

In Figure 2(b), the power consumption order of the scheduling algorithms from highest to lowest is: (1) Max-Min, (2) Qos-Min-min,(3)Min-Min,(4)Sufferage, (5)EDF, (6)MCT, and (7)PFM. PFM has the lowest power consumption because it takes into account the real time power consumption when scheduling tasks.

V. CONCLUSIONS

In this paper we propose a power aware job scheduling algorithm based on feedback control theory and characterize workload through data gathered from trace data to capture possible application behavior. The job scheduling model is independent of implementation platforms and therefore feasible for future applications on multi-processor systems. By exploring the nature of dependence of server performance on time-varying load and operating conditions, the proposed general framework is possibly applicable to a diverse spectrum of server-based applications.

In the future, several extensions can be made further under different dimensions. For example, the scheduled tasks will be extended to be real tasks whose actual execution time is not known until its completion. And the scheduler can be implemented as a kernel module in an OS to provide power aware job scheduling.

ACKNOWLEDGMENT

The funding support of this work by Natural Science Fund of China (Grant No. 60873023), Science and Technology Research and Development Program of Zhejiang Province, China (Grant No. 2009C31033), Key Program of Science and Technology Planning of Hydraulic Bureau of Zhejiang Province, China (Grant No. RB0927), and Hangzhou Dianzi University Startup Fund (Grant No. KYS 055608058) are greatly appreciated.

REFERENCES

[1] C. Lefurgy, K. Rajamani, F. Rawson, W. Felter, M. Kistler, and T.W. Keller, “Energy management for commercial servers,”Computer, Vol.36, pp.39-48,2003.

[2] X. Wang, M. Chen, C. Lefurgy,and T.W. Keller, “SHIP: Scalable hierarchical power control for large-scale data centers,”in Proceedings of 2009 18th International Conference on Parallel Architectures and Compilation Techniques(PACT’09),pp.91-100,2009.

[3] X. Wang,and M. Chen , “Cluster-level feedback power control for performance optimization,” in Proceedings of IEEE 14th International Symposium on High Performance Computer Architecture(HPCA 2008), pp.101-110, Feb. 2008.

[4] Report to Congress on Server and Data Center Energy Efficiency. U.S. Environmental Protection Agency, ENERGY STAR Program, August 2, 2007.Available at:http://www.energystar.gov/ia/partners/prod_development/downloads/EPA_Datacenter_Report_Congress_Final1.pdf

[5] Y. Wang,K. Ma, and X. Wang, “Temperature-constrained power control for chip multiprocessors with online model estimation,” in Proceedings of the 36th annual international symposium on Computer architecture (ISCA’09),pp.314-324,June 2009.

[6] C. Franke, J. Lepping, and U. Schwiegelshohn, “Greedy scheduling with custom-made objectives,”Annals of Operations Research,in press.

[7] T.D. Braun, H.J. Siegel, N. Beck,L. L. Bölöni,M. Maheswaran,A.I. Reuther,J.P. Robertson, et al., “A Comparison of Eleven Static Heuristics for Mapping a Class of Independent Tasks onto Heterogeneous Distributed Computing Systems, ” Journal of Parallel and Distributed Computing, Vol. 61,pp.810-837,2001.

[8] X. He, X. Sun, and G.V. Laszewski, “QoS guided Min-Min heuristic for grid task scheduling, ”Journal of Computer Science and Technology, Vol.18,pp. 442-451,2003.

[9] M. Maheswaran, S. Ali, H.J. Siegel, D. Hensgen, and R.F. Freund, “Dynamic mapping of a class of independent tasks onto heterogeneous computing systems,” Journal of Parallel and Distributed Computing, Vol. 59, pp.107-131,1999.

[10] L.D. Briceño , M. Oltikar, H.J. Siegel, and A.A. Maciejewski, (). “Study of an iterative technique to minimize completion times of non-makespan machines,” in Proceedings of 2007 IEEE International Parallel and Distributed Processing Symposium (IPDPS 2007),pp.1-14,2007

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 201-205

Networked Manufacturing Resources Optimization Deployment for Complicated Parts

Wenli Peng[1*], Zhongbao Qin[2], and Wenni Zhang[3]
[1]College of Electro-mechanics Engineering, Jiaxing University, Jiaxing, China
Email: peng_wenli@163.com
[2]Xi'an Institute of high-technology, Xi'an, China
Email: zhongb_qin@163.com
[3]College of Foreign Languages, Jiaxing University, Jiaxing, China
Email: zhangwenni.pc@163.com

***Abstract*—Agile of physical manufacturing unit is the competitive advantage in the networked manufacturing environment. It is believed that the agile can be reached by optimization deployment of networked manufacturing resources. To solve this problem, logical manufacturing unit (LMU) and logical manufacturing process (LMP) are put forward and defined to decompose and model networked manufacturing task according to the process of complex part. When selecting manufacturing resources for these manufacturing tasks, many factors should be taken into account. However, manufacturing cost, time to market and manufacturing quality are the most important factors. In this paper, networked manufacturing resources pre-allocate is carried out to find candidate manufacturing resources based on abilities of manufacturing resources, such as part family, geometric feature, material type, rough type, dimension range, machining method, precision grade and production type. Then, taking transportation time and cost besides manufacturing time, cost and quality into consideration, the target and constraint of manufacturing resources optimization deployment are analyzed, and manufacturing resources optimization deployment problem is considered as a multi-objectives optimization problem.**

***Index Terms*—manufacturing resources optimization deployment, physical manufacturing unit, logical manufacturing unit, logical manufacturing process, executive manufacturing process**

I. INTRODUCTION

With the development of network and information technologies, the manufacturing of complex products can be achieved based on agile manufacturing through collaborating of enterprises. In this paper, the model of networked manufacturing resources optimization deployment was built in considering the processing cost, time, manufacturing quality [1]. But in the networked manufacturing environment, the physical manufacturing units are geographically distributed, so transportation cost and time between different physical manufacturing units are significant enough not to be ignored, and should be taken into consideration[2]. Thus, the model of networked manufacturing resources optimization deployment was much more complicated [3]. This resembles the partner selection in virtual enterprise, but there were many differences. In our views, before those factors were considered, manufacturing resources abilities such as part family, geometric feature, material type, rough type, dimension range, machining method, precision grade and production type, which can be realized by physical manufacturing units, should be taken into consideration. We call this process as the networked manufacturing resources pre-allocate[4].

II. NETWORKED MANUFACTURING TASK DECOMPOSITION AND DESCRIPTION

Definition 1: Physical manufacturing unit (PMU) is composed of all physical equipments located at the same place, including machining equipments, fixture accessories, and computers and so on. The basic information of PMU includes location information, workshop name, enterprise name, typical products, personnel information, contact method and other main information.

A. Designing logical manufacturing unit

Definition 2: Logical manufacturing unit (LMU) expresses the manufacturing sub-process, and is defined as the aggregation of work procedures. The information model of LMU is described by eight elements: LMU(*Pf*, *F*, *Ma*, *R*, *D*; *Mach*, *Pre*, *Pro*).

Where: *Pf* expresses part family; *F* expresses geometric feature; *Ma* expresses material type;*R* expresses rough type; *D* expresses dimension range; *Mach* expresses machining method; *Pre* expresses precision grade; *Pro* expresses production type.

From the definition of LMU and the characteristics of these eight elements, a LMU has the following characteristics: All elements inside a LMU have close relationship each other, but the relationship between every two LMU-s is loose; each work procedure is included in a LMU, but a LMU can include several work procedures [5].

B. Designing logical manufacturing process

Definition 3: Logical manufacturing process (LMP) is defined as the abstract description of the whole production procedure of a complex part. It is composed of a sequence of some LMU-s and the dynamic information of each LMU, namely the restrictions of

*Corresponding author 86-573-83647635; Project Number: 70508003

AP-PROC-CS-10CN006

machining time, machining quality, machining cost and other especial machining tools or fixtures[6].

Because LMU includes eight elements, and each element has much different information, through combining different information of each element, many LMU-s will be produced and LMU model database will be set up. Thus, when designing the LMP, designers only choose the appropriate LMU-s according to the manufacturing process of a complex part from LMU model database, in which each LMU has been designed and defined, so the LMU information of LMP is relatively static[7]; after choosing the appropriate LMU, designers should evaluate and compute the machining time, machining quality, machining cost and point out the especial machining tools or fixtures for this LMU, and these information are dynamic, because they are designed according to this part[8]. The design of LMP is shown in Fig.1.

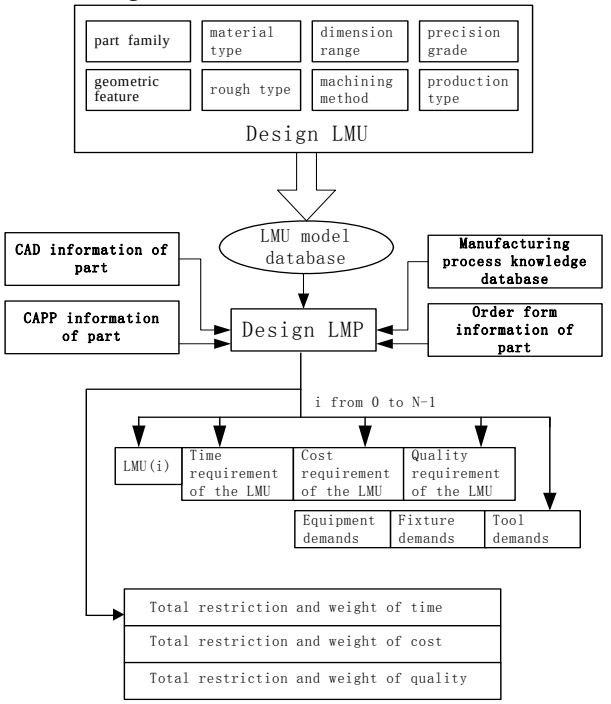


Fig.1 Designing flow of LMP

C. Information model of networked manufacturing task

So, for a complex part, which has N manufacturing sub-tasks, the information model of the whole manufacturing task is described by the following formula:

$$LMP=\{LMU(0),C_0,T_0,Q_0,F_0,E_0;LMU(1),C_1,T_1,Q_1,F_1,E_1;\cdots; \\ LMU(N-1),C_{N-1},T_{N-1},Q_{N-1},F_{N-1},E_{N-1}\} \quad (1)$$

Where: C_i,T_i,Q_i are the cost, time and quality restrictions for finishing the $LMU(i)$; E_i and F_i are the requirements for especial equipments and fixtures; i is an integer, $i\in[0,N-1]$.

III. FORMULATION FOR COMPLICATED PARTS

A. Problem description

Definition 4: Executive manufacturing process (EMP) expresses the whole practical manufacturing process of a complex part, and is the mapping result between PMU and networked manufacturing sub-tasks (LMU-s). EMP is corresponding to LMP. It is a sequence of several PMU-s, and includes the detail manufacturing resources information of PMU-s. The mapping process is shown in Fig.2.

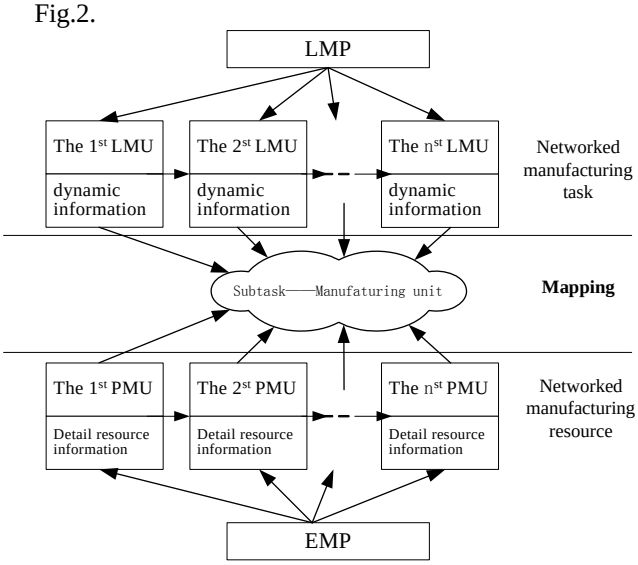


Fig.2 The mapping process of logical manufacturing tasks and

For example shown in Fig.3, if a physical manufacturing unit marked as PMU1 accepted a manufacturing order of a complicated part, after analyzing the process of this part, process planners decompose the whole task to N+1 sub-task in considering itself resources. Among these sub-tasks, two sub-tasks marked as LMU (0) and LMU (N) can be completed by PMU1. For other manufacturing sub-tasks, the cooperative PMU-s should be deployed in considering the total running time, cost and the total manufacturing quality.

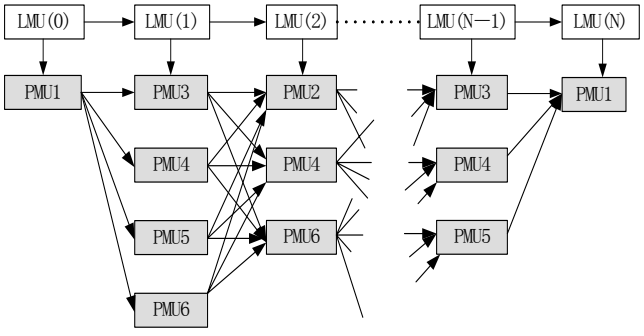


Fig.3 The problem description of networked manufacturing resources optimization deployment

B. Mathematics model

1) Selection fields of PMU and EMP

For a complicated part, after designing the LMP, according to the detail information of each LMU, PMU-s that can complete this LMU will be selected out and these PMU-s are called as PMU subset. The mapping result for the i*-the* subtask— $LMU(i)$ in LMP is marked as:

$$G_i=\{PMU(i,0),PMU(i,1),\cdots,PMU(i,m_i-1)\} \quad (2)$$

Where: G_i is the selection field of PMU corresponding to $LMU(i)$; m_i is the number of PMU in G_i; i is an integer and ranges from 0 to N-1; N is the number of sub-tasks, namely the quantity of LMU in LMP.

So, for the whole networked manufacturing task−LMP described in (1), all PMU subsets can be obtained and described as:

$$G = \{G_0, G_1, \cdots, G_{N-1}\} \tag{3}$$

This also is the selection fields of EMP,

2) Objective functions

For the whole networked manufacturing tasks formulated by (1), EMP is orderly composed of N PMU-s, marked as:

$$\begin{aligned} EMP = \{&PMU(0, k_0),\ PMU(1, k_1), \cdots, \\ &PMU(N-1, k_{N-1})\} \end{aligned} \tag{4}$$

Where: $k_i \in [0, m_i - 1]$; $PMU(i, k_i)$ is the k_i *-th* PMU in G_i; m_i is the number of PMU in G_i.

In order to obtain the optimal EMP, each EMP should be evaluated in terms of some qualitative or quantitative criterions. In this paper, the following three criterions will be considered: the least total running cost—C, the least total running time—T and the best total machining quality — Q. Thus, the objectives functions are formulated by (5).

$$\begin{cases} \min C = \min(C_{in} + C_{tr}) = \min(\sum_{i=0}^{N-1} C_i(j) + \sum_{i=0}^{N-2} C_{i,\ i+1}(j,k)) \\ \min T = \min(T_{in} + T_{tr}) = \min(\sum_{i=0}^{N-1} T_i(j) + \sum_{i=0}^{N-2} T_{i,\ i+1}(j,k)) \\ \min Q = \min \sum_{i=0}^{N-1} (1 - Q_i(j)) \end{cases} \tag{5}$$

Where: $PMU(i, j)$ is the j *-th* in G_i, $j \in [0, m_i - 1], k \in [0, m_{i+1} - 1], i \in [0, N-1]$

m_i is the number of PMU in G_i, $C_i(j)$ is the machining cost inside $PMU(i, j)$, $C_{i,i+1}(j,k)$ is the transportation cost between $PMU(i, j)$ in G_i and $PMU(i+1, k)$ in G_{i+1}, $T_i(j)$ is the machining time inside $PMU(i, j)$, $T_{i,i+1}(j,k)$ is the transportation time between $PMU(i, j)$ in G_i and $PMU(i+1, k)$ in G_{i+1}, $Q_i(j)$ is the machining quality, which is guaranteed by $PMU(i, j)$, expressed by the rate of certified parts quantity divided by all parts quantity.

3) Restriction conditions

For LMP, R and D are assumed as the released time and the delivery time. C^T and Q^T are respectively assumed as the total cost restriction and the total quality restriction. During *[R, D]*, all LMU-s in LMP must be completed. Similarly, LMU (i) has a released time R^i and a delivery time D^i, and has the cost, time and quality restrictions expressed by C^i, T^i and Q^i. During $[R^i, D^i]$, all work procedures of LMU (i) must be completed. Thus, the restriction conditions of networked manufacturing resources optimization deployment for complicated part include:

The sequence restriction of sub-tasks：if LMU(i) must be completed before LMU(j), then

$$R^i + T^i \le D^i \le R^j. \tag{6}$$

The running time restrictions of LMU (i):

$$\begin{cases} \max T_i(j) \le T^i \le (D^i - R^i) \\ \max(\sum_{i=0}^{N-1} T_i(j) + \sum_{i=0}^{N-2} T_{i,\ i+1}(j,k)) \le (R - D) \end{cases} \tag{7}$$

The running cost restrictions of LMU (i):

$$\begin{cases} \max C_i(j) \le C^i \\ \max(\sum_{i=0}^{N-1} C_i(j) + \sum_{i=0}^{N-2} C_{i,\ i+1}(j,k)) \le C^T \end{cases} \tag{8}$$

The total quality restrictions of LMU (i):

$$\begin{cases} \min \sum_{i=0}^{N-1} Q_i(j) / N \ge Q^T \\ \min Q_i(j) \ge Q^i \end{cases} \tag{9}$$

Ⅳ. OPTIMIZING DEPLOYMENT ALGORITHMS BASED ON GENETIC ALGORITHM

GA is widely applied in advanced manufacturing fields, such as the optimum scheduling of workshop resources, the partner selection of virtual enterprise. In this paper, for the convenience and speediness of computation, we transform the multi-objectives optimization problem formulated by (5) into a single objective optimization problem to solve with GA. Exterior information is not almost used when searching the best solution by GA, and only the value of fitness function is taken into consideration. In order to make the optimization result impersonal and precise, we introduce the relative membership degree to the best member of fuzzy mathematics into the designing of relative fitness function.

A. Coding method

According to the characteristics of networked manufacturing resources optimization deployment, the integer coding method is applied in GA. The coding regulations include:

- Coding the manufacturing subtask: the code is 0,1,…,N-1 (N is the number of subtasks, namely the number of LMU in LMP).
- A gene of a chromosome corresponds to a LMU in LMP, and the length of a chromosome, namely the quantity of genes in a chromosome, is equal to the quantity of LMU in LMP.
- Coding the manufacturing resources: the code is 0, 1,…, $m_i - 1$, where m_i is the quantity of PMU in G_i.
- A gene value of a chromosome corresponds to a manufacturing resource, namely a PMU.
- x_i (the i *-th* gene value of a chromosome) expresses the x_i *-th* PMU in G_i, and the corresponding manufacturing subtask is the i *-th* LMU in LMP. x_i is integer and $0 \le x_i \prec m_i$.
- A chromosome corresponds to an EMP. For the chromosome expressed by (10), the corresponding EMP expressed by (11).

$$X = [x_0 x_1 \cdots x_i \cdots x_{N-1}] \tag{10}$$

$$EMP=\{PMU(0,x_0),\ PMU(1,x_1),\cdots PMU(i,x_i)\cdots,\ PMU(N-1,x_{N-1})\} \quad (11)$$

Where, $PMU(i,x_i)$ stands for the x_i-*th* PMU in G_i.

B. Designing the relative fitness function

1) Influence degree and influence *degree coefficient*

Usually, EMP that simultaneously satisfies the three objectives does not exist. So the weights should be assigned to the three objectives. The three weights are expressed by W_C, W_T and W_Q:

$$W_C+W_T+W_Q=1 \quad (12)$$

According to the coding regulations, a chromosome (an individual) is described as $X=[x_0x_1\cdots x_i\cdots x_{N-1}]$, which stands for an EMP. Where: x_i is an integer, $0\le x_i\prec m_i$, and m_i is the number of PMU in G_i. For each chromosome (individual), the (C,T,Q) value can be computed through (13) in every generation. Then, the objective eigenvalue matrix is computed and described as (14).

$$\begin{cases} C(X)=C_{in}+C_{tr}=\sum_{i=0}^{N-1}C_i(x_i)+\sum_{i=0}^{N-2}C_{i,i+1}(x_i,x_{i+1}) \\ T(X)=T_{in}+T_{tr}=\sum_{i=0}^{N-1}T_i(x_i)+\sum_{i=0}^{N-2}T_{i,i+1}(x_i,x_{i+1}) \\ Q(X)=\sum_{i=0}^{N-1}(1-Q_i(x_i)) \end{cases} \quad (13)$$

$$\begin{bmatrix} C(X_1) & C(X_2) & \dots & C(X_M) \\ T(X_1) & T(X_2) & \dots & T(X_M) \\ Q(X_1) & Q(X_2) & \dots & Q(X_M) \end{bmatrix} \quad (14)$$

$C_{\max},C_{\min},T_{\max},T_{\min},Q_{\max}$ and $Q_{\min}$ can be computed based on the eigenvalue matrix of a generation.

For a new generation, each individual is newly produced. The authors define this influence as the influence degree of the objective non-standardization, marked as E_C, E_T and E_Q, computed by (15):

$$\begin{cases} E_C=W_C\cdot\left(\frac{C_{\max}-C_{\min}}{C_{\max}+C_{\min}}\right)^{\alpha} \\ E_T=W_T\cdot\left(\frac{T_{\max}-T_{\min}}{T_{\max}+T_{\min}}\right)^{\alpha} \\ E_Q=W_Q\cdot\left(\frac{Q_{\max}-Q_{\min}}{Q_{\max}+Q_{\min}}\right)^{\alpha} \end{cases} \quad (15)$$

Where: $\frac{C_{\max}-C_{\min}}{C_{\max}+C_{\min}}$, $\frac{T_{\max}-T_{\min}}{T_{\max}+T_{\min}}$ and $\frac{Q_{\max}-Q_{\min}}{Q_{\max}+Q_{\min}}$ respectively express the relative changing range of C, T and *Q*. α is the adjusting factors. Considering that the weights have the more direct influence on the optimization result than the eigenvalue relative changing range usually α is given a positive value less than 1. In this paper, $\alpha=0.5$.

Then, the three influence degrees expressed by (15) are compared with each other, and the maximum and minimum influence degree can be obtained, marked as $E_{\max}$ and $E_{\min}$. The influence degree coefficient $e(e_c,e_t,e_q)$ is computed by (16).

$$\begin{cases} e_c=\frac{E_C-E_{\min}}{E_{\max}-E_{\min}} \\ e_t=\frac{E_T-E_{\min}}{E_{\max}-E_{\min}} \\ e_q=\frac{E_Q-E_{\min}}{E_{\max}-E_{\min}} \end{cases} \quad (16)$$

2) Relative membership degree based on influence degree coefficients

From (5), we can know that the less the three objective values of an individual are, the better the individual is. So, if $C_{\min},T_{\min}$ and $Q_{\min}$ are far to $C_{\max},T_{\max}$ and $Q_{\max}$ respectively, the strong relative membership degree of an individual respectively corresponding to $C_{\min}$, $T_{\min}$ and $Q_{\min}$ is defined as (17):

$$\begin{cases} r_{1C}(X)=\frac{C_{\max}-C(X)}{C_{\max}-C_{\min}} \\ r_{1T}(X)=\frac{T_{\max}-T(X)}{T_{\max}-T_{\min}} \\ r_{1Q}(X)=\frac{Q_{\max}-Q(X)}{Q_{\max}-Q_{\min}} \end{cases} \quad (17)$$

But, if $C_{\min},T_{\min}$ and $Q_{\min}$ are near to $C_{\max},T_{\max}$ and $Q_{\max}$ respectively, another function is defined to compute the relative membership degree, which is called as the weak relative membership degree shown in (18):

$$\begin{cases} r_{C2}(X)=1-\frac{C(X)}{C_{\max}+C_{\min}} \\ r_{T2}(X)=1-\frac{T(X)}{T_{\max}+T_{\min}} \\ r_{Q2}(X)=1-\frac{Q(X)}{Q_{\max}+Q_{\min}} \end{cases} \quad (18)$$

Finally, the relative membership degree based on influence degree coefficients $e(e_c,e_t,e_q)$ can be obtained by linearly superposing (17) and (18), shown as (19):

$$\begin{cases} r_C(X)=e_cr_{C1}(X)+(1-e_c)r_{C2}(X) \\ r_T(X)=e_tr_{T1}(X)+(1-e_t)r_{T2}(X) \\ r_Q(X)=e_qr_{Q1}(X)+(1-e_q)r_{Q2}(X) \end{cases} \quad (19)$$

When $e(e_c,e_t,e_q)$ is equal to $e(1,1,1)$, (19) is becoming (17), which expresses the objective changing range has the most great influence on the optimization result.

From (17), (18) and (19), the objective eigenvalue matrix described as (14) can be transformed into the relative membership degree matrix described as (20):

$$\begin{bmatrix} r_C(X_1) & r_C(X_2) & \dots & r_C(X_M) \\ r_T(X_1) & r_T(X_2) & \dots & r_T(X_M) \\ r_Q(X_1) & r_Q(X_2) & \dots & r_Q(X_M) \end{bmatrix} \quad (20)$$

3) Relative fitness function based on relative membership degree

Extracting out the maximal value and the minimal value from each row of the matrix described as (20), which are marked as r_g and r_b:

$$r_g=\{r_{gC},r_{gT},r_{gQ}\}=\{\max_{i=1}^{M}r_C(X_i),\max_{i=1}^{M}r_T(X_i),\max_{i=1}^{M}r_Q(X_i)\}=\{1,1,1\} \quad (21)$$

$$r_b=\{r_{bC},r_{bT},r_{bQ}\}=\{\min_{i=1}^{M}r_C(X_i),\min_{i=1}^{M}r_T(X_i),\min_{i=1}^{M}r_Q(X_i)\}=\{0,0,0\} \quad (22)$$

If there is X_j in a generation, which meets the condition described as (23), the individual X_j is called the ideal excellent individual.

$$r_j = \{r_C(X_j), r_T(X_j), r_Q(X_j)\} = r_g \tag{23}$$

Similarly, if there is X_j in a generation, which meets the condition described as (24).

$$r_j = \{r_C(X_j), r_T(X_j), r_Q(X_j)\} = r_b \tag{24}$$

Thus, for individual X_i, $u_g(X_i)$ and $u_b(X_i)$ are defined as the membership degree of X_i to the ideal excellent individual and the ideal inferior individual. Both of them meet with the conditions described as (25):

$$\begin{cases} 0 \le u_g(X_i) \le 1 \\ 0 \le u_b(X_i) \le 1 \\ u_g(X_i) + u_b(X_i) = 1 \\ i = 1,..., M \end{cases} \tag{25}$$

In order to get the excellent membership degree of X, namely $u_g(X)$, we set up the optimization rule: the sum of the square of hamming weighted distance of X to the ideal excellent individual and the square of hamming weighted distance of X to the ideal inferior individual is minimal.

C. Algorithm design

1) Genetic Operators design

The first step in GA is computing the fitness value. Each chromosome has a selection probability, which is decided by the distribution of all individual fitness in a generation. The roulette wheel selection method is adopted to compute the selection probability in this paper. Supposing the fitness value for each chromosome is $f_k, (k = 1,2,\cdots,M)$, and the total fitness value in a generation is $\sum_{k=1}^{M} f_k$. Then, the selection probability for the k-th individual is $f_k / \sum_{k=1}^{M} f_k$; The single point crossover is applied to produce new individual for next generation. One integer from 0 to $N-1$ is decided by random functions, which expresses the crossover position. The typical crossover probability P_c ranges from 0.6 to 0.9; The mutation operator has some restriction conditions. The mutation is carried out at a position in the reasonable integer range.

2) Selection strategy and ending condition of algorithm

The M individuals selected from the current generation and the father generation according to the fitness value form the next generation. This new generation reserves the better individuals of the current generation and the father generation. The ending condition of GA is the fitness value in a new generation is convergent. At last, several optimized individuals, namely several EMP-s, will be exported, and the decision-maker can select one of them.

V. CONCLUSION

To research the networked manufacturing resources optimization deployment, the idea of LMU and LMP is put forward, and the information model of networked manufacturing assignments for a complex part is set up based on them; For convenience of networked manufacturing resources optimization deployment, the manufacturing ability information for PMU is encapsulated and expressed by LMU; To obtain the feasible candidate PMU-s, the pre-deployment between LMU and PMU is implemented by information mapping in considering the resources manufacturing abilities and the sub-tasks manufacturing requirements.

To realize networked manufacturing resources optimization deployment, manufacturing cost, time to market and manufacturing quality are the most important factors. In considering these criteria we model the optimization deployment problem by a multi-objectives optimization problem. Then, transforming the multi-objectives optimization problem into a single objective optimization problem, the mathematics model and implementing algorithm of manufacturing resources optimization deployment are studied based on GA. The recommended GA with a relative fitness function to decrease the influence of non-standardization on the optimization result can fast achieve the optimal solution of the mentioned problems with a probability. A typical example shows the algorithm has the better synthetic performance in both computational speed and optimality, and the computation results show its potential to the networked manufacturing resources optimization deployment.

REFERENCES

[1] Yao Changfeng, Zhang Dinghua, Peng Wenli, Hu Chuangguo, 2004, Study on Networked Manufacturing Resource Model Based on Physical Manufacturing Unit.CHINA MECHANICAL ENGINEERING, 15, 414-417.

[2] Camarinha-Matos, Afsarmanesh, Camarinha-Matos. 1999, Infrastructures for Virtual Enterprises: Networking Industrial Enterprises, Kluwer Academic Publishers, Boston, pp.3-14.

[3] Talluri, S., Baker, R.C., 1999, A framework for designing efficient value chain networks. International Journal of Production Economics, 62, 133-144.

[4] Dingwei Wang, Yung, K.L., Ip, W.H., 2001, A heuristic genetic algorithm for subcontractor selection in a global manufacturing environment. IEEE Transactions on Systems, Man and Cybernetics Part C, 31(2), pp. 189-98.

[5] Gunasekaran A., 1998, Agile manufacturing: enablers and an implementation framework. International Journal of Production Research, 36, 1223-1247.

[6] Brucker, P., Drexl, A., Mohring, R., Neumann, K., Pesch, E., 1999. Resource-constrained project scheduling: Notation, classification, models, and methods. European Journal of Operational Research, 112, 3-41.

[7] RAO Yun-qing，ZHU Chuan-ju，ZHANG Chao-yong, 2003, Resource Integration and Execution System in Shop Floor Supporting Networked Manufacturing. Computer Integrated Manufacturing System, 9, 120-1125.

[8] Silva C M, Biscaia E C, Jr. Genetic Algorithm Development for Multi-objective Optimization of Batch Free-radical Polymerization Reactors. Computers and Chemical Engineering [J], 2003, 27: p.1329-1344.

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 206-209

The Annihilator and its Structure in Lattice Implication Algebras

Hua Zhu[1], Weifeng Du[2]*, and Jianbin Zhao[1]
1.Department of mathematics, Zhengzhou University, Zhengzhou , China;
Email: zhuhua@zzu.edu.cn, zhaojianbin@zzu.edu.cn
2.School of Mathematics and Information Engineering, Jiaxing University, Jiaxing, China
Email: woodmud@tom.com

***Abstract*—The notion of annihilator of lattice implication algebras is proposed. An annihilator is proved to be an ideal and a *sl* ideal. Then the special characteristics of an annihilator are obtained. The relationships between an annihilator and an ideal, between the lattice implication homomorphism image of annihilator and the annihilator of lattice implication homomorphism image are discussed, respectively.**

***Index Terms*— lattice implication algebra; annihilator; ideal**

Ⅰ. INTRODUCTION

To establish an alternative logic for reasoning and knowledge representation, in 1993, Xu [7] proposed the concept of lattice implication algebra by combining lattice and implication algebra. Lattice implication algebra can respectively describe the comparable and the incompletely comparable properties of truth value, which can more efficiently reflect the people's thinking, judging and reasoning. Since then, many researchers[2], [4], [6], [8],[9], [12], [3], [11], [13], [14] have investigated this important logic algebra. For example, Jun [1] defined the notions of LI-ideals in lattice implication algebras and investigated its some properties. In 2003, Liu et al.[4] proposed the notions of ILI-ideals and maximal LI-ideal of lattice implication algebras, respectively, investigated their properties and obtained the extension theorem of ILI-ideals. In 2006, Zhu et al.[12] introduced the notions of prime ideal and primary ideal of lattice implication algebra respectively, researched their properties and discussed their relations. In 2008, Pan [11] introduced the notions of lattice implication n-ordered semigroup and lattice implication p-ordered semigroup, then discussed the properties of sl ideals in lattice implication n-ordered semigroups and lattice implication p-ordered semigroups. In this paper, as an extension of aforementioned work, In section 2, we list some basic concepts of lattice implication algebra which are needed for this topic. In section 3, we introduce the concept of annihilator and obtain some special characteristics of annihilator in lattice implication algebras. Then we prove that an annihilator is an ideal and a sl ideal , and discuss the relationships between an annihilator and an ideal, between the lattice implication homomorphism image of annihilator and the annihilator of lattice implication homomorphism image in lattice implication algebras, respectively.

Ⅱ. PRELIMINARIES

By a lattice implication algebra we mean a bounded lattice $(L,\vee,\wedge,O,I)$ with order-reversing involution $'$, I and O the greatest and the smallest element of L respectively, and a binary operation $\to$ satisfying the follow axioms:

(I_1) $x\to(y\to z)=y\to(x\to z)$;

(I_2) $x\to x=I$;

(I_3) $x\to y=y'\to x'$;

(I_4) if $x\to y=y\to x=1$, then $x=y$;

(I_5) $(x\to y)\to y=(y\to x)\to x$;

(L_1) $(x\vee y)\to z=(x\to z)\wedge(y\to z)$;

(L_2) $(x\wedge y)\to z=(x\to z)\vee(y\to z)$

for all $x,y,z\in L$.

If a lattice implication algebra L satisfies $\forall x,y,z\in L$, $x\vee y\vee((x\wedge y)\to z)=I$, then we call it a lattice *H* implication algebra.

Definition 2.1([10]). Let L be a lattice implication algebra. An ideal A is a non-empty subset of L such that for any $x,y\in L$,

(1) $O\in A$;

(2) $(x\to y)'\in A$ and $y\in A$ imply $x\in A$.

Lemma 2.2([10]). Let A be an ideal of a lattice implication algebra.

If $\forall x,y\in L$, $x\le y$ and $y\in A$, then $x\in A$.

Theorem 2.3([10]). Suppose A is a non-empty family of ideals of a lattice implication algebra L. Then $\bigcap$A is also an ideal of L.

Let A be a subset of a lattice implication algebra L. The least ideal containing A is called the ideal generated by A, denoted by $\prec A\succ$.

Specially, if $A=\{a\}$, we write $\prec A\succ=\prec a\succ$.

* Corresponding author: Du Weifeng , School of Mathematics & Information Engineering, Jiaxing University, Jiaxing, Zhejiang, China, Email:woodmud@tom.com

AP-PROC-CS-10CN006

Theorem 2.4([10]). Let L_1 and L_2 be lattice implication algebras, $f: L_1 \to L_2$ be a mapping from L_1 to L_2, if for any $x, y \in L_1$, $f(x \to y) = f(x) \to f(y)$ holds, then f is called an implication homomorphism from L_1 to L_2. If f is an implication homomorphism and satisfies

$$f(x \vee y) = f(x) \vee f(y),$$
$$f(x \wedge y) = f(x) \wedge f(y),$$
$$f(x') = (f(x))',$$

then f is called a lattice implication homomorphism from L_1 to L_2.

A one-to-one and onto lattice implication homomorphism is called a lattice implication isomorphism.

We can define a partial ordering $\leq$ on a lattice implication algebra L by $x \leq y$ if and only if $x \to y = I$.

In a lattice implication algebra L, $\forall x, y, z \in L$, the following hold(see[10]):

（1） $x \to y \leq (y \to z) \to (x \to z)$,
$x \to y \leq (z \to x) \to (z \to y)$;

（2） if $x \leq y$, then
$y \to z \leq x \to z, z \to x \leq z \to y$;

（3） $x \vee y = (x \to y) \to y$;

（4） $x \leq (x \to y) \to y$.

In the follows, if not special noted, L denotes a lattice implication algebra.

Ⅲ. THE ANNIHILATOR IN LATTICE IMPLICATION ALGEBRA

Firstly, we introduce the notion of the annihilator.

Definition 3.1. Let B be a non-empty subset of L, if

$$B^* = \{x \mid \forall b \in B, x \wedge b = O, x \in L\},$$

then B^* is called an annihilator of B.

The following example shows that the annihilator of lattice implication algebra exists.

Example 3.2. Let $L = \{0, a, b, c, d, 1\}$, $0' = 1, a' = c, b' = d, c' = a, d' = b, 1' = 0$, the Hasse diagram of L be defined as Fig.1 and its implication operator be defined as Table 1, then $(L, \vee, \wedge, ', \to)$ is a lattice implication algebra.

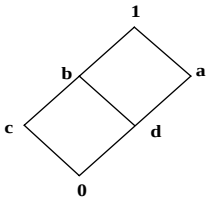


Fig.1

Table 1

→	0	a	b	c	d	1
0	1	1	1	1	1	1
a	c	1	b	c	b	1
b	d	a	1	b	a	1
c	a	a	1	1	a	1
d	b	1	1	b	1	1
1	0	a	b	c	d	1

Let B = {0, c}, then $B^* = \{0, a, d\}$ is the annihilator of B by simple computing.

Remark. Obviously, the annihilator of {O} is L.

Now, we give the important characteristics of an annihilator in lattice implication algebras.

Theorem 3.3. Let B be a non-empty subset of L, if B^* is an annihilator of B, then $\forall x \in L$, $b \in B$, $x \to b = x' \Leftrightarrow x \in B^*$.

Proof. Suppose that $\forall x \in L$, $b \in B$, $x \to b = x'$, then $(x \to b) \to x' = I$. It follows that $(b' \to x') \to x' = b' \vee x' = I$. Thus $x \wedge b = O$. Hence $x \in B^*$ by the Definition 3.1.

Conversely, if $x \in B^*$, then $\forall b \in B$, $x \wedge b = O$. It follows that $(b \wedge x)' = b' \vee x' = (b' \to x') \to x' = I$. Thus we have $b' \to x' \leq x'$. $x' \leq b' \to x'$ is trivial. Hence $b' \to x' = x'$, and so $x \to b = x'$.

Theorem3.4. Let $a \in L$, if $(a)^*$ is an annihilator of $\{a\}$, then $\forall x \in (a)^*$, $a \leq x'$.

Proof. Suppose that $(a)^*$ is an annihilator of $\{a\}$, then $\forall x \in (a)^*$, $x \to a = x'$ by Theorem 3.3. Since

$$x' \vee a \to (x \to a)$$
$$= (x' \to (x \to a)) \wedge (a \to (x \to a))$$
$$= (a' \to (x' \to x')) \wedge (x \to (a \to a)) = I,$$

we have $x' \vee a \leq x \to a$. It follows that $x' \leq x' \vee a \leq x \to a = x'$. Hence $x' \vee a = x'$, and so $a \leq x'$.

The following theorem prove that an annihilator is an ideal and a *sl* ideal in lattice implication algebras.

Theorem3.5. Let B be a non-empty subset of L. If B^* is an annihilator of B,then B^* is an ideal of L.

Proof. If B^* is an annihilator of B, then $O \in B^*$ is trivial by the Definition 3.1. Assume that $\forall x, y \in L$, $(x \to y)' \in B^*, y \in B^*$, then we have $\forall b \in B$, $y \to b = y'$ and $(x \to y)' \to b = x \to y$ by Theorem 3.3. It follows that

$$x' = I \to x' = ((y \to b) \to y') \to x'$$
$$= ((b' \to y') \to y') \to x'$$

$= (b' \vee y') \to x' = (b' \to x') \wedge (y' \to x')$

$= (b' \to x') \wedge (x \to y)$

$= (b' \to x') \wedge ((x \to y)' \to b)$

$= (b' \to x') \wedge (b' \to (x \to y))$

$= (b' \to x') \wedge (b' \to (y' \to x'))$

$= (b' \to x') = x \to b$. Thus $x \in B^*$ by Theorem 3.3 . Hence B^* is an ideal of L.

Corollary 3.6. Let B be a non-empty subset of L, if B^* is an annihilator of B, then B^* is a sl ideal of L.

Proof. it is trivial by Theorem 3.5 and Theorem 4.2 of [11].

Next, the special properties of an annihilator are obtained in lattice implication algebras.

Theorem3.7. Let B and C be non-empty subsets of L. Then the following hold

(1) if $B \subseteq C$, then $C^* \subseteq B^*$;

(2) $B \subseteq B^{**}$;

(3) $B^* = B^{***}$;

(4) $(B \cup C)^* = B^* \cap C^*$.

Proof. (1). Suppose that $B \subseteq C$. Then $\forall x \in C^*$, $c \in C$, $x \wedge c = O$, and so $\forall b \in B$, $x \wedge b = O$. Thus $x \in B^*$.

(2). Because $\forall b \in B$, $x \in B^*$, $x \wedge b = O$, then we have $b \in B^{**}$. Hence $B \subseteq B^{**}$.

(3). By (2), we have $B^* \subseteq B^{***}$ and $B \subseteq B^{**}$, and so $B^{***} \subseteq B^*$ by (1). Hence $B^* = B^{***}$.

(4). Because $B \subseteq B \cup C$ and $C \subseteq B \cup C$, then $(B \cup C)^* \subseteq B^*$ and $(B \cup C)^* \subseteq C^*$ by (1). Thus $(B \cup C)^* \subseteq B^* \cap C^*$.

On the other hand, $\forall x \in B^* \cap C^*$, we have $x \in B^*$ and $x \in C^*$. It follows that $\forall b \in B, c \in C$, $x \wedge b = O$ and $x \wedge c = O$. Then we have $x \wedge (b \vee c) = O$, and so $x \in (B \cup C)^*$. Hence $B^* \cap C^* \subseteq (B \cup C)^*$.

Corollary 3.8. If A is a non-empty subset of L, then $A^* = \bigcap_{a \in A} (a)^*$.

Proof. By Theorem 3.7 (4) and $A = \bigcup_{a \in A} \{a\}$, we have $A^* = (\bigcup_{a \in A} \{a\})^* = \bigcap_{a \in A} (a)^*$.

Corollary 3.9. Let A and B be non-empty subsets of L. Then $A^* \cap B^* \subseteq (A \cap B)^*$.

Proof. By Theorem 3.7 (4), we have $(A \cup B)^* = A^* \cap B^*$. Because $A \cap B \subseteq A \cup B$, then $(A \cup B)^* \subseteq (A \cap B)^*$ by Theorem 3.7 (1). It follows that $A^* \cap B^* \subseteq (A \cap B)^*$.

Theorem 3.10. Let B be a non-empty subset of L. If $< B >=< B >^{**}$, then $< B >= B^{**}$.

Proof. Because $B \subseteq < B >$, then we have $< B >^* \subseteq B^*$, and so $B^{**} \subseteq < B >^{**}$ by Theorem 3.7 (1). Then by $< B >=< B >^{**}$, we have $B^{**} \subseteq < B >$.

On the other hand, by Theorem 3.7 (2), we have $B \subseteq B^{**}$, and by Theorem 3.5, we know that B^{**} is an ideal of L. Thus $< B > \subseteq B^{**}$ by the meaning of the generated ideal. Hence $< B >= B^{**}$.

Theorem 3.11. Let A and B be non-empty subsets of L, then $< A^* \cup B^* > \subseteq (A \cap B)^*$.

Proof. Because $A \cap B \subseteq A$ and $A \cap B \subseteq B$, then we have $A^* \subseteq (A \cap B)^*$ and $B^* \subseteq (A \cap B)^*$ by Theorem 3.7 (1). Thus $A^* \cup B^* \subseteq (A \cap B)^*$. By Theorem 3.5, we know that $(A \cap B)^*$ is an ideal of L. Hence $< A^* \cup B^* > \subseteq (A \cap B)^*$ by the meaning of the generated ideal.

The relation between an annihilator and an ideal is given in the following.

Theorem 3.12. If B is an ideal of L, then $B \cap B^* = \{O\}$.

Proof. That $O \in B \cap B^*$ is trivial.

On the other hand, $\forall x \in B \cap B^*$, we have $x \in B$ and $x \in B^*$. Then $x = x \wedge x = O$ by the Definition 3.1.

Theorem 3.13. Let B and C be ideals of L. Then $B \cap C = \{O\}$ if and only if $B \subseteq C^*$.

Proof. Suppose that $B \cap C = \{O\}$, then $\forall x \in B$, $c \in C$, $x \wedge c = O$. Or $x \wedge c \neq O \in B \cap C$ contradicts that $B \cap C = \{O\}$. Thus $x \in C^*$ by the Definition 3.1. Hence $B \subseteq C^*$.

Conversely, if $B \subseteq C^*$, then $B \cap C \subseteq C^* \cap C$. By Theorem 3.12, we have $C \cap C^* = \{O\}$. Hence $B \cap C = \{O\}$.

Corollary 3.14. Let B and C be ideals of L. If $C = C^{**}$, then $B \subseteq C$ if and only if $B \cap C^* = \{O\}$.

Proof. Suppose that $B \subseteq C = C^{**}$. We know that C^* is an ideal of L by Theorem 3.5. Hence $B \cap C^* = \{O\}$ by Theorem 3.13.

Conversely, if $B \cap C^* = \{O\}$, then we have $B \subseteq C^{**} = C$ by Theorem 3.13.

The relation between the lattice implication homomorphism image of annihilator and the annihilator of lattice implication homomorphism image is investigated in the following.

Theorem 3.15. Let $(L,\wedge,\vee,\rightarrow,',O,I)$ and $(L_1,\wedge_1,\vee_1,\rightarrow_1,\neg_1,O_1,I_1)$ be lattice implication algebras, B be a non-empty subset of L and $f:L\rightarrow L_1$ be a lattice implication homomorphism. If B^* is an annihilator of B, then $f(B^*)\subseteq f(B)^*$.

Proof. Suppose that B^* is an annihilator of B. Because $\forall y\in f(B^*)$, there exists $x\in B^*\subseteq L$ such that $f(x)=y$, thus $\forall b\in B$, $x\wedge b=O$, and so $f(x\wedge b)=f(x)\wedge_1 f(b)=O_1$. In other words, we have $\forall f(b)\in f(B)$, $f(x)\in L_1$, $f(x)\wedge_1 f(b)=O_1$. Hence $y=f(x)\in f(B)^*$ by the Definition 3.1.

Theorem 3.16. Let $(L,\wedge,\vee,\rightarrow,',O,I)$ and $(L_1,\wedge_1,\vee_1,\rightarrow_1,\neg_1,O_1,I_1)$ be lattice implication algebras, B be a non-empty subset of L and $f:L\rightarrow L_1$ be a lattice implication isomorphism. If B^* is an annihilator of B, then $f(B^*)=f(B)^*$.

Ⅳ. Conclusion

Lattice implication algebra supplies an alternative theoretical basis for lattice-valued logic. As we know, in order to investigate the structure of an algebraic system, the ideals with special properties play an important role. In this paper, we introduce the notion of annihilators, and prove that an annihilator is an ideal and a sl ideal in lattice implication algebra. Then we give some special properties of the annihilator and discuss the relationships between an annihilator and an ideal, between the lattice implication homomorphism image of annihilator and the annihilator of lattice implication homomorphism image in lattice implication algebras. We hope that above work would supply certain theoretical basis for lattice implication algebras and develop corresponding lattice-valued logical system.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of P.R. China (Grant no. 60875034) and Zhejiang province fatal project (priority subjects) key industrial project (2008C11011).

References

[1] Y.B. Jun, Eun Hwan Roh,and Y.Xu, ”LI-ideals in lattice implication algebras”, Bull. Korean Math. Soc., vol. 35(1),pp.13-24,1998.

[2] J. Liu, Y. Xu, ”Filters and structure of lattice implication algebra,” Chinese Science Bull, vol.42 pp. 1517-1520, 1997.

[3] J. Ma, W.Li, D.Ruan, Y.Xu, ”filter-based resolution principle for lattice-valued proprsition logic LP(X)”, Inform. Sci., vol.177 pp. 1046-1062, 2007.

[4] Y.L. Liu, S. Y. Liu, Y. Xu, K. Y. Qin,”ILI-ideals and prime LI-ideals in lattice implication algebras,” Information Sci, vol.155,pp.157-175, 2003.

[5] Y.B. Jun, Y. Xu, J.Ma, ”Redefined fuzzy implative fliters,” Information Sci, vol.177,pp.1422-1429, 2007.

[6] X.F.Wang,”Study on lattice-valued logic based on lattice implication algebra and its model theory,” Southwest Jiaotong Univ, 2004.

[7] Y. Xu, ”Lattice implication algebras,” Journal of Southwest Jiaotong Univ, vol.28 (1),pp.20-27,1993.

[8] Y. Xu, K. Qin, J. Liu, Z.M. Song,”L-valued propositional logic,” Information Sci, vol.114,pp.205-235,1999.

[9] Y. Xu, J. Liu, Z.M. Song, K.Y.Qin, ”On semantics of L-alued first-order logic Lvpl,” Internation Journal of General Systems, vol.29(1),pp.53-79 ,2000.

[10] Y. Xu, D. Ruan, K.Y. Qin, J. Liu, ”Lattice-Valued Logic,” Berlin: Springer-Verlag, 2003.

[11] X. Pan, Y.Xu,”Lattice implication ordered semigroups,” Information Sci, vol.178,pp.403-413, 2008.

[12] H.Zhu, J.B.Zhao, Y.Xu ”The primary ideal of residuated lattice implication algebra,” Journal of Zhengzhou Univ, vol.38(2),pp.25-28,2006.

[13] H.Zhu, J.B.Zhao, Y.Xu,”The n-fold Prime Filter of Residuated Lattice implication algebras,” Journal of Zhengzhou Univ, vol.40(1),pp.19-22, 2008.

[14] H.Zhu, S.W. Chen,et al,”Multi-fold Fuzzy Positive Implicative Filter of Resduated Lattice Implication Algebras,” Jouenal of Zhengzhou Univ, vol.41(2),pp.19-23,2009.

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 210-213

Combining Permission Based Delegation Model with Chinese Wall

Chunxiao Ye[1], Fawen Yu[1], Tinsen Huang[1], Haibo Hu[2], and Hong Xiang[2]
[1]College of Computer Science, Chongqing University, Chongqing, China
Email: yecx@cqu.edu.cn
[2]School of Software Engineering, Chongqing University, Chongqing, China
Email: {oram, xianghong}@cqu.edu.cn

***Abstract*—Chinese Wall Security Policy (CWSP) is a widely applied access control policy in many fields, especially in commercial world. Delegation is one of the hot topics of access control technologies. Delegation with CWSP means delegation must satisfy not only delegation constrains but CWSP as well. There exist many delegation models, such as RBDM, RDM2000 and PBDM et al, but few focus on it. This paper proposed an approach of how to delegate permission with the restriction of CWSP. Although CWSP is part of delegation constraint, it does not mean that existing delegation models can be easy applied to this kind of delegation. In our approach, we first define two types of delegation constraints consisting of CWSP. Then we discussed different types of revocation and found that automatic revocation can make delegation safer than user revocation. Also, we found that there exists security vulnerability in multi-step delegation and gave some feasible solutions. Finally, this paper gave system implementation architecture and some examples to show how our approach works properly in a situation with CWSP.**

Index Terms*—Permission Based Delegation Model, Delegation, Chinese Wall Security Policy, CWS*P

I. INTRODUCTION

Delegation means that a user (delegator) can give her permissions to other person (delegatee). There are three types of situations in which delegation can take place: backup of roles, decentralization of authority and collaboration of work. Human-to human delegation has received considerable attention recently [1] ~ [5].

RBDM [1] is the first delegation model based on role. RBDM deals with flat and hierarchical role, multi-step delegation. RDM2000 [2] is an extension of RBDM. A rule-based declarative language has been proposed to specify and enforce policies in delegation. PBDM [3] is a flexible delegation model that supports multi-step delegation and revocation in role and permission level. RPRDM [4] only addresses repeated and partial delegation. ABDM [5] is a strict and more secured delegation model both in temporary and permanent delegation.

Chinese Wall Security Policy model (CWSP model) [6] was first defined by Brewer and Nash in 1989. It has been widely applied in commercial world where there exists conflict of interest among different companies. CWSP can be seen as a mixture of DAC and MAC and is seem as a history-based access control policy.

In a commercial environment, delegation occurs frequently, especially when someone is on a business trip. That means a user (delegator) can transfer her permissions to other person (delegatee) and delegatee can perform these transferred permissions just like the delegator herself. Of course, delegation is constrained by system security policy. Above researches have focus on delegation constraints, but few have elaborated delegation with CWSP.

II. MOTIVATION

Let us consider the example shown in figure 1: Alex can access to data of Bank A and Oil Company B, and Mike can access to data of Bank B and Oil Company A. Because Bank A and B, Oil Company A and B are in different Conflict of interest classes (COI) respectively, according to CWSP, there is a "wall" between Alex and Mike.

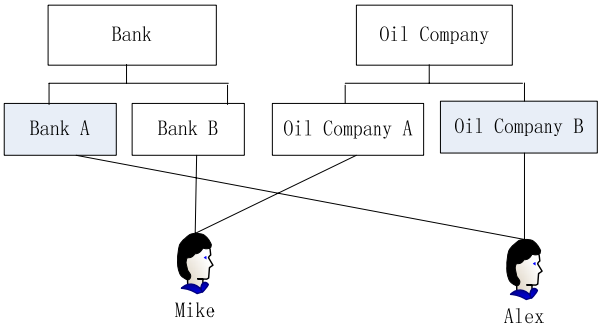


Figure 1. User access to different datasets

Figure 2 illustrates examples of delegation among different persons. $P_{Bank\ A}$ and $P_{Oil\ Company\ B\ stand}$ for access right to Bank A's and Oil Company B's datasets respectively. According to CWSP, Alex cannot choose Mike and delegate permissions to him. So, Alex can choose another person, say Tom, and perform delegation. Assume Alex delegates $P_{Bank\ A}$ to Tom, as shown in figure 2 (a). If Tom has accessed to those datasets that Mike is able to access to, the delegation will be prohibited; otherwise the delegation will succeed. In this case, because Tom has no permissions before delegation, the delegation succeeds.

Let us consider another situation: assume Alex and Mike delegate some of their permissions to Tom separately, as shown in figure 2(b). According to CWSP,

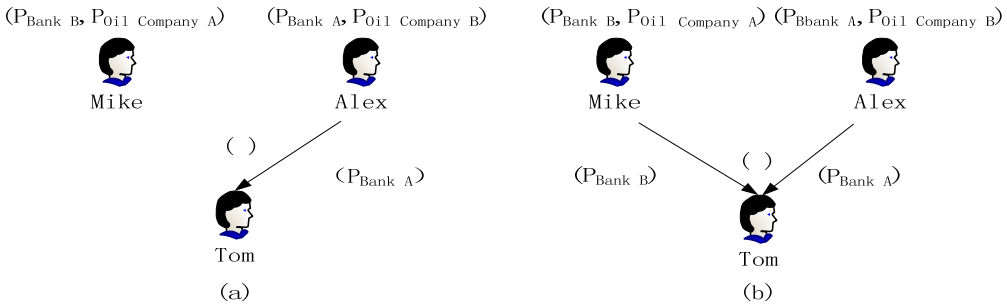


Figure 2. Example of single-step delegation

Tom can accept permissions within one COI, say Bank A and B, and once he choose to access to one bank's dataset and a "wall" will prevent him from the other's. So, delegator does not know whether the delegation is success after the delegation finished, for it depends on delegatee's selection of permissions. For example, if Tom chooses to access to bank A's data, then Alex's delegation is success and Mike's failed. This is a kind of "uncertain delegation" but is actually allowed according to CWSP. Thus we have two choices: to make delegation easy, we can prohibit it, or to make delegation flexible, we can allow it. The former is easy to implement and less safe and the latter is safer and makes the implementation somewhat difficult. In the former case, CWSP can be expressed by delegation constrains, such as prerequisite condition (CR) and in the later by activation constrains. If "uncertain delegation" is allowed, for the purpose of not violating CWSP, once a delegatee activates some permissions, the delegatee is prohibited from activating the corresponding conflict permissions for ever or those permissions will be revoked immediately.

III. DELEGATION WITH CWSP

CWSP can be regarded as a kind of SOD and looks like dynamic SOD [7]. We also believe CWSP is an object-based separation of duty as mentioned in [7], but the difference is that we use role-based separation of duty to describe CWSP.

A. Concepts

Here we give some concepts definitions:

Definition 1

- T, U, AR, RR, TDR, P, O, OBJ and R are set of time, user, administrative role, regular role (created by system administrator), temporary delegation role (created by delegator), permission, operation, object and role respectively. t_0, $t_c \in T$. t_0 is the initial time point and t_c is current time point of system.
- $P \subseteq O \times OBJ$, $R = RR \cup TDR$
- $PRA \subseteq R \times P$
- $URA \subseteq U \times RR$
- $UDA \subseteq U \times TDR$
- $own(u): U \rightarrow 2^{TDR}$ and $\neg\exists$ (u_1, $u_2 \in U$, $tdr \in TDR$) $\wedge$ ($u_1 \neq u_2$) $\wedge$ ($tdr \in own(u_1) \wedge tdr \in own(u_2)$): a function mapping a user to a set of delegation roles which she created.
- $UA = URA \cup UDA$
- UA_{ti} is the UA at ti time point.
- $UA = U_{tc}$.
- $UAH = UA_{t0} \cup \ldots \cup UA_{tc}$: a relation of user-role assignment history.

Other concepts' definitions can be found in [3] and [5].

Definition 2 for two permissions $p_1 = (o_1, obj_1)$ and $p_2 = (o_2, obj_2)$, if obj_1 and obj_2 belong to the same COI, we call p_1 and p_2 are conflict with each other, denoted as $p_1 \leftrightarrow p_2$.

Definition 3 we call r and r' are conflict with each other, if ($\exists$ p, p'$\in$P)$\wedge$ (r, p) $\in$PRA $\wedge$ (r', p') $\in$PRA $\wedge$ p $\leftrightarrow$ p', denoted as r $\leftrightarrow$ r'. A role r's conflict roles can be denoted as $\Re r$, where $\forall r_i \in \Re_r$, r $\leftrightarrow$ r_i.

Definition 4

$$\text{hasAct}(u,r) = \begin{cases} \text{true } (u,r) \in UAH, u \text{ has activated/ is activating} r, \\ \text{false otherwise.} \end{cases}$$

Definition 4 defines the function to judge whether a role has been activated or being activated.

B. Delegation

Delegation constraint is vital for delegation security. In our work, delegation constraint consists of three parts: the first one is prerequisite role (CR), the second is SOD and the third is CWSP. Although CWSP looks like SOD, we separate it from SOD for there are two types of CWSPs: weak and strong CWSP:

- Weak CWSP can be seen as a normal delegation constraint. This means a user can keep two conflict roles at the same time but she can only activate one role and lost the right of activate the other one for ever. So, if a delegatee has roles that conflict with delegation role but she has not activated or is activating these roles, delegation will be allowed.
- Strong CWSP is stricter than weak CWSP. This means if delegatee already has roles that conflict with delegation role, delegation will not be allowed even those roles has not been activated before. Strong CWSP beyond the basic security restrictions of CWSP and is useful in a situation of multi-step delegation.

Accordingly, to obtain different delegation security level, delegation can be defined as two different types:

delegation with weak CWSP and strong CWSP. Delegation relations are defined as follows:

Definition 5 a delegation with weak CWSP can be denoted as relation can-delegateW⊆U×R×U, where U and R are user and role set respectively. a delegation that a user u delegates r to u' succeed means (u, r, u') ∈can-delegateW⇔ ((u, r) ∈UA) ∧ (∀r_i∈$\Re_r$, (u', r_i) ∈UAH⇒hasAct (u', r_i) =false)).

A successful delegation means it satisfies all delegation constraints, such as CR, SOD and CWSP.

Definition 6 a delegation with strong CWSP can be denoted as relation can-delegateS⊆U×R×U. a delegation that a user u delegates r to u' succeed means (u, r, u') ∈can-delegateS⇔ ((u, r) ∈UA) ∧ (∀r_i∈$\Re_r$, (u', r_i)∉UAH).

C. Revocation

We propose two types of revocation: user and automatic revocation. User revocation means revocation only can be performed by delegator or system administrator manually, while automatic revocation means revocation is triggered and performed by system automatically. User revocation can be found in most delegation model, but few support automatic revocation.

Definition 7

- Revocation performed by delegator is denoted as relation can-revokeD ⊆ U×R×U. (u, r, u') ∈can-revokeD ⇔ (u, r, u') ∈can-delegateS or (u, r, u') ∈can-delegateW.
- Revocation performed by system administrator is denoted as relation can-revokeS ⊆ U×AR×U×R×U. (u, ar, u', r, u'') ∈can-revokeS ⇔ (u, ar) ∈URA ∧ ((u', r, u'') ∈can-delegateS or (u', r, u'') ∈can-delegateW).

For example, Alex has delegated rBank A to Annie and he can revoke it from Annie whenever he wants to. A system administrator can also revoke the role from Annie

We define automatic revocation as:

Definition 8 revocation triggered and performed by system automatically can be denoted as relation can-auto-revoke⊆U×R×U, (u, r, u') ∈can-auto-revoke ⇔ (u, r, u') ∈can-delegateW ∧ (∃r_i∈$\Re_r$, (u', r_i) ∈UAH∧hasAct (u', r_i) =true).

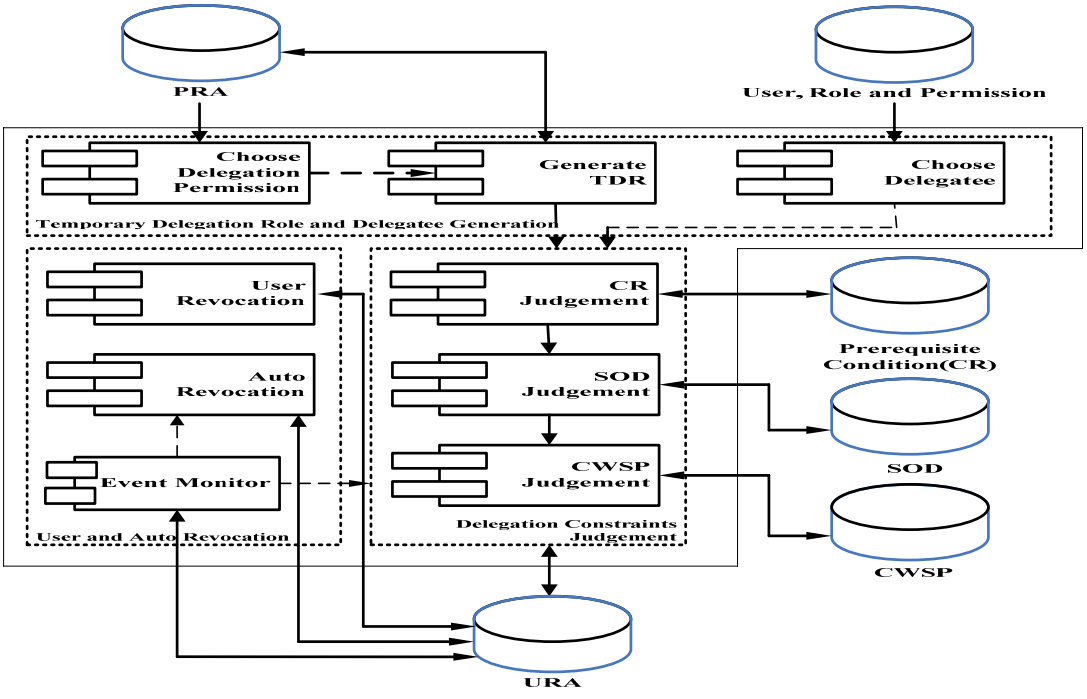
As we can see in definition 8, automatic revocation can only be triggered and performed in the situation of delegation with weak CWSP.

IV. Implementation

A. System Architecture

Figure 3 shows the main components of delegation and revocation with CWSP. The solid arrow lines denote data flow between components and databases, while dashed arrow lines denote data flow among components.

Each component's function is listed in table 1. We can divide these components into four different groups: temporary delegation role and delegatee generation, delegation constraints judgment, user and auto revocation and access control data and results (consists of PRA, User, Role and Permission, Prerequisite Condition (CR), SOD, CWSP and URA). The first group generates temporary delegation role and delegatee, the second group performs delegation by judging CR, SOD and CWSP, the third group deals with user and auto



Fingure 3. Delegation and Revocation Components

revocation and the fourth group stores basic access control data and delegation and revocation results.

B. Delegation and Revocation Steps

We give some sequence diagrams to show how delegation and revocation works, including single and multi-step delegation, user and automatic revocation. Involved components and their sequence in those processes are also given.

1. Single-step delegation

1） Delegator chooses which permissions or roles she will delegates to other person (CDP);

2） Delegator creates a temporary delegation role (GTDR);

3） Delegator chooses a user and delegates a tdr to her (CD);

4） tdr and user's information are forward to Delegation Constraints Judgement (DCJ) ;

5） DCJ judges whether delegation satisfies CR, SOD and CWSP simultaneously. If not, delegation failed; otherwise DCJ writes delegation results into URA.

2. User revocation

1） Delegator decides to revoke tdr from delegatee, she calls UR;

2） UR determines whether revocation succeeds. If it is successful, UR saves results into URA; otherwise it returns revocation failed message to delegator.

3. Automatic revocation

3） EM detects an event that one of a delegatee's delegation roles which are conflict with each other is now activating;

4） EM forwards this message including delegatee, conflict roles, et. al. to AR;

5） AR determines whether revocation succeeds and saves results into URA or returns revocation failed message to delegator accordingly.

V. CONCLUSIONS AND FUTURE WORKS

In this paper, we firstly give an introduction and overview of Chinese Wall Security Policy and delegation model and find that few works are focus on delegation with CWSP. Then, we show the motivation of our work and find that delegation with CWSP is different from regular delegation. Current delegation models cannot deal with these issues efficiently. So, we propose some mechanisms to meet these delegation requirements, such as single and multi-step delegation, user and automatic revocation and activation rule. Thus ensure security of delegation with CWSP. Finally, we give the system implementation architecture and some examples to show how our mechanisms work. In this paper, permissions of a temporary delegation role are only consisting of the whole or part of a role's permissions. It is difficult to delegates permissions coming from different roles, especially in multi-step delegation, for weak and strong CWSP will be more complicated. In future work we would like to deal with this issue.

ACKNOWLEDGMENT

This research is supported by National Natural Science Foundation of China (Project No. 60803027), the Natural Science Foundation of Chongqing, China (Project No.CSTC, 2008BB2320) and the National High Technology Research and Development Program of China (Project No. 2007AA01Z445)

We acknowledge the generous help of Dr. Jason Crampton, who working in the Information Security Group (ISG), Royal Holloway, University of London.

REFERENCES

[1] Ezedin S. Barka, "Framework for Role-Based Delegation Models" [PhD Dissertation], George Mason University, Fairfax, Virginia, summer 2002.

[2] Longhua Zhang, Gail-Joon Ahn, and Bei-Tseng Chu, "A rule-based framework for role-based delegation", *ACM Transactions on Information and System Security (TISSEC)*,2003,6(3), pp.404-441.

[3] Xinwen Zhang, Sejong Oh, and Ravi Sandhu, "PBDM: A Flexible Delegation Model in RBAC", In *Proc. of the SACMAT'03*, Como, Italy: ACM press, 2003,pp. 149 – 157.

[4] ZHAO Qing-Song, SUN Yu-Fang, and SUN Bo, "RPRDM: A Repeated-and-Part-Role-Based Delegation Model", *Journal of Computer Research and Development*, 2003, 40(2), pp.221-227.

[5] Ye Chun-Xiao, Wu Zhong-Fu, Fu Yun-Qing, et al, "An Attribute-Based Extended Delegation Model", *Journal of Computer Research and Development*,2006,43(6), pp.1050-1057.

[6] Brewer David D.C. and Michael J. Nash, "The chinese wall security policy", In *Proc. of the IEEE Symposium on Security and Privacy*, Oakland, IEEE Press,1989, pp. 215-228.

[7] J. Crampton, "Specifying and enforcing constraints in role-based access control," in *Proc. of SACMAT'03*, Como, Italy: ACM press, 2003, pp. 43–50

Chunxiao Ye was born in Chongqing, China, on April 12th, 1973. He received his BEng in computer software from Sichuan University, China in 1995 and MEng in computer architecture from Chongqing University, China in 2002. In 2005, he received his PhD in computer science from Chongqing University, China.

He is now an associate professor at the College of Computer Science, Chongqing University, China. He has published more than 50 journal papers and conference articles in access control, software engineering and database. His research interests include access control, grid, software engineering and database.

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 214-217

Research on Security Policy and Framework

Dongliang Jiao[1,2], Lianzhong Liu[1], Shilong Ma[2], and Xiaoni Wang[1]
[1]Key Laboratory of Beijing Network Technology, Beihang University, Beijing, China
Email: {jiao_dl, lianzhong-liu, xiaoni_wang}@163.com
[2]State Key Laboratory of Software Development Environment, Beihang University, Beijing, China
Email: slma@nlsde.buaa.edu.cn

***Abstract*—Policies are rules that govern the choices in behaviors of a system. Security policies define what actions are permitted or not permitted, for what or for whom, and under what conditions. In this paper, the present situation of research on policy is overviewed, including policy and security policy definition, policy language, the conflict detection of policy and policy framework. Finally, the summary and the future trends of policy and security policy study are given.**

***Index Terms*—Policy, Security policy, Policy Framework, Interval Temporal Logic (ITL), Ponder Language**

Ⅰ. INTRODUCTION

As systems become increasingly complex, the development of methodologies for their security and effective management becomes more and more critical. One important aspect of this general systems management problem is that of being able to raise the level of abstraction [1]. This requirement is commonly described as the ability to deal with systems in terms of policies rather than explicit controls.

Policy means different things to different people. For our purposes, the term "policy" is defined as high-level rules for implementing the goals, beliefs, and objectives of an enterprise. There are three types of policies, and you will use each type at different times in your information security program and throughout the organization to support the business process or mission. The three types of policies include [2]:

1) Global policies. These are used to create the organization's overall vision and direction.

2) Topic-specific policies. These address particular subjects of concern. We discuss the information security architecture and each category.

3) Application-specific policies. These focus on decisions taken by management to control particular applications (financial reporting, payroll, etc.) or systems (budgeting system).

Ⅱ. SECURITY POLICY

A security policy establishes what must be done to protect information stored on computers. A well-written policy contains sufficient definition of "what" to do so that the "how" can be identified and measured or evaluated [3].

Supported by Co-Funding Project of Beijing Municipal Education Commission under Grant No. JD100060630

AP-PROC-CS-10CN006

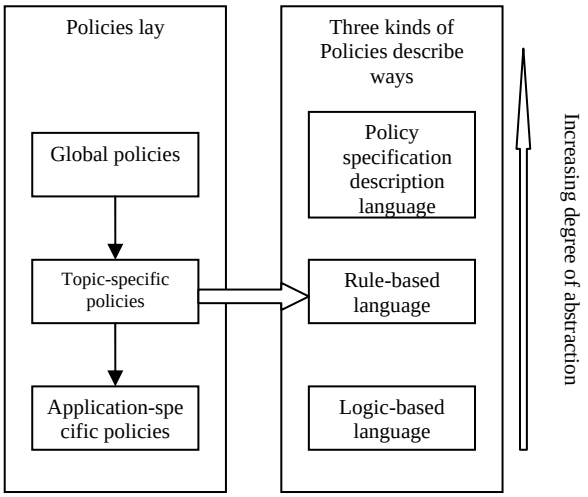


Figure 1. Three types of policies

The goal of a security policy is to maintain the integrity, confidentiality, and availability of information resources [4].Confidentiality requires that the information or resources in a computer system only be disclosed to authorized parties. Integrity includes data integrity (the content of the information) and origin integrity (the source of the data, often called authentication). Availability is concerned with the ability to use the information or resource desired. The aspect of availability that is relevant to security is that someone may deliberately arrange to deny access to data or to a service by making it unavailable to legitimate parties.

To enforce the above requirements, authentication, access control and auditing, three mutually supportive technologies are used [5].

1) Authentication: deals with identification of users.

2) Access control: concerned with limiting the activity of users who have been successfully authenticated by ensuring that every access to information or resources is controlled and that only authorized accesses can take place.

3) Auditing: is the process of recording information about accesses to resources so as to be able to establish responsibilities in the event of a security breach.

Ⅲ. SECURITY POLICY REPRESENTATION

A. Policy Language

One of the main factors constrains the policy applications is that policy description language is not uniform. Each language can only support one or several models. At present, the policy language of the research

focused mainly on rule-based forms of expression. The following are examples of some well-known policy language:

1) PDL (Policy Description Language) is a Bell Labs-developed event-based language [6]. They use "event - condition - behavior" rules of the form to define a policy. PDL's syntax is simple, with clear semantics and can express a strong sexual, However, PDL does not support Access Control Policy, and does not support the roles of complex strategies.

2) Ponder is a declarative, object-oriented language for specifying security policies with role-based access control, as well as general-purpose management policies for specifying what actions are carried out when specific events occur within the system or what resources to allocate under specific conditions [7]. Ponder have four basic policy types: authorizations, obligations, refrains and delegations and three composite policy types: roles, relationships and management structures that are used to compose policies. The dependencies between the various types are shown in figure 2. Ponder also has a number of supporting abstractions that are used to define policies: domains for hierarchically grouping managed objects, events for triggering obligation policies, and constraints for controlling the enforcement of policies at runtime.

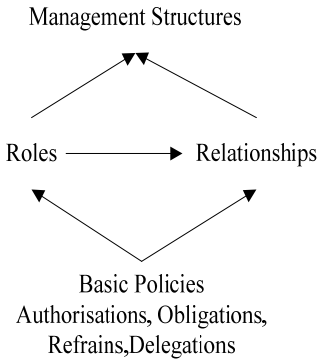
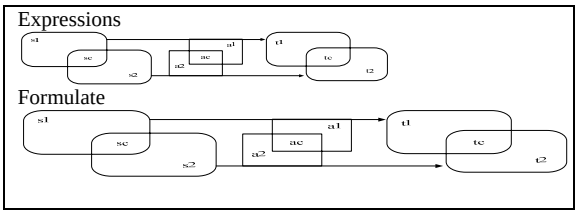


Figure 2. Policy Types

3) *Interval Temporal Logic (ITL)* is a linear-time temporal logic with a discrete model of time. A system is modelled by a set of relevant state variables An interval is considered to be a (in)finite, nonempty sequence of state $\partial_0, \partial_1, \ldots\ldots$, where a state ∂_i is mapping from a set of variables to a set of values [5]. The first work investigates logic-based languages for the specification of authorization is Woo and Lam [8]. Jajodia et al. worked on a proposal that provides specific rules for resolving conflicts among positive authorizations and negative authorizations. Their policy language, called Authorization Specification Language (ASL), provides a specific predicate, cando, for expressing explicit authorizations and a different predicate, dercando, for expressing derived authorizations [9]. Bertino et al. propose a model supporting the specification of periodic authorizations, that is, authorizations that hold in specific periodic intervals [10]. The syntax of ITL is defined in Table 1 where:

TABLE Ⅰ. SYNTAX OF ITL



Z is an integer value,

a is a static integer variable (doesn't change within an interval),

A is a state integer variable (can change within an interval),

v a static or state integer variable,

g is a integer function symbol,

4) XACML is used OASIS (Organization for the Advancement of Structured Information Standards) to represent the policy. The current syntax for policy in XACL is oriented around the 3-tuple {subject, object, action} [11]. The subject primitive allows user IDs, groups, and/or role names. The object primitive allows granularity as fine as a single element within an XML document. The action primitive consists of four kinds of actions: read; write; create; and delete. The main advantage of using XML is: XML is a widely used standard.

B. Policy language Evaluation Criteria

Each policy language has its own advantages and disadvantages. Formal Logic language is not enough intuitive, not easy to map to the realization of the mechanism, so is difficult to use and understand; Ponder language is more in line with common standards, could applies to a variety of applications, but Ponder Toolkit does not open source; XACML is ideal for on-line transmission and browse, but have a lot of redundant information. So when we use policy language, we should compare their advantages and disadvantages.

Through analysis and research of the existing policy language, it is obvious that a good policy language should meet the following requirements:

1) Be easy to understand, The policy written in the policy language should be easy to understand, which is very important. Only when the policy could easily understood can achieve a wide range of applications.

2) Avoid absolutes, It is not easy to map abstract representations into the implementation mechanism, thus too abstract representations should be avoided.

3) Be do-able, It is easy to map to a description of the underlying measurement specifications, it is compiled into some kind of low-level (e.g. C) language code.

4) Good scalability, According to demand, it can mix new policy types with the existing policy , namely, support for complex strategies, and supports reusability。

5) Policy classification high covering， The language can cover all the current international standards of information security policy.

6) Opining source, the language definition itself is an open source, which can be modified to cater for the new changes.

Ⅳ. SOLUTION TO THE CONFLICT

In the policy-based applications, it is difficult to avoid conflicts between policies. So the tools to provide detection of conflict and conflict resolution approach are necessary. We discuses two kinds of conflict here: modality conflicts and application specific conflicts [12].

A. modality conflicts

The modality conflict is caused by inconsistencies in policy specification. When two or more policies use the same subject, behavior or goals that define the form of the opposite, then such conflicts occur. As shown in figure 3. it can be achieved by parsing policies to predict [13]. Because space is limited, we do not make a specific description.

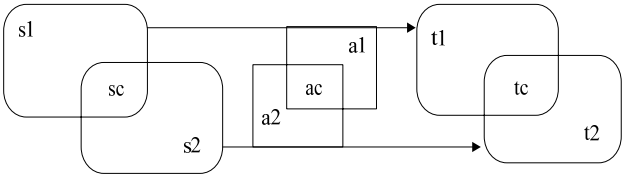


Figure 3. Form conflict

B. application specific conflicts

Application specific conflict refers to the conflict with the consistency of the elements, these elements that is related to the subject, objectives and standards of behavior and external. Such conflicts can not predict directly through the definition of policies, need to specify additional information that lead to conflict. The additional information is called meta-policy. There are two kinds of proposal of static and dynamic to resolve the semantic conflict by Application of meta-policy. Static proposal refers to the policies of conflict can not exist together in a system. Dynamic proposal is the policies of conflict can not be run at the same time to activate.

Ⅴ. POLICY FRAMEWORK

A. Introduction

SHAN [14] divided security policy framework into three categories in his dissertation: based on policy description language; based on security attributes and based on the unified model. The following are some well-known policy Framework:

1) FAM Framework. This is based on policy description language, is proposed by Sushil Jajodia, Pierangela Samarati et al. in 1997. They defined a flexible authorization manager, could to implement multiple security policies in the same system. But it is the biggest drawback is that can not adapt to dynamic changes. An overview of FAM framework is shown in figure 4.

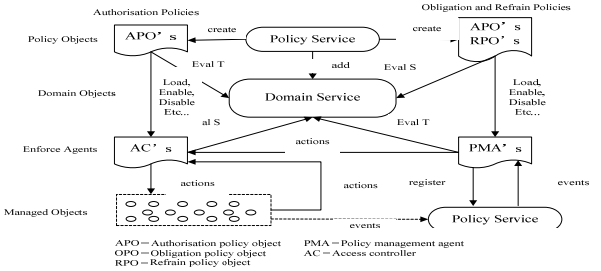


Figure 4. FAM Framework

2) Ponder policy deployment model. The framework is an extension of IETF policy management framework. It includes 3 supporting services: a domain service, a policy service, and an event service. The Policy Service acts as the interface to policy management, it stores compiled policy classes, creates and distributes new policy objects. The Domain Service manages a distributed hierarchy of domain objects and supports the efficient evaluation of subject and target sets at run-time. Each domain object holds references to its managed objects but also references to the policy objects that currently apply to the domain. The Event Service collects and composes events from the underlying systems and from the managed objects in the system, and forwards them to registered policy management agents triggering obligation policies. An overview of the policy deployment model is shown in figure 5.

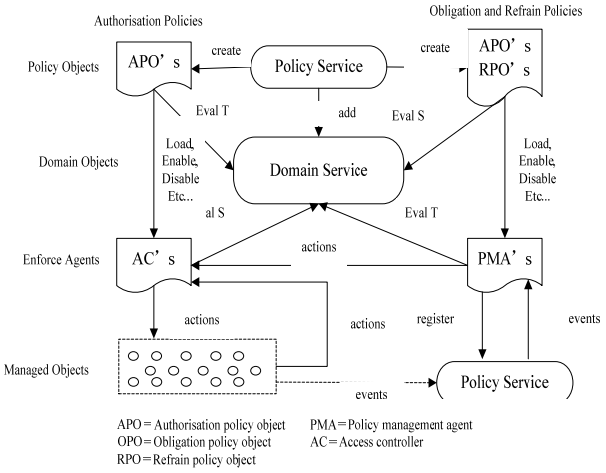


Figure 5. Overview of policy deployment model

3) FLASK Framework. The Flask security architecture, as shown in Figure 6, describes the interactions between subsystems that enforce security policy decisions and a subsystem which makes those decisions, and the requirements on the components within each subsystem. The primary goal of the architecture is to provide for flexibility in the security policy by ensuring that these subsystems always have a consistent view of policy decisions regardless of how those decisions are made or how they may change over time. Secondary goals for the architecture include application transparency, defense-in-depth, ease of assurance, and minimal performance impact.

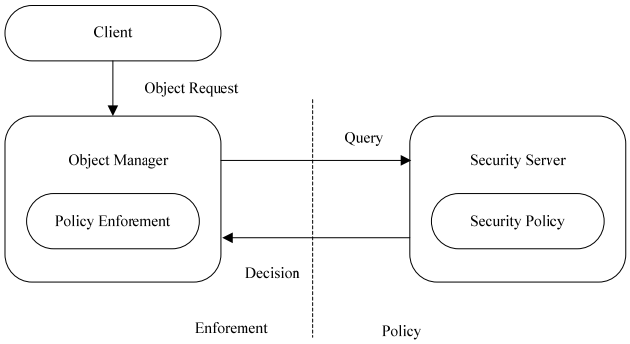


Figure 6. The Flask architecture

B. Evaluation criteria

Because of the diversity of policy, policy framework should to support the characteristics of different policies, the more features supported, the more policies to support. Michael Carney and Brian Loe have proposed five evaluation criteria of policy switching mechanism, we quote as follow:

1) Policy Flexibility, because the current policy description language is not uniform, a good policy framework should be able to support multiple policy language, so as not to limit its application in large-scale systems.

2) Function Flexibility, Policy framework itself should be configurable, and when it need other features, it can be dynamically added.

3) Safety, The security of a policy framework is that policies can be correctly stored and implementation, and will not be tampered with.

4) Reliability, Policy framework should take into account the different policy operating conditions change, the framework itself should be able to adapt to this change, rather than stop running, waiting for administrators to manually manage.

5) Efficiency, Especially for smaller granularity of time, must focus on the efficiency of policy enforcement, if policy is not configured or implemented, the time has passed.

C. Comparison of Policy Framework

We have the framework of the above-mentioned three according to the evaluation criteria for comparison. We compare mainly from the following three aspects: support for policy flexibility, function flexibility, efficiency, and reliability.

Ponder policy deployment model is the best model on policy flexibility, but unable to adapt to environmental change; FAM Framework can not support function flexibility and unable to adapt to environmental change, because the access policy is execute by hard-coded language, but its efficiency and policy flexibility is good; Flask Framework is very good at support function flexibility and is better at efficiency, but can't support DAC and unable to adapt to environmental change.

VI. CONCLUSION

Policy-based management is one of the latest developments, and achieved some success. From the practical application point of view, however, this is not enough. Problems at this stage included as the following aspects:

1) Lack of standards

2) Lack of good policy management framework

3)Lack of policy conflict detection and policy elimination of technical

4) Lack of a unified Policy description Language

Overall, currently the study of policy should focus on the described in the policy specification, policy framework development, as well as some key technologies. Only in this way, the policy could be applied more widely.

REFERENCES

[1] M.J. Maullo, S.B. Calo "Policy management: an architecture and approach," Systems Management, vol., no., pp.13-26, Apr 1993.

[2] S. S. Greene, "Security Policies and Procedures," PEARSON Education, 2008.

[3] L. Cholvy, F. Cuppens, "Analyzing consistency of security policies," Security and Privacy, IEEE Symposium on, p. 0103, 1997 IEEE Symposium on Security and Privacy, 1997 .

[4] M.Bishop, "Computer Security," Addison-Wesley, 2002.

[5] Francois Siewe, "A Compositional Framework for the Development of Secure Access Control Systems" 2005 Ph.D.

[6] L. Jorge, B. Randeep and N. Shamim, "A policy Description Language," Orlando, Florida: AAAI, 1999.

[7] N. Dulay, E. Lupu, M. Sloman, and N. Damianou, "A Policy Deployment Model for the Ponder Language," Integrated Network Management Proceedings, IEEE/IFIP International Symposium on , vol., no., pp.529-543, 2001.

[8] M. Winslett, C. Zhang, and P. A. Bonatti,"A logic for distributed authorization," The ACM Conf. on Comput and Communications Security. New York: ACM Press, 1993, PP. 124-133.

[9] S. Jajodia, Pierangela. Samarati, V. S. Subrahmanian and E. Bertino "A unified frame-work for enforcing multiple access control policies," ACM transaction on Database Systems. 2001, PP. 474 - 485 .

[10] S. D. Stoller, P. Yang "Efficient policy analysis for administrative role based access control," The ACM Conf. on Computer and Communications Security. New York: ACM Press, 2007, PP. 445-455.

[11] T. MOSES "eXtensible Access Control Markup Language (XACML) Version 2.0,". OASIS Standard, 2005.

[12] E. C. Lupu and M. Sloman, "Conflicts in Policy-based Distributed Systems Management," Software Engineering, IEEE Transactions on , vol.25, no.6, pp.852-869, Nov/Dec 1999.

[13] Lu Shifeng, Liu Xuemin, Liu Taoying and Wang Qin,"Overview on Policy-Based Management". Computer Engineering and Applications, vol 9, pp.85-89. 2004(09).

[14] SHAN Zhiyong, "Research on the Framework for Multi-Policies and Practice in Secure Operation System," Institute of Software Chinese Academy of Sciences, 2002.

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp.218-222

An Adaptive Scheduling for QoS Enhancement Mechanism in IEEE 802.11 WLAN

Aiyun Zhan, Aihan Yin, Yueli Jiao, and Qingmiao Zhang
School of Information Engineering, East China Jiaotong University
Nanchang, China 330013
E-mail: yinaihan@126.com

***Abstract*-For the problem of QoS in the IEEE802.11DCF, proposed an enhanced adaptive scheduling distributed EASDCF mechanism. Through the largest contention window and retry times, this mechanism provides differentiated services and guarantees the real-time service priority access to the channel with higher probability. Besides, it provides a reliable guarantee for the packet drop rate sensitive service. Simulation results indicate that the classification method adopted by the EASDCF has a good effect, enhancing the overall throughput performance of the WLAN and making reasonable use of system resources.**

***Index Terms*-largest contention window; QoS; distributed scheduling mechanism; throughput; packet drop rate**

I. INTRODUCTION

Wireless communication technology has brought us more and more convenient communication services. However, with the development of technology, the system performances could not meet the demand of requirements for the increasing services from us. As is well-known, the wireless channel is a limited resource, it is a problem needed to be resolved in MAC (Medium Access Control) layer that how to share and distribute the channel fairly, effectively and reliably. CSMA/CA (Carrier Sense Multiple Access with Collision Back-off) mechanism, as the elementary access way in IEEE802.11 protocol [1], adapts to the change of station numbers and the communication traffic. Due to CSMA/CA lack of effective QoS (Quality of Service) mechanism, makes it hard to support the services with high requirements of delay performance [2]. A basic application of CSMA/CA is the IEEE802.11DCF (Distributed Coordination Function) mode. Based on DCF model, the paper analyzes the traditional back-off mechanisms of 802.11 protocol and compares EDCF (Enhanced DCF) with frequently used DCF, then put forward a performance-enhanced mode---EASDCF (Enhanced Adaptive SDCF). In this way, we execute the network simulations under NS-2, whose results have certain guiding significance in the transmission of real time services, like voice in the wireless local area network.

II. DCF AND EDCF ACCESS MODE

DCF is the basic media access control mechanism of 802.11, it adopts CSMA/CA and binary exponential back-off mechanism. DCF access mode is shown in figure 1. In the IEEE802.11 model, each station accesses channels through competition [3]. The initial size of contention window is CWmin. When the transmission fails, it doubles its size and chooses another random back-off time in the new window. Otherwise, the contention window returns to minimum size CWmin directly. Checking again whether there are still data to be transmitted in the station, if so, after a period of DIFS (DCF Interframe Space), reenter the back-off waiting state again until it gets the authorization for media access [4]. The way which DCF adopted to allocate channels is simple and effective. Particularly when collisions increase, it can enlarge CW (Contention Window) promptly. The larger the contention window is, the more powerful the system is to solve collisions.

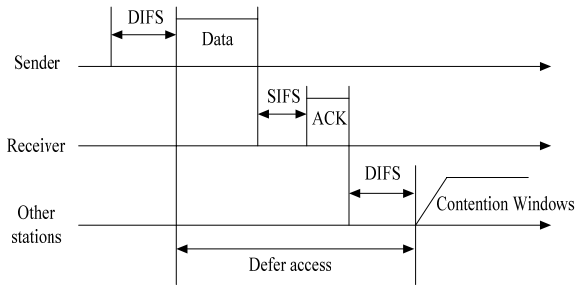


Figure 1 Basic DCF access model

EDCF model is the improvement of MAC protocol proposed by IEEE802.11e work group, and the program introduced the Enhanced Distributed Coordination Function [5]. The support of QoS in the EDCF is realized by the introduction of TC (Traffic Categories).The priority of channel access is differentiated by the QoS parameters of TC. After detecting the channel idle time reaches an AIFS (Arbitration InterFrame Space), the mobile station starts back-off. The relationship between various IFS which defined by EDCF is shown in figure 2. AIFS is at least a DIFS time. The back-off counter range is [0, CW-1].

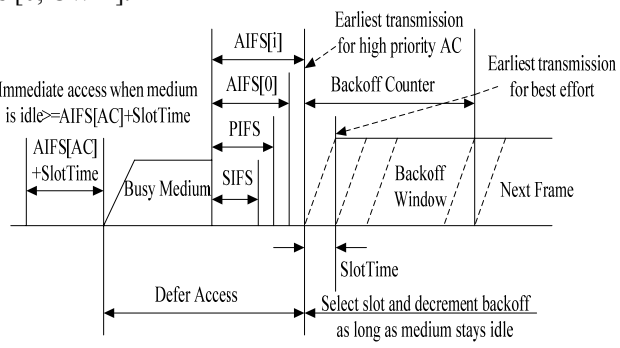


Figure 2 EDCF access model

EDCF in comparison with DCF, has some important differences, when detecting that the channel is busy in AIFS period, the back-off counter reduce 1 at the last time slot of AIFS. However in the DCF, the back-off counter reduces 1 at the first time slot after DIFS. When the transmission fails, EDCF makes use of the PF (Persistent Factor) to control the selection of CW, CWnew [TC]=PF×CWold [TC]. However in the DCF, after the failing transmission, the CW always doubles the size, that is PF = 2.

In summary, EDCF provides different access priorities by modifying three control parameters of DCF competitive mechanism: 1) InterFrame Space; 2) Minimum contention window; 3) Persistent Factor.

III. PERFORMANCE ENHANCEMENT SCHEME EASDCF

A. *Improved back-off mechanism*

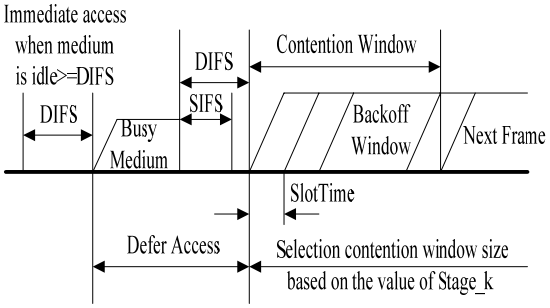


Figure 3Improved back-off mechanism

The improved back-off mechanism as shown in figure 3, we set a variable whose stage is stage_k in our scheme, to some extent, the stage_k could reflect the system state. While the stage number is high, it is indicated that the number of stations is many and the congestion of the network is serious. While in a lower stage shows a better network condition. Because the stage of each station is different before a successful transmission, so we can use the different stages of stations to control different back-off methods.

Different schemes can be chosen through controlling the stage_k. With fewer stations, adopting the DCF back-off mechanism because of it's simple and easy to achieve. If there are many stations, we adopt the Slow CW Decrease scheme [6]. For the successful transmission, don't return to the CWmin to select back-off counter, otherwise results in unfairness and collision problems. Setting a threshold value K, when stage_k <K, we select DCF mechanism, and select Slow CW Decrease scheme when stage_k>K. If previous back-off stage is big enough, stage_k>K, it adopts stage_k-g back-off window before the next transmission (g is a constant, the value is 1 or 2), the window descends after each successful transmission. When stage_k<K, it adopts DCF back-off mechanism again [7]. This scheme not only reduces collision probability when there are many stations, but also the unnecessary delay due to back-off window is too large when there are few stations. Thereby, the use of system resources is optimized.

B. *EASDCF scheme*

A great variety of the improved DCF protocols in IEEE802.11 can not provide different packet drop rate for the frame with different service levels [8]. Voice and data services have large difference in the sensitivity of delay. In 802.11DCF mechanism, both of them have the same retry limit and packet drop rate. The voice service is sensitive to delay and insensitive to packet drop rate, but in opposite the date service is sensitive to packet drop rate and insensitive to delay, so DCF mechanism is inappropriate. Basing on service classification provided by IEEE802.11e draft, we put forward the EASDCF scheme.

EASDCF combines the advantages of EDCF with adaptive Slow CW Decrease, which focuses on resolving the problems of access delay, packet drop rate and throughput. From the idea of service classification in EDCF, making different service level frame have distinction in access delay and packet drop rate through setting different minimal contention window and maximal retry limit. Based on adaptive Slow CW Decrease, this scheme introduced a parameter adjustment method, thus throughput is greatly improved.

First, the method of delay classification in this scheme is the same as EDCF protocol in 802.11e draft [9]. The frames are divided into M grades. The relationship among minimal contention window of each grade is:

$$CWimin > CWjmin \quad j>i$$

Each grade has the same size of maximal back-off stage, that is: $mi= mj=m=5$, $i, j \in [0,M]$

According to the above principles, the minimal contention window of each grade is divided in table 1 as following:

Table 1: The sizes of contention window of each grade

Level	1	2	3	4	5	6	7	8
CWmin	32	64	128	256	512	1024	2048	4096
CWmax	1024	2048	4096	8192	16384	32768	65536	131072

Besides, packet drop rate also take into consideration, we divide the frame into N grades. The relationship of maximal retry limit among each grade is: $mq > mp$, $q>p$, $p, q \in [0,N]$.The division method is as following in table 2:

Table 2: Maximal retry limit of each grade

leveli	1	2	3	4	5	6	7	8
Shot frame	4	6	8	10	12	14	16	18
Long frame	7	9	11	13	15	17	19	21

Delay and packet drop rate belong to two aspects of evaluation network QoS performance. Through the combination of them, from Table 1 and Table 2, we know that EASDCF obtain 64 different QoS classification services in all.

IV. SIMULATION AUTHENTICATION

NS2 is a simulation tool for optimizing real network traffic transmission by establishing network devices and links [10]. In the process of simulation, adopting right scale to evaluate the QoS of the WLAN is a key point. It

is paramount for results evaluation and reliability of performance of the entire network. Therefore, we authenticate performance improvement in three aspects, throughput, average delay and packet drop rate.

A. *Simulation parameters setting*

For Simple, simulation model conforms to IEEE802.11b standard, transfer rate of 2Mbit/s for all simulations. The following is other parameters' setting:

- ●Transmission delay: 1μs
- ●Timeslot size: 20μs
- ●SIFS: 10μs
- ●back-off stages: 5
- ●In each scenario, the number of station increases from 10 to 60 at the increment of 10.

B. *Simulation scenarios setting*

Simulation scenarios as shown in figure 4.

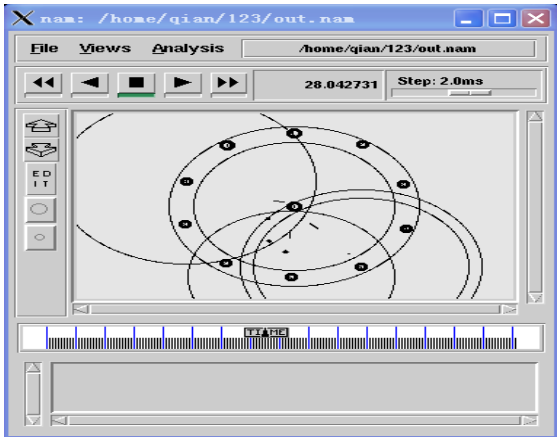


Fig.4 Simulation scenarios

All stations access a WLAN server, which has the function of AP (Access Point), carrying FTP service, transmitting in TCP protocols, and each data frame size is fixed at 1000 bytes, 60 seconds for each simulation time as well. In addition, for observing the stable state, we should eliminate the impact caused by transient variables generated in the course of initialization. Therefore, only the states after 10 seconds of each simulation should be considered.

C. *Brief description*

In simulation scenarios, in order to simplify the model, we make sure that each station supports one grade, that is to say, it only can send data frames in the same type. What's more, for the convenience of simulation, both EDCF and EASDCF merely take two QoS grades into consideration. Table 3 describes the parameter and situations of division in detail.

Table 3: Parameter in three scenarios

	DCF	EDCF	EASDCF
Class 1 CWmin	32	32	32
Class 2 CWmin	32	64	64
Class 1 RLmax	7	7	7
Class 2 RLmax	7	7	9

V. RESULTS AND ANALYSIS

A. *Simulation results and analysis of throughput*

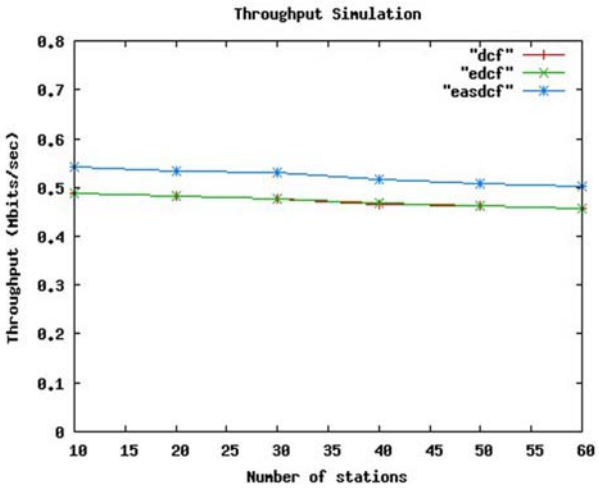


Figure.5 Overall throughput without services differentiation

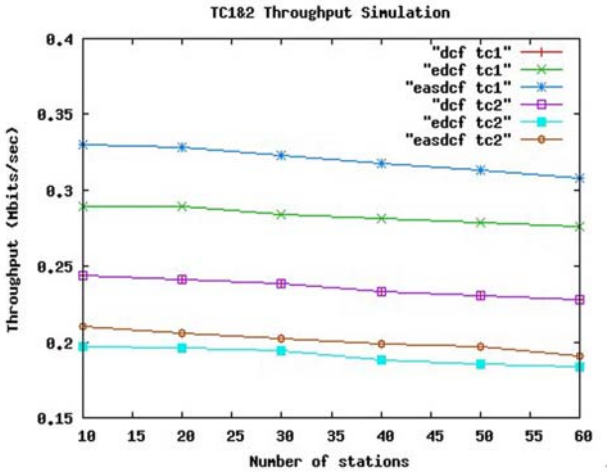


Figure.6 Throughput of TC1 and TC2

As is shown in figure 5 for the performance of throughput, DCF is close to EDCF, but EASDCF is superior to both of them. EDCF mechanism sacrifices low priority services to satisfy the need of delay sensitive services. However, it could not enhance the overall throughput of the system.

From table 3 we know, TC1 and TC2 don't differentiate priority in DCF model, they have the same contention window and maximal retry limit, so they have the same probability in media access authority, TC1 and TC2 have the same throughput, whose lines are two superposition curves. Nevertheless, it differentiates priority in EDCF and EASDCF, in comparison with TC2, TC1 has a smaller contention window and a greater opportunity to media access, which lead to a larger throughput. The two services in EASDCF have larger throughput than in EDCF, especially in the greater enhancement of TC1. In this way, not only ensure the priorities of transmission real-time services, but also improve the overall throughput performance about 10% - 15%.

B. *Simulation results and analysis of delay performance*

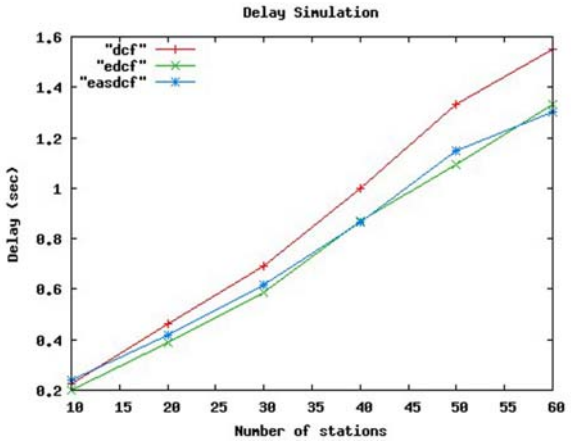


Figure 7 Overall delay without services differentiation

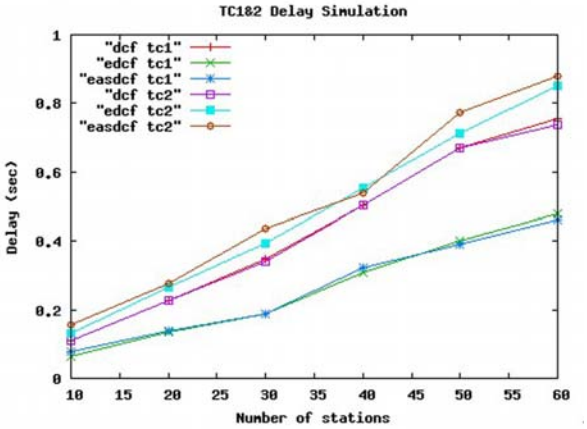


Figure 8 Delay of TC1 and TC2

Delay simulation curves as shown in figure 7, totally speaking, the delay of DCF is the largest, especially as the number of station increases it becomes more and more evident. Because DCF always minimizes the window after a successful transmission, when there are many stations, the time spent on back-off may obviously increase. But there aren't great differences in EDCF and EASDCF.

Figure 8 depicts the relationship between the station numbers in the network and the frames of TC1 and TC2. The frame of TC1 in EDCF and EASDCF have less delay than TC2, it is mainly caused by the small contention window of TC1. Besides, in EASDCF the maximal retry limit of TC2 is 9, so its delay is a little larger than EDCF. Meanwhile, due to adopting the adaptive back-off strategy, the probability of collision reduces and the time spent on back-off waiting shortens. Generally speaking, TC2 frame of EASDCF has few differences with that of EDCF. Namely, the adoption of adaptive back-off strategy compensates the delay to some extent which is caused by the increment of the maximal retry limit.

From the curves of figure 8, we can know, EASDCF and EDCF make a classification of delay. The delay curves of TC1 and TC2 have evident difference. It offers a guarantee of QoS for services with different delay requirement.

C. *Simulation results and analysis of packet drop rate*

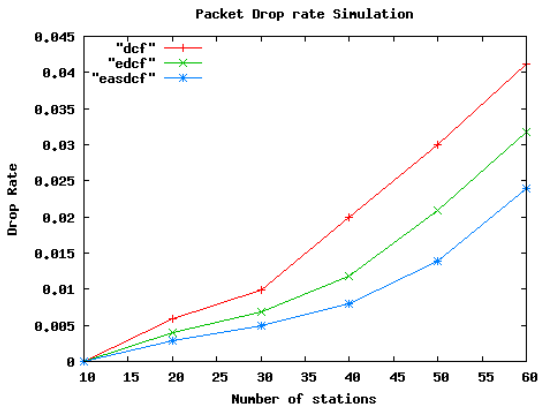


Figure 9 Simulation of packet drop rate

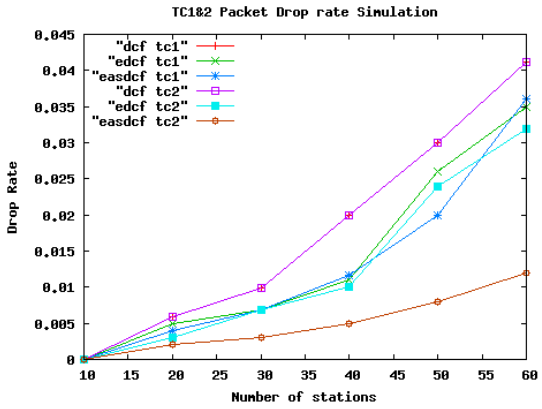


Figure10 Packet drop rate of TC1 and TC2

Figure 9 shows the packet drop rate of the three modes. Conspicuously, from the curves DCF has the biggest packet drop rate. When the number of station in the network surpasses 40, the packet drop rate of EASDCF is lower than DCF and EDCF evidently.

Likewise, we now analyze the packet drop rate of each TC. As shown in figure 10, the frame of TC2 in EASDCF has larger retry limit than that of TC1, so the packet drop rate of TC2 is fewer than TC1. EASDCF has the lowest packet loss rate, there are two reasons. One is that TC2 frame of EASDCF has maximal retry limit, which increases the retry limit leading to the reduction of packet drop rate. Another is that the property of adaptive back-off, it will select a back-off way based on the working condition of the network, thus it will reduce the possibility of collision. In all frames, the TC2 of EASDCF has the least packet drop rate, thus it can offer a reliable guarantee for the date services.

VI. CONCLUSION

This paper proposed a new algorithm for the MAC access, and it synthesized the two methods of slow decrease back-off and services classification. Compared with the proposed back-off algorithm in WLAN, its advantage is that: (1) EASDCF mechanism can combine with the existing IEEE802.11 access protocol without need to introduce any additional expenses and other

hardware devices. (2) Through the mechanism of largest contention window and retry limit, it can offer services classification. (3) It can efficiently reduce the number of collision under high load network environment, and improve efficiency in the use of wireless channel. The simulation results proved that EASDCF scheme can effectively enhance throughput performance of real-time services, improve packet drop rate of data service to a great extent. It is meaningful for us to enhance the QoS performance in WLAN.

REFERENCES

[1] IEEE Standard 802.11, Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications[S]. 1999

[2] Ng P.C, Liew.S.C. Throughput Analysis of IEEE802.11 Multi-hop Ad hoc Networks[J]. IEEE/ACM Transactions on Networking 2007 15(2), p:309-322.

[3] PE Engelstad, ON Osterbo. The Delay Distribution of IEEE 802.11e EDCA and 802.11 DCF[C]. Performance, Computing, and Communications Conference, 2006,IPCCC 25th IEEE International 10-12, April 2006, p:96

[4] H.L.Vu, T.Sakurai. Accurate Delay Distribution for IEEE 802.11 DCF[J]. Communications Letters, IEEE Apr 2006 10(4), p:317-319.

[5] Jiang T, Shao S, Wei K etc. Research on an Improved Algorithm Based on IEEE802.11e EDCF Mechanism[C]. Wireless Communications, Networking and Mobile Computing, 2008. 4th International Conference on Dalian China, p:1-3.

[6] Qiang Ni, Imad Aad, Chadi Barakat etc. Modeling and Analysis of Slow CW Decrease for IEEE 802.11 WLAN[C]. Personal, Indoor and Mobile Radio Communications, PIMRC 2003. 14th IEEE Proceedings on 7-10 Sept. p:1717-1721

[7] Kuppa S, Prakash R. Adaptive IEEE 802.11DCF Scheme with Knowledge Based Backoff[C]. Wireless Communications and Networking Conference IEEE. March 2005(1), p:63-68

[8] Zhang M, Gong C, Lu Y. Dynamic Priority Backoff Algorithm for IEEE802.11 DCF[C]. Computer Science and Software Engineering International Conference 2008,Wuhan, Hubei, (4), p:956-958.

[9] IEEE 802.11 WG, Draft Supplement to Standard for Telecommunications and Information Exchange Between Systems-LAN/MAN Specific Requirements Part 11: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Medium Access Control (MAC) Enhancements for Quality of Service (QoS), IEEE 802.11e/Draft 4.2, 2003.

[10] The Network Simulator-NS-2. http://www.isi.edu/nsnam/ns/

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 223-226

Research on Resource Locator in P2P Network

Zhihong Zhao[1], Guangxue Yue[2,3], and Libo Yang[1]
[1] School of Information Science & Engineering,Jiangsu Polytechnic University,Changzhou 213164,China
[2] College of Mathematics and Information Enginering,Jiaxing University,Jiaxing 314001,China
[3]Guangdong University of Business Studies ,GuangZhou, China
Email:zhao-zhihong@163.com

***Abstract*—It is a key problem of P2P network to how to find resources efficiently and accurately, and this issue will give affect to the application of P2P system in other fields directly. Through inducing the relevant literature systematically this paper gives the research of current situation. This study summarizes the major models of resource locator in P2P networks and analyses the advantages and disadvantages of these models systematically, and gives a model of probabilistic transmittance based on Gnutella. This paper also discusses the prospects of future studies.**

***Index Terms*—P2P; Distribute network; Resource locator; Gnutella; Probability of Transmission**

Ⅰ. INTRODUCTION

As the development of networking, the scale of the network is become larger. Too many nodes in the network lead the mode of C/S to discover its shortage. In this mode the server can't meet the demands of the clients. How to integrate the various network devices which are laid in different places and have heterogeneous system is an effective way to solve the problem.

The appearance of P2P[1,2] network is to overcome the weak point of the C/S. All of nodes in P2P network are on equal terms and they don't only have the same liability but also complete transaction in cooperation (this node is called servent). Every node in this system is not only a client but also a server. Through the way of direct connection between servents, the nodes in system can share information, processor, memory or cache and other resources in order to remove information block and the bottleneck of the C/S[3]. The substance of P2P is to change the mode of management in network from concentration to separation. In this way the resources can be used effectively and the capability of the network will be enhanced.

But due to nodes have the lack of knowledge of system, It is a key problem of P2P network that how to find resources efficiently and accurately, and this issue will give affect to the application of P2P system in other fields directly.

Ⅱ ANALYSIS THE CURRENT SITUATION OF P2P RESOURCE LOCATER

It can be divided in four categories according to the topology of P2P network: centralized topology P2P, distributed unstructured P2P, distributed structured P2P, hybrid topology P2P.

A. Centralized Topology P2P

There is a server in this system which has a catalog of the shared resources about this system. Every node in this system relies on this catalog to search resource. The main steps can be list as follows:

(1) The node just joined the system has to connect the directory server to provide its own information of shared resources. The server put the information into its database.

(2) If a node needs to search some resources, it gives a sign to the server first and the server will give the address of the node which has the resources based on its catalog database.

(3) The node making request can connect to the certain node according the address from the server and the transaction will be disposed by the two nodes. In this transaction the server doesn't take effect any more.

In the process we can see that the server only takes part in the one and two step. So the server has more capacity to deal other things. Napster is the typical case.

The main advantage of this topology is that the high efficiency of resource locator and the algorithm is simple to realize. But there are some shortages of this topology can be list as follows:

(1) This network relies on the directory server too largely. If the server in the system has broke down whole system will paralysis.

(2) As the scale of the network become more and more large, the server will be the bottleneck of the system.

B Distributed Unstructured P2P

Unlike centralized P2P topology, distributed unstructured P2P is a fully distributed structure. In this system, nodes connect with other nodes(neighbors) directly. Each node has a local database to maintain information about its shared resources. While node provides service to other nodes, it can send request to other nodes. System can complete the search and download through the communications between nodes. The overlay network uses random graph and the dimensionality of the nodes obey the law of "Power-law". So it can find resources efficiently and has strong ability of fault-tolerant and to face dynamic changes of the network. At the same time it can support complex queries, such as multiple keywords with a regular expression queries, fuzzy queries, etc. The typical case is Gnutella [4,5]. The topology of distributed unstructured P2P is shown in figure 2.

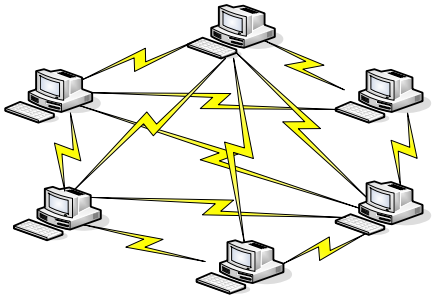


Figure 1. : Topology of distributed unstructured P2P

In Gnutella, nodes use flooding to forward messages. If the average of the degrees of the nodes is n_c and the hops of the messages of the Ping or Query is *h*, then when active node sends a message of Ping or Query there would be $n_c \cdot \frac{(n_c-1)^h-1}{n_c-2}$ messages of Ping or Query at the worst[6]. The number is an exponential growth function on the *h*. It would consume a lot of bandwidth while the *h* is big because there would be a lot of messages. But when the *h* is little, the search scopes would be small and lead to can't find the needed resources.

To reduce the number of messages in the distributed unstructured network, there are three methods about this problem listed as follows.

(1) Based on improved the mechanism of forwarding method. This method is improving the algorithm of flooding by doing not forward message to all neighbors. Literature[7] put forward Directed BFS. In this way all nodes preserve statistical data such as the delay to neighbors. Peer would select some optimal nodes to sends messages. The number of messages can be reduced largely by this way. Intelligent-BFS put forward by Literature[4] is similar to Directed BFS. It sends messages to the neighbors which have most return results. k-Walker[5] is that every active nodes sends k different Walker and every Walker can select next node by random. If the hops are *h* there would be k×h messages at worst and reduce the number of messages largely. Literature[8] gave a method called hybrid periodical flooding, HPF.

(2) Based on cache. This way contains index cache and content cache. Literature[1] found that the network communications on the edge of ISP are reduplicated mostly and the ratio of hitting the target in the cache reaches 67 percent while web cache is only 30 to 60 percent. Some systems put the index to normal nodes. All nodes preserve the indexes about files on the nodes when there are certain hops between them. When the Peer receives the query, it can provide response for the nearby Peer. With locality of query, literature [2,9] put forward every peer cache the results of search through itself. Literature[10] studied the KaZaa network by simulation and showed that content cache could reduce the demand of bandwidth.

(3) Based on topology optimization. Literature [11] discussed the mismatch between overlay network and physical network, has designed a method called Location-Aware Topology Matching, LTM and assessed the efficiency of LTM in dynamic and statistically network. The result of the experiment showed that LTM can reduce 75 percent of traffic and 65 percent the query response time. Literature[12] proposed a distributed solution of non-structural overlay network does not match the problem ACE (Adaptive Connection Establishment). The simulations show this method can reduce the cost of 65 percent for each query and 35 percent response time.

The methods above are cutting high-cost connections, establishing low-cost connections, but there are short of proof of the correctness of matching operations, and no analysis of their matching method would bring forth the negative impact of net. In addition, all these methods eliminate mismatching after the mismatching appearance. This will produce a lot of messages for eliminate mismatching.

C Distributed Structured P2P

There are a lot of researches about how to construct a highly structured system because of the poor scalability of unstructured systems. The point of current researches is how to find information effectively and most recent results are based on DHT (Distributed Hash Table) which is a distributed discovery and routing algorithm. These algorithms don't use flooding like Gnutella but through a distributed hash function. While enter the keywords it maps to a certain node, and establish a connection with the node through different rules of routing. The typical representatives are Chord, CAN [13,14].

Chord use the compatible with the hash function to assign *m* bits NodelID to every node and m bits DataID to every resource. All nodes make up a logistic hoop by NodleID mod 2^m**.** The data of *k* is stored in the first one whose NodelID>=DataID and the node is called the successor nodes of the data *k* and labeled successor(*k*). When there are *N* nodes in system each node preserves the other O(log (N)) nodes information and store it to routing table called Finger Table. In this table, the item *i* contains the information of node *j* and meet *j*=successor(n+2^{i-1}) where 1<*i*<=m. Chord uses the Finger Table and the method of similar binary search to find the certain node of particular data. So every search needs the O(log(N)) messages. It needs O($\log^2$(N)) messages when a node take part in or quit the system to update the table[13]. The topology of Chord is shown in figure 3.

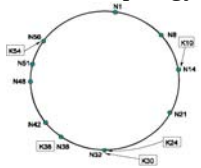


Figure 2: Topology of Chord

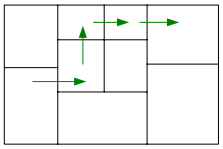


Figure 3: Topology of CAN

CAN which is designed by the Center of AT&T ACIRI uses a d-dimensional coordinate space. Each node in system is mapped to a subspace. It use hash function to map shared data to a point in the space and this data will be stored in the node which contains the subspace. Every node has the address information of its neighbors stored

in routing table. When search resources CAN uses greedy algorithm and send the messages to the neighbor node which is nearest to target according the routing table. Each node has O(d) neighbor nodes, and the hops of searching is $O(dN^{1/d})$[14]. The topology of CAN is shown in figure 3.

The biggest problem of DHT structure is that the mechanism of maintenance is too complex. In particular while there are nodes always join or quit the system, it becomes difficult to maintain. The other shortage of DHT is that it supports exact keywords search only. It can't support the content of complex queries.

D Hybrid Structure P2P

Hybrid structure P2P integrates the advantages of the centralized topology and distributed topology. In this system, there are some supernodes which have high performance such as the ability of compute, bandwidth, memory and so on. These supernodes have stored the information of other nodes. Discovery algorithm is carried out between supernodes only and then the certain supernode forward the query to the specific bottom nodes. Hybrid structure P2P is a semi-distributed and hierarchical structure. Super nodes make up a high-speed forwarding layer and this super nodes and normal nodes constitute several levels. The most typical case is KaZaa. This topology is shown in figure 4.

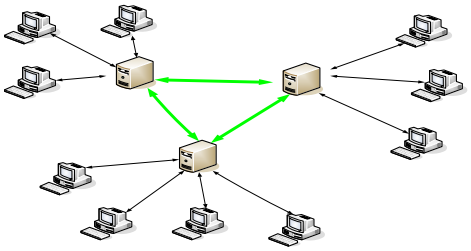


Figure 4: hybrid structured P2P

The advantages of hybrid structured are that the performance and scalability of it can reach a preferable level. Of course it also relies on the supernodes and it is easy to be attacked. But as a compromising method although all aspects of its performance has failed to reach the optimal level of the current structures, it has more practical value and attracted many researchers.

The company CA's statistics indicate that there are more than 250 million times download on KaZaa. Using KaZaa to transport files consumes 40 percent of the Internet bandwidth. It was so successful because it combines the advantages of Napster and Gnutella. It uses the distributed structure of Gnutella and so it has high scalability. It selects nodes which have high peformances and let these nodes become super nodes. These super nodes store the information about the bottom nodes nearest to them and they connect together to make up an overlay network. As the supernodes have indexing about certain bottom nodes, the performance of searching has been improved greatly.

Now, many studies are based on hybrid structure. The focus is using relevant estimate system and specific trust computing to select right nodes to be supernodes and how to dispose the problem while supernodes are attacked or disabled. Literatures[15] is a typical document about this problem.

III AN IMPROVEMENT BASED ON GUETELLA

It is found that searching always focus on a few popular resources. So there is a model of probabilistic transmittance based on Guntella been put forward.

In this model, every node in this system builds a routing table about popular resources on the basis of the searching which relevant to itself. Nodes send messages according to the table.

While they echo ping or query the messages of Pong or Queryhit should take some additional information such as processing power, bandwidth, current load, online time, delay of connection and so on. Source node renews its routing table after received a response from target node. The additional information make up a vector *Y*, and all components can be divided in two kinds, positive component (bandwidth), negative component (delay of connection) based on practical meaning. In order to be easy to dispose we let all negative components to be inverse to themself and there a vector *X* in which all components is positive components. Routing table has build n item of routing of most popular resources based on the ability of the node such as memory. The popular resources are towards to a certain node. Every keywords has *m* targets and it can be controlled to send *c* request ($c \le m$, while $c=m$ sends request to all targets directly). With respect to a certain target i we pull-in a parameter of transmission k^i. it can be compute by iteration shown in equation ①.

$$k^i_n = f(X, k^i_{n-1}) \quad ①$$

The function f() needs to be able to correctly reflect the impact to the probability of every component as accurately as possible. There are two basic conditions listed as follows.

Condition 1: $\frac{\partial f}{\partial x_i} \ge 0$

Condition 2: $\frac{\partial^2 f}{\partial x_i^2} \le 0$

（x_i is any component of X）

The probability of the target i is :

$$P_i = \begin{Bmatrix} (k^i \times c) / \sum k^j (P_i \le 1) \\ 1(else) \end{Bmatrix} \quad ②$$

(j∈N, N is the set of target nodes toward the certain keywords)

In theory, nodes in this model transmitting messages are based on the routing tables not using flooding and so it has these advantages such as reducing the demand of bandwidth, improving the efficiency of searching and

controlling the load-balance. The actual performance of the model remains to be proved by experiment.

Ⅳ SUMMARY AND OUTLOOK

There are a lot of research results of P2P network nowadays. But there are some problems needs to be resolved such as how to finding correct resources effectively and controlling the number of the messages in searching process. Beside this, the problem of churn and semantic query is key problem of distributed unstructured P2P and how to select supernodes effectively and how to do while supernodes are disable.

ACKNOWLEDGMENT

This work is supported by Economic Commence Market Application Technology Foundation Grant by 2007gdecof004.

REFERENCES:

[1] Dou wen,etc. A Recommendation-Based Peer to Peer Trust Model. [J] Journal of Software, 2004, 15(4):571~583.

[2] V. Kalogeraki,etc. A Local Search Mechanism for P2P Networks. In CIKM, 2002.

[3] Fox G. Peer-to-Peer networks. [J]Computing in Science & Engineering. 2001.3(3):75~77.

[4] Huang Daoyin,etc. A Distributed P2P NetworkModel Based on Active Network. [J] Journal of Software, 2004,15(7):1081~1089.

[5] Xu Fei, etc. Design of Distributed Storage System on Peer-to-Peer Structure. [J] Journal of Software, 2004, 15(2):268~277.

[6] Mihajlo A, Jovanovice B S. Modeling Lager-scale Peer-to-Peer Network Architecture[C], In: International Conference on Commumication Technology (ICCT2003) Proceeding,2003.1222~1232.

[7] Wang Qingbo, etc. An Infrastructure for Attribute Addressable P2P Network: Barnet. [J] Journal of Software, 2003,14(8):1481~1488.

[8] Ma Yan. Design and Implementation of DoubleID Space Based Peer-to-Peer File System. [J] Computer Engineer, 2005,31(10):12~19.

[9] B. Yang H. Garcia-Molina. Efficient search in Peer-to-Peer networks. Proceedings of ICDCS'02, 2002.

[10] Q. Lv, etc. Search and replication in unstructured P2P networks, Proceedings of the 16th ACM International Conference on Supercomputing, 2004.

[11] Xiong Wei,etc. A Structured Peer to Peer File Sharing Model with Non-DHT Searching Algorithm.[J] Journal of Computer Research and Develop 2009, 46(3): 415~424

[12] Z. Zhuang,etc. "Hybrid Periodical Flooding in Unstructured P2P Networks". Proceedings of International Conference on Parallel Processing, 2003.

[13] Stoica I,Morris R,Karger D,et al.Chord:A scalable peer-to-peer lookup service for Internet applications.[C] Proc of the ACM SIGCOMM 2001. New York:ACM,2001:149-160.

[14] Stoica I,Morris R,Karger D,etc.Chord:A scalable peer to peer lookup service for Internet applications. [C] Proc of the ACM SIGCOMM 2001. New York: ACM, 2001: 149-160.

[15] Yue Guangxue, etc. An Efficient Virtual Co-Allocation Service Model in P2P Networks.[J] Journal of Computer Research and Develop. 2008,45(11):1928~1938.

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 227-229

Emotion Recognition of EMG Based on BP Neural Network

Xizhi Zhu
School of Physical & Electronics Engineering, Taizhou University, Taizhou, China
zxztzc@126.com

***Abstract*— This paper presents the emotion recognition of BP neural network classifier. Experiment analyzes wavelet transform of surface Electromyography (EMG) to extract the maximum and minimum multi-scale wavelet coefficients firstly. And then we enter the two kinds of structural feature vector classifier for emotion recognition. The experimental results show that both classifiers shows a effective recognition on the four kinds of emotions, joy, anger, sadness, pleasure, etc. Also the experimental results show that feature vector extracted by wavelet transform can characterize emotional patterns, and it has a stronger emotional recognition effect than the traditional classify method.**

***Index Terms*—Surface Electromyography(EMG)Signal; Emotional Recognition; Wavelet Transform; BP Neural Network**

I. INTRODUCTION

The study of emotion recognition has an important significance in understanding human emotions in the role of human intelligence. In daily life, human intelligence not only shows in the normal rational thinking and logical reasoning ability, but also in the normal emotional capabilities. In computer science, this ability to promote the establishment of a friendly man-machine interface is of great significance. As the deepening of affective computing, the request for emotion recognition technology will be correspondingly enhanced. [1, 2]

The MIT media experiment team led by Professor Picard collected a 20-day physiological signal from eight kinds of emotions of an actor when performing deliberately, to extract the value of demographic characteristics using way of Fisher Projection [2]. At last they get 83% of the emotion recognition rate [3]. In Augsburg University, Germany, Johannes Wagner and others got approximately 80% of the emotion recognition rate on a subject's evoked in the music under joy, anger, sadness, pleasure of the four kinds of EMG signal by the use of physical K-nearest neighbor, linear discriminant function and multi-layer in 25 days [4]. Among them, Yang Ruiqing and Liu Guangyuan led the feature selection by the use of discrete binary particle swarm optimization (BPSO), proving that the method of BPSO on physiological signal for feature selection is feasible [5]. Also Niu Xiaowei and Liu Guangyuan selected the most representative characteristics of the corresponding optimal combination of emotional states by using the genetic algorithms, proving that the optimization problem, the genetic feature selection algorithm being used to identify the choice of the combinational optimal portfolio of the physiological signal recognition optimal emotional characteristics, is feasible [6].

In this paper, the method of Wavelet Transform is used in surface EMG aiming for non-stationary features of surface EMG signal in order to extract more effective, reliable, robust signal characteristics. This will help improve the recognition rate of the surface EMG. Also this paper uses the surface EMG signal with objective data for six-scale decomposition of surface EMG with the method of wavelet transform and extract the maximum of multi-scale wavelet coefficients, constructing 14-dimensional feature vector, then we input into the BP neural network for emotion pattern recognition.

II. RECOGNITION METHODS OF EMOTIONAL MOTION

BP neural network is a multi-level error feedback network proposed by Rumelhart and Mc Clelladn in 1985. It uses the difference between the actual output and the desired output to correct the connection weight of the network and the threshold of each node in each layer from back to front. BP neural network contains an input layer, a middle layer (hidden layer) and an output layer. There is a full connectivity between the upper and lower layers and no connections between neurons in each layer. For the input signal, it needs to spread towards to hidden layer nodes and transformed by the function, then transmit the input signal of hidden layer nodes to the output layer nodes. Usually, the transfer function of BP neural network is Sigmoid Type differentiable function, which can achieve arbitrary non-linear mapping between the input and output, so BP network has been widely applied in pattern recognition, function approximation and other areas [7, 8].

The three nodes of the BP network is represented as:

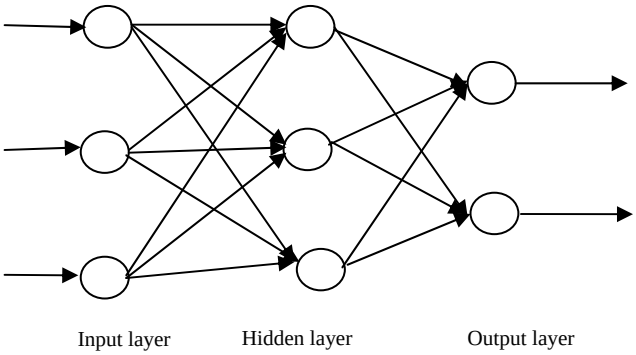


Figure 1. BP neural network structure

input node x_j, hidden node y_i, output node σ_j, network weight of input node w_{ij}, network weight of hidden node and output node T_{li}, the expectation output the output node t_l, Figure 1 shows the BP neural network structure.

The basic learning algorithm of BP neural network:

l) Determine the various learning parameters on the basis of the known network structure, including input layer, hidden layer, output layer neuron number, learning rate and error parameters.

2) Initialize the network weights and thresholds.

3) Provide learning sample: input vector and target vector.

4) Start to learn, and do the following for each sample:

① Forward-calculation of the j unit in the l layer:

$$v_j^{(l)}(n)=\sum_{i=0}^{T} w_{ji}^{l}(n)y_i^{l-1}(n) \tag{1}$$

Equation (7) $y_i^{l-1}(n)$ is the signal transmitted from the i (i=0, set $y_0^{l-1}=-1$, $w_{j0}^{l}(n)=\theta_j^{l}(n)$) unit of the $(l-1)$ layer.

If function of the j unit activation is sigmoid function, then

$$y_j^{(l)}(n)=\frac{1}{1+\exp\left(-v_j^{(l)}(n)\right)} \tag{2}$$

And

$$f'\left(v_j(n)\right)=\frac{\partial y_j^{(l)}(n)}{\partial v_j(n)}=y_j(n)\left[1-y_j(n)\right] \tag{3}$$

If the j unit belongs to the first hidden layer $(l=1)$, then

$$y_j^{(0)}=x_j(n) \tag{4}$$

If the j unit belongs to the output layer $(l=L)$, then

$$y_j^{(L)}(n)=O_j(n) \tag{5}$$

$$e_j(n)=d_j(n)-O_j(n) \tag{6}$$

② back-calculation of δ:

For the output units,

$$\delta_j^{(l)}(n)=e_j^{(L)}(n)O_j(n)\left[1-O_j(n)\right] \tag{7}$$

and for the hidden layer units,

$$\delta_j^{(l)}(n)=y_j^{(l)}(n)\left[1-y_j^{(l)}(n)\right]\sum_k \delta_k^{(l+1)}(n)w_{kj}^{(l+1)}(n) \tag{8}$$

③ Fix the right values according to the following:

$$w_{jk}^{(l)}(n+1)=w_{ji}^{(l)}(n)+\eta\delta_j^{(l)}(n)y_i^{(l-1)}(n) \tag{9}$$

5) Enter a new sample until it reaches the error requirement, and the input order of each cycle in training samples needs a re-random order.

The specific program flow chart of training network using BP algorithm shows as fig.2.

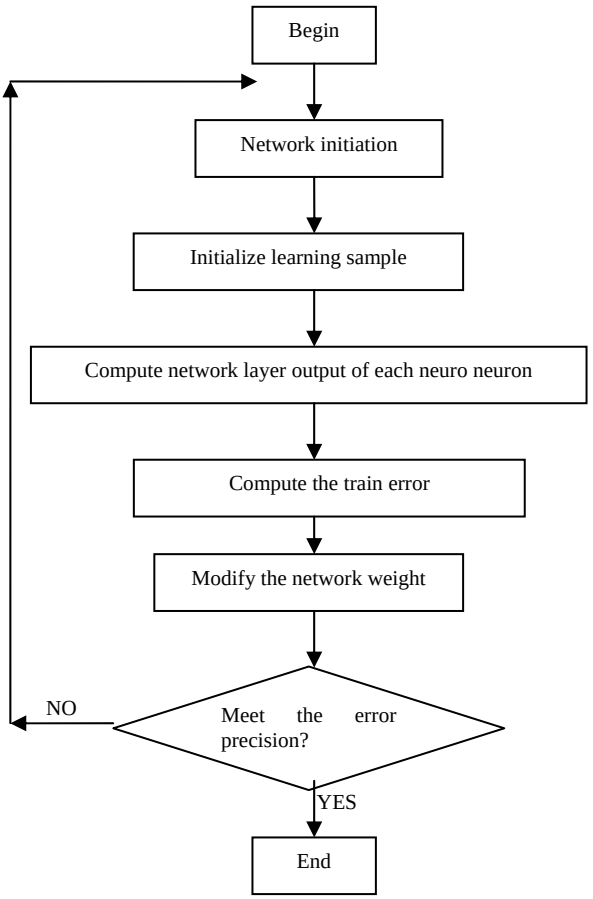


Figure 2. BP algorithm Flow Chart

III. EXPERIMENT RESEARCH

Experiment was carried out in the Matlab7.6 environment. The physiological signal data of EMG is from the Augsburg University in Germany, it is four kinds of emotions, joy, anger, sadness and pleasure, generated by a subject's conduct of music by Johannes Wagner and

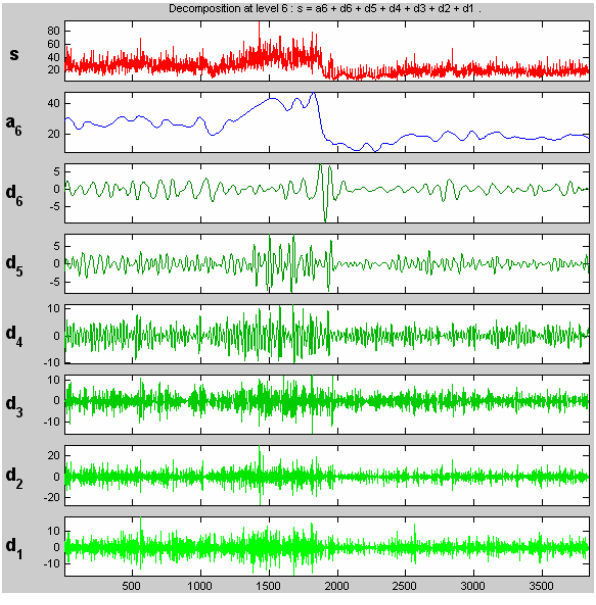


Figure 3. The EMG joy signal of wavelet transform in six scales

others through the selective emotional music, with a total of Record a 25-day EMG physiological signals whose signal sampling frequency is 32Hz.

This paper uses a quadratural, compact Daubechies wavelet as the base function for six-scale decomposition of the EMG physiological signal data each day. And extract the maximum and minimum values composition vector of each layer in wavelet decomposition as the feature vector of the surface EMG signal vector, constituting a 14-dimensional feature vector. The EMG signals Waveform of joy and wavelet transform coefficients Wf (a, b) in different scales are shows as fig.3.

For emotional joy pattern recognition next procedure, Table I shows the maximum value of 6 level wavelet decomposition in five experiments.

TABLE I. the maximum of typical surface EMG joy signal

	joy						
	a6	d6	d5	d4	d3	d2	d1
1	137.9	29	11.99	14.38	16.67	13.76	7.097
2	130.6	24.51	10.11	40.3	27	50.58	27.75
3	209.8	20.83	40.21	18.28	18.53	17.68	18.77
4	133.4	24.95	11.21	23.47	16.94	18.01	9.231
5	91.27	21.91	9.811	13.92	8.421	8.224	8.942

Then a three-layer BP neural network is used in this paper. Input nodes number is 14 and the output nodes number is 4, which represents four kinds of emotional states, respectively, joy(1000), anger(0100), sadness(0010) and pleasure(0001). For how number of the nodes selection in the middle hidden layer, experiments show that the hidden layer nodes have a significant impact for the performance of neural networks. If we select too few nodes, each category can not be separated by the network, and if too many nodes, the operation is too big, there maybe "over learning". Therefore system performance and efficiency must be taken into comprehensive consideration to determine the hidden layer nodes. In this study, after many experiments comparison, finally we select 14 as the hidden layer nodes and the effect is quite good. The training sample set is closely related to network performance. To design a good set of training samples, it is necessary to select the sample size and the quality of the sample is important. That is, the determination of the samples number and sample selection organization is important. In this paper, data of 19 days are selected as the training set. Through experimental selection and comparison, the remaining data of six days are selected as test set. Learning rate has a great impact on network performance too. The experiments show that when the learning rate is 0.01 and the precision control parameter will under 0.01.

Then we get the result of emotion recognition by the use of BP neural network. After many experiment, the recognition rate can reach more than 83.33%. We can see that the emotion recognition using the BP neural network is feasible. BP neural network for solving linear equations used to achieve the classification makes the training time greatly reduced.

IV. CONCLUTION

Emotion recognition has a promising development prospect. Identifying the person's emotional state through the physiological signal has drawn increasing attention. This experiment introduce multi-scale decomposition wavelet of EMG signals by wavelet transform and extract the maximum and minimum of wavelet decomposition coefficients to construct signal feature vector to present the original EMG. Then enter it into the standard BP neural network classifier for emotion recognition. The kind of classify is able to detect and identify the surface EMG of four kinds of emotions, joy, anger, sadness and pleasure. Compared to the classical classifier, the emotion recognition has a better classification effect, higher recognition rate and better robustness. Experimental results show that the surface EMG feature extraction based on wavelet transform and emotional type recognition method using BP neural network as a classification tool is feasible and effective in application of emotion recognition.

REFERENCES

[1] Luo Senlin, Pan Limin. Affective Computing Theory and Technology [J]. Systems Engineering and Electronics, 2003,25 (7) :905-909.

[2] Picard R. Affective Computing [M]. England: MIT Press, 1997:68.

[3] Kim KH, Bang SW, Kim S R. Emotion recognition system using short-term monitoring of physiological signals [J]. Medical Boilogical Engineering Computing, 2004, 42 (3) ,419-427.

[4] Wagner J, Kim J, AndréE. From physiological signals to emotions: implementing and comparing selected methods for feature extraction and classification [J]. IEEE International Conference on Multimedia & Expo.New York: IEEE Computer Society Press, 2005:940 -943.

[5] Yang Rui request, Liu Guangyuan. BPSO four physiological signals based on the emotional state identification [J]. Computer Science, 2008，35(03),137-139.

[6] Niu Xiao-wei, LIU Guang-yuan. Physiology signals based on genetic algorithm emotion recognition [J]. Southwest University (Natural Science Edition), 2007, (9) ,134-138.

[7] Wu Fang-liang, Shi Zhong Kun, Yang Xiang-hui, et al. Based on BP neural network with LM algorithm submarine sonar self-noise prediction [J]. Chinese shipbuilding, 2006,47 (3): 45 - 50.

[8] Kun, Wang Zhizhong. Using LM algorithm is an improved BP network in Surface EMG Signal in the applied research [J]. R & D on the, 2005,29 (6) :399-401.

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 230-233

Design of Electron-agriculture System Base on Information Visualization and Windows Live Platform

Meiyu Zhang, and Fengfan Zou
Zhejiang University of Technology, Hang Zhou, China
zmy@zjut.edu.cn
Zhejiang University of Technology, Hang Zhou, China
zfftxwd@163.com

***Abstract*—In order to solve our country existing agricultural information system defects: such as function sole, interactive poor, difficult for users to operation, as well as the information resources is sole and redundant. The paper has developed an so-called electron-agriculture system using information visualization technology, electronic geography information technology, Visual C#.NET, Silverlight, and SQL Server. The design simplifies the user operation, friendly human-machine interface and involved in interaction.**

***Index Terms*—information visualization, windows live, silverlight, electronic geography information technology**

I. INTRODUCTION

Taking into account the main user is farmers, and the farmers do not well in using complicated PC and surfing the Internet, while the traditional systems had not considered this problem, leading to the farmer have not enough enthusiasm to participate in countryside informatization. So the design of the Electron-agriculture system based on Windows Live platform and information visualization mainly consideration the man-machine interface functions, and the business function of traditional system operation is replaced by virtual earth and google earth to come true. By using the technology of information visualizing users use simple 、intuitive and other easy operate ways, such as using mouse to click, to use this system. User doesn't need to have the professional knowledge of geographic information, can be surfing Internet to view agriculture informations, such as space information, graphics and pictures, and can use the information to help users to analyze.

The system realizes a four tier system structure based on B/S (browser/server). The whole system should have good expansibility and commonability, so it is convenient to carry on the secondary development and expand on windows live id platform that it includes virtual earth, unified user identity authentication (Live ID), online communication by Windows Live Messenger, Windows Live Search and Silverlight streaming. Combining Silverlight technology to compose the rich client, by using this technology, we will have a rich content and wonderful visual effect of flowery interactive experiences.

II. OVERALL DESIGN

A. System function

The system mainly has four modules. As shown in Fig.1.

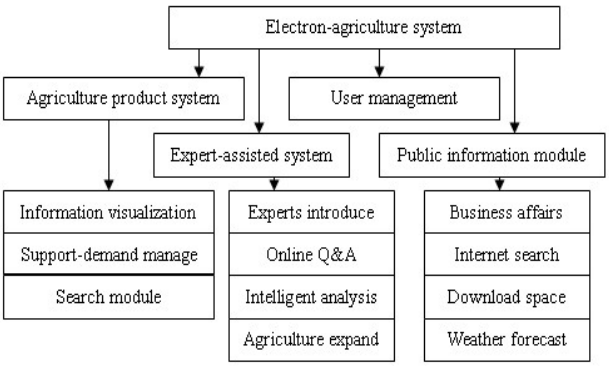


Figure 1. The main function modules

Agricultural information module can be divided into three smaller modules, agricultural information visualization module, agricultural supply management module and search module. The main function of agricultural information visualization module is view and contrast agricultural information by using silverlight that it shows an intuitive comparison through the wonderful visual effect chart. The main function of agricultural supply and demand management module is to look over and release supply and demand information. Search module mainly have supply and demand search, price search, technology search, market search and comprehensive search. Allowing users to search the entire system resources so as to quick location resources.

Expert-assisted module mainly introduces various experts, on-line question and answer with experts, intelligent analysis. Interactive information through QQ and MSN. You can ask for help from expert online or send leaving message to experts. By using the technology of visual/Google earth and Silverlight to show the analysis of distribution line on the map and to show the price forecast analysis through the chart.

The user management module manages the permission of system user and group management that it helpful for multi-user system management tasks.

Public information module provides users with some common traditional electron-agriculture service functions,

AP-PROC-CS-10CN006

such as weather information, technology information, links and other information.

B. Design and development the platform based on Windows Live

This system will use Windows Live platform for secondary development, including a lot of Microsoft's technology [3]; design scheme is shown in Fig.2.

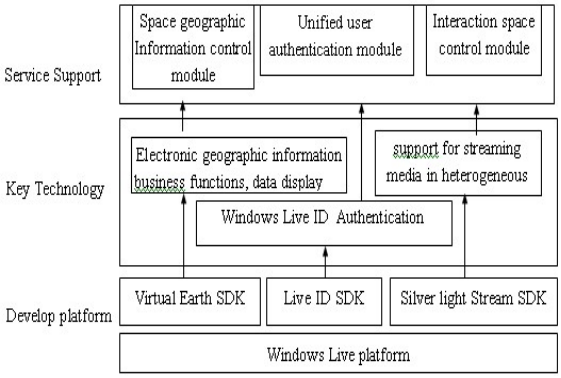


Figure 2. system design

The windows live id platform support the interface of Virtual Earth SDK to show the graphic information data, and using Google earth to show the detail information, because Google earth have a better graphics detail. Realize unified identity authentication through Live ID SDK interface, then user could login the system without registering, but also can register in the system and login in. Silverlight Stream SDK interface is used to access heterogeneous information that support streaming. The three key technologies have always been throughout the business system applications.

C. Visual design on electronic agriculture information

The system uses virtual/google earth map information and silverlight applications to realize information visualization Design shows in Fig.3.

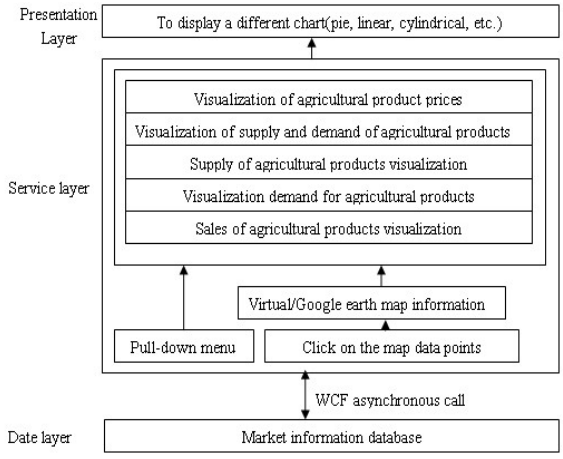


Figure 3. Informatization information visualization design

In the agriculture information visualization design, the user through the drop-down menu that shows a certain period of time or a particular market for a product or by clicking on the virtual/google earth map information to display the data [2] [3]. Among them, visualization of agricultural product supply and demand, the different regions or markets, on the map, using different colors to indicate the relationship between supply and demand, when click on a specific region, it will use chart to display the specific data, such as the Fig.4 shows.

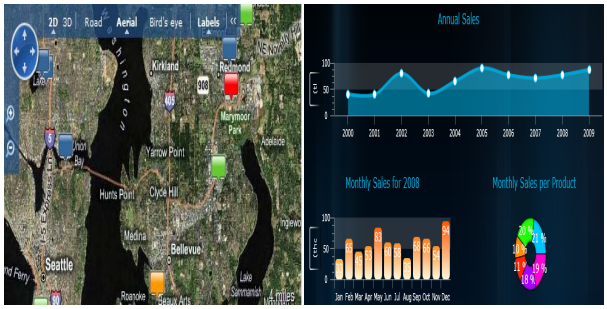


Figure 4. Display data through the map and silverlight chart

In the left may, the red region means demand exceeds supply, blue means oversupply, green represents a considerable supply and demand. When the user double-click on these information points, when it will carry out various forms of graphical charts to show in Fig.5.

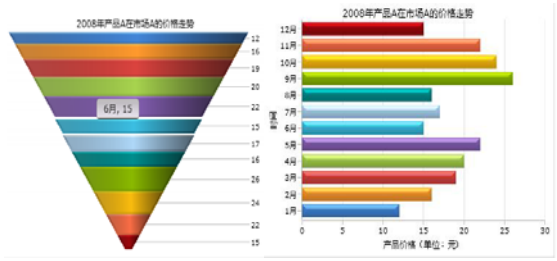


Figure 5. Graphic chart showing

You can also through the drop-down menu to choose, market, price, supply and demand, displaying mode, etc. The selection can be demonstrated as shown in the graphic charts.

Judging from what has been argued above, we have come to see that the design support more interactive behavior, rich the presentation tier logic. It allows the user to achieve common dynamic data view by normal clicking operation, reducing the interruption and waiting. It is helpful for user to analysis the market, and improves the experience significantly.

III. WINDOWS LIVE PLATFORM AND SILVERLIGHT FUNCTIONAL DESIGN

A. The application summary of the Windows Live platform in the business functions layer

In order to achieve the system's target, simplifies the operation, friendly system interface and participate interactively, we use Windows Live platform for secondary development, mainly contain three techniques, Virtual Earth, Live ID, and Silverlight stream.

In addition to Internet access, there are cell phone accesses, telephone, GPS positioning and so on. At the same time, user could log in through the site registered, MSN, QQ and so on in the same access channels.

The system use Windows Live ID, internet unified authentication system, and it is simply to operation and support interaction. It combines the system application

software seamlessly, so as to achieve the premise of a unified authentication, sharing the multiple functions of application platform cross-domain and on-line.

Windows Live ID is Microsoft's identity and authentication services, there are more than 380 million worldwide users have Windows Live ID. It means that the system is integrated Windows Live ID authentication, these users will become potential customers, the system also develop a login system, when you use the Windows live id to log in, the system's login authentication will be synchronized automatically to log in.

The system also applied some other applications on Windows Live platform, such as: Live Search, Live Messenger and Live Space.

Virtual Earth in agricultural production and circulation of information management applications

In recent years, with the developments of technical, map service and globalization is more and more popular, the system incorporates Microsoft's Virtual Earth, with a 2D/3D visual effects, display the distribution of agricultural markets on the map, because the block of information in China is poor in Virtual Earth, so the system use Google Earth to make up for this shortcoming, because the Google Earth in the localization is better that it provides a wealth of geographic information. Taking into account the unity of the platform, we still use Virtual Earth in simple geographic information or large-scale map displays, related to the specific area using Google Earth, so as to achieve the best visual effect.

Virtual Earth and Google Earth combine the effects shown in Fig.6.

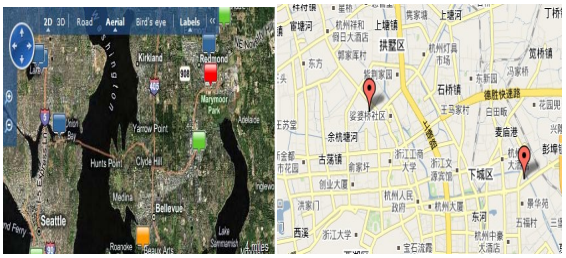


Figure 6. Virtual Earth and Google Earth

Silverlight Streaming is apply in the system interface and streaming media player

In streaming video, it occupies a great deal of server resources. It can reduce the pressure on the server space by using Silverlight Streaming, while the video streaming speed has a very good guarantee. Microsoft provides 10G storage space for Silverlight Streaming application, each file upload increased to about 100M that satisfy large files to upload, the video stream transmission rate is 1400kbps that satisfy the high-definition multimedia video playback smooth, and it is beautiful and easy to use.

B. Silverlight apply in the expert-assisted

Silverlight mainly have two aspects of the application in this system, one at the interface beautify, the other at the expert-assisted that analysis the data through chart. By using the technology of Silverlight to beautify the interface. This system gives a fresh sense of the impact and increase the human-machine interaction.

The system includes Silverlight applications and server applications. Silverlight applications consist of user

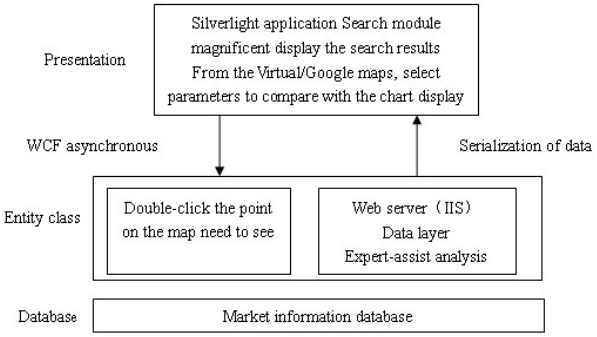


Figure 7. Silverlight showing process

control and style. Presentation tier use asynchronous calls through WCF to communicate with the server. Using the WCF service reference and operation in accordance with service agreement and data agreement to achieve Silverlight applications and services of communication. Data agreement disclosed the entity class structure which is in the server application, and the data combines the Silverlight application to dynamic showing out by using chart or graphical image in the application tier. Above process is shown in Fig.7.

Rich client development techniques have a highly interactive, provides user with a higher and more comprehensive interactive experience, greatly improve user participation rate. Silverlight technology, the important component of rich client development techniques, can develop beautiful graphics and animation function, and supports the interaction with the database. It's not only a simple graphical interface showing, but also implementation the front-end computing power required for the application, the data source through the network to communicate with the background.

In the interface and search module, using Silverlight magnificent style, with simple operations, the visual effect is good for the purpose of display information. Click the picture or information interface to flip the form to convert [1]; search module search results also show a non-traditional manner, the effect shown in Fig.8

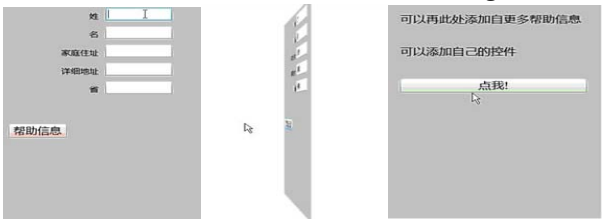


Figure 8. The effect image

Examples: In the registration window, when you click on the help information, the registration box to flip to the back of the display help information, then click click-me Button, and will be re-flip back.

Selecting parameters from Virtual/Google map and to display through chart.

Technical Overview: To view the market price demand or sales' trend on a particular agricultural product, first to click the matching icon, and click the matching link on

the opening window, a related page will be open, then enter the time period that you want to see, results will be come out from the database, displaying the results with Silverlight and selecting the chart type from the pull-down menu. As Fig.9.

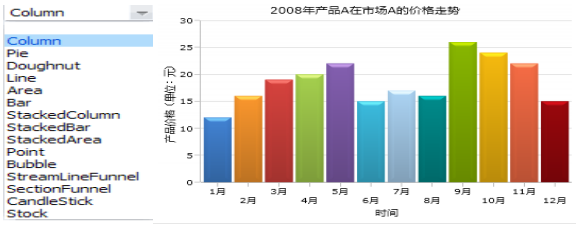


Figure 9. Silverlight charting display

Double-clicking an icon, selecting the current market and comparing the parameters for the selected market, then displaying the results with silverlight directly.

Technical Overview : Double-clicking the market icon in the picture, it shows the current market participating in the comparison if the shape of icon changes, Selecting all the markets that you want to compare with, then clicking the compare button, there will jump out a message box, Let users fill the related compare parameters and time period, server will get the data from database, then displaying the date to the users with silverlight directly.

III. SUMMARY

In this paper, it has totally discussed the overall design on the windows live platform, developing and achieving the design by using key technologies. This system uses the silverlight and WCF service of RIA technology from rich client. The latest RIA service presents a new information systems man-machine interfaces and operating modes which has advantages of simple operation, human system interface and interactive user-oriented. It pursuit for the biggest satisfy to clients.

The system uses NET3.0 Framework, VS2008 development software, Silverlight 3.0 and other applications to achieve the goal which help improving user's experience and participation rate. Though it has much defectiveness, with the quick updated speed of Microsoft and continue developing of internet technology, the functions will be much stronger, at last the perfect agricultural information visualization could be achieved and farmers and society will be got more help from it.

REFERENCES

[1] Li-min Zhang. Master Lecture "Silverlight 2.0 development of technologies". Science Public.

[2] Zhao-feng Tian, Chu-Liang Yan. Agricultural equipment information network platform based on silverlight and resource management. Journal of Agricultural Machinery, 2008. 11:151-155.

[3] Kun Fu. Explore the Silverlight technology in the application of customer experience [J]. SCIENCE & TECHNOLOGY INFORMATION. 2009, 39 (11) :340-341.

[4] Aimin Li, Jing Liu, WeiGang Ma. Silverlight-related research [J]. Computer technology and development. 2009, 19 (6) :117-120.

[5] Hua Fu, Yong Zhong. Rich Internet applications displayed in the application of Web data [J]. Computer Applications. 2009, 29 (6) :292-295.

[6] McCormik A M,Defanti T A. Visualization in scientific computing. Computer Graphics,1987,21(6)

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 234-237

Research of System-Switch Mechanism Based on sCPU-dBUS Architecture

Fengjing Shao, Xiaoning Song, and Rencheng Sun
College of Information Engineering
QingDao University, QDU
Qingdao, China
sfj@qdu.edu.cn, helensongmail@163.com, qdsunstar@163.com

***Abstract*—sCPU-dBUS architecture is a solution for the problem of network security from the angle of changing the computer's architecture. In order to manage the resources on this architecture efficiently, we adopt the operating system with double-subsystems. In this thesis, as an important part of the operating system for the sCPU-dBUS architecture, a system-switch mechanism is designed. Firstly, the kernel bus-switch mechanism is designed. Secondly, the system-switch flows of each stage during the OS's starting and running are designed.**

***Index Terms*—sCPU-dBUS, double-subsystem, system-switch, clock interrupt**

I. INTRODUCTION

As the problem of network security is becoming more and more serious, the solutions aren't confined to the existing ways any more, such as virus scanning technology, firewall technology and IDS. Many experts hope to solve network security problem from a new angle. The single-cpu and dual-bus secure computer architecture (sCPU-dBUS) is a solution from the angle of improving computer architecture. The architecture of sCPU-dBUS is shown in Figure1 [1].

sCPU-dBUS secure computer architecture contains one CPU and two independent fast system sub-buses which are called local bus and network bus. The CPU is connected to the shared bus, and also can be connected with local bus or network bus through controlling the bus bridge [2, 3]. The bridge makes sure that only one bus can be connected with the CPU at any time. Network bus and all the devices which are connected to the network bus consist of network sub-area, while local bus and all the devices which are connected to the local bus consist of local sub-area. In the sCPU-dBUS secure computer architecture, all network devices are connected to the network bus. When network intrusions happen, the destructions will be limited to the network sub-area as the bridge is not connected to the local bus. User's important data are stored in the local-area, so it can avoid being destroyed. As the sCPU-dBUS secure architecture has one CPU and two sub-areas, there should be a befitting operating system to support this architecture. In order to enhance the security of the system, the operating system is designed containing two sub-kernels: the local sub-kernel and the network sub-kernel. The two sub-kernels run in two sub-areas, and they are coordinating relationship rather than subordinate relationship. In order to implement communication between two subsystems, there should be a good system-switch mechanism [4]. In this article, a system-switch mechanism is designed.

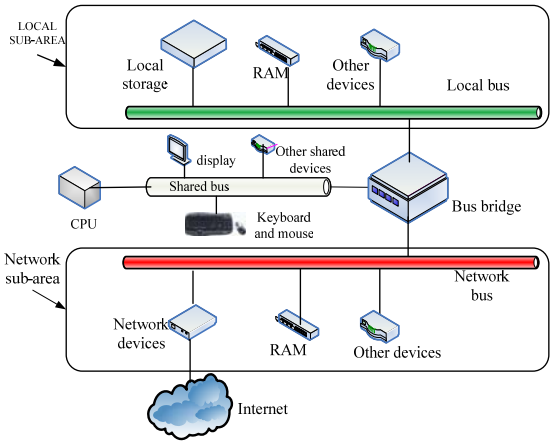


Figure 1. sCPU-dBUS secure architecture

II. SYSTEMIC ANALYSIS OF SYSTEM-SWITCH MECHANISM

System switch is an important part of the double-kernel operating system. At the booting stage of operating system, there are two subsystems needing to be booted, so the system-switch mechanism is needed. And when the operating system is running after it is started, there still exist two situations that need system-switch mechanism to realize system-switch. Firstly, when the user wants to do some operations in the other subsystem, system-switch occurs and this situation is called user-control switch. Secondly, when some data are needed to be transferred to the other subsystem, system-switch occurs and this one is called data-transfer switch. In this situation, system should be switched automatically until the data transfer is done.

In order to make one sub-system continue running from the breakpoint after CPU is given back from the other sub-system, two aspects should be conserved. The first is all of process contexts and the second is address of the next instruction. So two blocks of spaces should be open up to store the information, and the spaces should be on the instruction cache which is mounted on the shared bus for the sake that both of two sub-systems can visit. The two blocks of spaces are shown as follows:

AP-PROC-CS-10CN006

1) Process context storage areas. It is used to store process contexts before present subsystem is interrupted. There are two process context storage areas, which are separately used to store the local subsystem's contexts and the network subsystem's contexts. Certainly, they have different starting address.

2) Switch back registers. It is used to store address of the next instruction that present sub-system is going to execute at the time of being interrupted. There are also two registers in order to store two sub-systems' address of the next instruction separately.

In the next, systemic design of switch mechanism is elaborated. And then, system-switch flows of booting and running stages of operating system are illustrated.

III. DESIGN OF SYSTEM-SWITCH MECHANISM

In order to implement the system-switch between two sub-systems, a system-switch mechanism is designed. Firstly, as it has two sub-buses and they are both connected to the shared bus according to the bridge, some values can be assigned to the bus-bridge controller register to realize connection with different sub-systems. Secondly, as mentioned above, process contexts should be conserved for the sake that sub-systems could run from the breakpoint. And in our mechanism, clock interrupt processor is chosen as the breakpoint of system-switch in order to make system switch in time [5]. Thirdly, one piece of space in instruction cache is opened up to conserve the switch-flag. The switch-flag is a variable which is used to point out whether system need to be switched or which switch direction should be adopted. Both of local sub-system and network sub-system can visit it and change its value. Besides, switch command is designed so that users can control system-switch through command. Specific design is shown as follows:

1) Trigger of system-switch.

System switch mechanism adopts command as the trigger. Users control systems' switch through inputting some commands. After users input switch commands, a process called switch process will be created. This process will set the switch flag according to the parameters of switch command.

2) Entrance of system-switch.

As clock interrupt occurs periodically, it is chosen as the breakpoint. In our mechanism, clock processor is modified. Some procedures are added into it which are used to judge switch flag and then determine whether system should be switched and which direction should be changed to.

3) The core switch program.

The core switch program is called bus-switch program, and it is core module to realize bus switch. Bus-switch program is located in the instruction cache. And apparently it has two parts which are used in the different directions of system switch: local-network bus-switch program and network-local bus-switch program [6]. They are stored in different starting addresses of instruction cache. In this bus-switch program, some functions are implemented as follows:

a) Store clock interrupt processor's context of present sub-system to its process context storage area, such as the data segment, relevant registers, stack and so on.

b) Assign specific values to bus bridge controller register to make shared bus connect to target sub-system's bus.

c) Restore the values of target sub-system's process context storage area to the relevant registers, stack and so on. And then, PC register is assigned as the value of target sub-system's switch back register.

4) Exit of system-switch.

After system bus is switched and the context and PC are restored, the system enters into target sub-system's clock interrupt processor and goes on running from the breakpoint. Figure2 is the flow chart of system-switch.

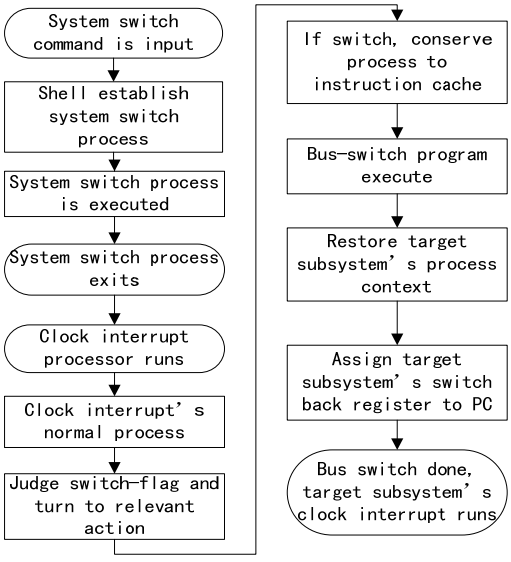


Figure 2. Flow chart of system-switch

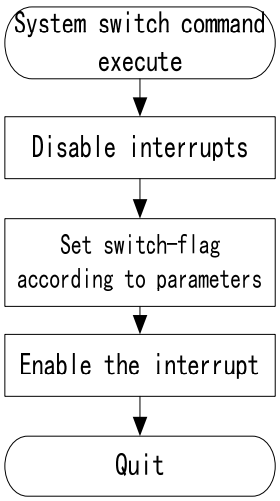


Figure 3. Flow chart of system-switch command

A. *design of system-switch command*

When system-switch command is executed, shell will automatically establish system-switch process. This process is used to set the switch-flag according to parameters of system-switch command. Besides, in order to prevent switch-flag from being changed unexpectedly when next clock interrupt comes, interrupts should be disabled during system-switch process is executed. After switch-flag is set successfully, interrupts will be enabled again [7]. At last, system-switch process exits. Flow chart of system-switch command is shown in Figure3.

In order to accomplish the system-switches in different situations, system-switch command is designed to have different parameters. And system-switch process can refer to these parameters to set switch-flag. System-switch command is called BusSwitch. Parameters of BusSwitch, its parameters' meaning and value of relevant switch flags are shown in table 1.

TABLE 1.
BusSwitch' parameters and values of relevant switch-flag

BusSwitch –p command		**values of switch-flag**	
Parameter's value (-p)	***Meaning of the parameters***	***Local subsystem***	***Network subsystem***
n	Stop system-switch, CPU works at present sub-area.	0	0
u	System-switch under users' control	3	3
c	System-switch during data transmitting	1	2

B. *clock interrupt processor*

The clock interrupt processor is the entrance and exit of system-switch and the controller. So it is the core of system-switch mechanism. Some procedures are added to the clock interrupt processor to judge the switch-flag. Then every time clock interrupt comes, clock interrupt processor firstly does some conventional processing, secondly judges switch-flag and then decides what kind of actions should be adopted [8]. The process flow of local-subsystem's clock interrupt processor is shown in Figure4 (The flow in the braces is the process flow of network-subsystem's clock interrupt processor):

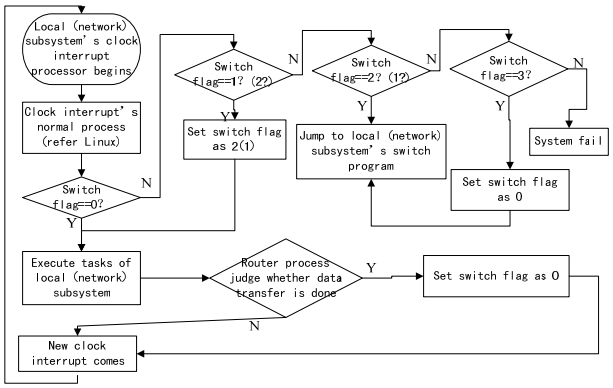


Figure 4. Flow of clock interrupt processor

IV. SYSTEM-SWITCH FLOW ANALYSIS IN DIFFERENT STAGES

System-switch occurs both at the stage of system booting and system running. Absolutely, there are some differences between two stages.

A. *system-switch flow in booting stage*

At this stage, we firstly start local-subsystem in order to assure the system's safety. Local-subsystem's bootloader and kernel are located in flash of local sub-area. System begins to run from the first address of local sub-area's flash automatically. And system-switch occurs when we need to operate network sub-area. In order to assure that system will continue running from the bootloader of network subsystem after system switch, bootloader of local subsystem should do some extra jobs except its normal guidance and load:

1) Initialize all the devices mounted on the shared bus, load two system-switch programs to instruction cache.

2) Initialize process context storage area of both subsystems as zeros.

3) Assign network sub-system's switch back register as the first address of network subsystem's bootloader.

According to system-switch mechanism, users can input the command "BusSwitch -u" to boot network sub-area and operate it.

B. *system-switch flow in running stage*

There are two system-switch situations during the running stage of operating system: user-control switch and data-transfer switch. User-control switch means that system-switch occurs when the users want to switch to the other subsystem and data-transfer switch means that the system is switched automatically when there are some data need to be transmitted to the other subsystem. In this situation, switch between two subsystems are automatically proceeding until data transmission is done. At last, target subsystem goes on running.

According to system mechanism, users can input command "BusSwitch -u" to realize user-control switch while input "BusSwitch –c" to realize data-transfer switch. In situation of data-transfer switch, if local subsystem is running now, system-switch flag is set as 1. When next clock interrupt comes, clock interrupt processor finds switch-flag is 1, and then it sets switch-flag as 2 according to Figure4. Next, it goes on processing local subsystem's tasks. When next clock interrupt comes, system is switched as it judges that switch-flag is 2. After network subsystem starts to run, interrupt return occurs and tasks of network subsystem are executed. In next clock interrupt, it judges switch-flag as 2, and then set it as 1. Tasks of network subsystem are executed again, that is receiving the data. Router process [9] judges whether data transfer has done, if yes, switch-flag is set as 0, and if no, nothing will done. When next clock interrupt comes, switch-flag is judged again, if it is 1, system-switch occurs, if it is 0, system stays at present subsystem.

V. CONCLUSION

In this thesis, sCPU-dBUS secure computer architecture and the requirements of its operating system are introduced at the beginning. Then as the core part of its OS, requirements of system-switch between two sub operating systems are analyzed. According to these, system switch mechanism is designed. At last, the flows of operating system's booting and running are illustrated. This system switch mechanism in this thesis has some reference values to the researchers who have relevant applications.

Secure architecture intelligence network terminal technology (2006AA01Z110), 863 project, 800 thousand, 2006.12-2008.12.

Secure architecture network terminal technology supporting remote reconfiguration, (09-2-3-19-chg), QingDao achievement transformation project, 250 thousand, 2009.09-2011.09.

REFERENCES

[1] Fengjing Shao, Rencheng Sun, Kegang Diao, and Xiaopeng Wang, *"A secure architecture of network computer based on single CPU and dual bus"*, In Proc. Fifth IEEE International Symposium on Embedded Computing, Beijing, China, Oct. 2008.

[2] Tiedong Wang, Fengjing Shao, Rencheng Sun, He Huang. *"A hardware implement of bus bridge based on single CPU and dual bus architecture,"* International Symposium on Computer Science and Computational Technology, Shanghai, China, December 2008.

[3] Le Zhang, Fengjing Shao, Rencheng Sun, *"Design and realization of embedded system developing platform hardware abstraction layer,"* Journal of Qingdao University Engineering & Technology Edition, Vol.21, Mar, 2007.

[4] Wenze Huang, Fengjing Shao, Rencheng Sun, *A New Kind of Task Scheduling of Operating System Based on Dual-Bus Secure Architecture.* 2009'Workshop on the Computer Network and Telecommunication. May 2009.

[5] Rongkai Zhan. *Linux kernel's clock interrupt mechanism.[EB/OL]*. http://www.linuxforum.net.

[6] Ze Tian. *Exploiture and Application of Embedded System [M]*. Bet Jing:Bet Jing University of Aeronautics and Astronautics Press. 2005.1.

[7] Jikun Sun, Xiaoquan Zhang. *Detailed explanation of Embedded linux system development technology- based on ARM [M]*. Bet Jing: Posts and telecom press. 2006.8.

[8] Dongshan Wei. *Embedded Linux application exploitation [M]*. Bet Jing: Posts and telecom press. 2008.8.

[9] Wei Wang，Fengjing Shao. *An Inter-subsystem Data Transfer Mechanism Based on A New Computer Architecture with Single CPU and Dual Bus[C]*. 2009 International Symposium on Intelligent Systems and Application. 2009,10.

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp.238-241

Research and Application of Enterprise Search Based on Database Security Services

Yi Huang[1], Xinqiang Ma[1], and Danning Li[2]
[1] School of Computer Sciences, Chongqing University of Arts and Sciences, Chongqing, P.R China
Email: huangyicq@sohu.com, mxq345@sohu.com
[2] Guizhou Academy of Science, Guiyang, P.R China
Email: lidn121@hotmail.com

***Abstract*—With the development of internet and the build of enterprise information, How to organize the information and make it accessible and useful? Especially, how to support the information security services? They have been become a hot topic of internet and enterprise information. This paper discussed the enterprise search systems based on database security services. Such as Oracle Ultra Search, Sybase Enterprise Portal and LogicSQL-based enterprise archive and search system. Ultra Search enables a portal search across the content assets of a corporation, bringing to bear Oracle's core capabilities of platform scalability and reliability. "Sybase Enterprise Portal Security Services", which describes how Enterprise Portal security works, a description of the security APIs that you customize to accommodate your Enterprise Portal system, and a description of the Web-based Security Administration tool. LogicSQL is independently developed based on the Linux high-level security database system. The development of enterprise internal Search System is based on LogicSQL security database. The enterprise information resources integration and security search are completed in the search system.**

***Index Terms*—security services, enterprise search, Internet search, database security, Network security**

I. INTRODUCTION

In the age of the Internet, proliferation of information is causing a new information management crisis for enterprises. Using the World Wide Web, workers become their own information retrieval experts. But searching for the right answers can be more than frustrating: Information is ultimately stored in corporate databases, Web pages, files in various popular document formats and in email or groupware systems. Servers are located throughout the country and across the globe. Corporate information is distributed across enterprises in both structured and unstructured form - structured relational databases, unstructured Word-processing documents, spreadsheets, presentations. As applications demand transactional consistency, coordinated multi-user access, administration and maintenance for content, a natural gradient is created to move more and more corporate information into databases. However, even when multiple databases are involved, searching across databases needs a robust solution.

Corresponding author: Xinqiang Ma and Danning Li

Many businesses manage both distributed computing environments and disparate data sources. They must gather information, integrate it into the business process, and distribute it to users with different needs. Ultra Search enables a portal search across the content assets of a corporation, bringing to bear Oracle's core capabilities of platform scalability and reliability [1]. The Sybase Enterprise Portal solves the complex problems of providing prompt, secure, personalized, and integrated information to employees, customers, and partners [2]. The enterprise information search management tools based on LogicSQL security database is consistent with the security functions of database [3].

II. INTERNET AND ENTERPRISE SEARCHES

How to organize the information and make it accessible and useful? Especially, how to provide the information security services? There are two mainly searching methods in this information technology times. They are Internet and Enterprise Searches.

A. Internet Search

If you're trying to find a particular page on the Internet, you can use one of the many available online Search Engines. These search engines allow you to search for information in many different ways -- some engines search titles or headers of documents on the net, others search the documents themselves, still others just search other indexes or directories [4]. Internet search sites can search enormous databases of Web pages, using titles, keywords or text. You can maximize the potential of search engines by learning how they work, and how to use them safely, quickly and effectively. Internet search engines are categorized by topic in our searchable directory of general and specialty search engines. Also listed are resources and tools for exploring the deep web, performing advanced research, and for learning about using search engine tools and technology [5].

B. Enterprise search

Enterprise search is the practice of making content from multiple enterprise-type sources, such as databases and intranets, searchable to a defined audience [6]. The term "enterprise search" is used to describe the application of search technology to information within an organization (though the search function and its results may still be public). This is in contrast to the other two

main types of horizontal search environment: web search and desktop search (though the areas overlap, and systems may address these scenarios as well).

The major challenge faced by Enterprise search is the need to index data and documents from a variety of sources such as: file systems, intranets, document management systems, e-mail, and databases and then present a consolidated list of relevance ranked resources from these various sources. In addition, many applications require the integration of structured data as part of the search criteria and when presenting results back to the users. And of course access controls are vital if users are to be restricted to data and documents which they are granted access by the various document repositories within the enterprise. These major challenges are unique to enterprise search.

C. Internet versus enterprise search

The internet and enterprise domains differ fundamentally, such as contents, user behavior and economic motivations.

- The notion of a "good" answer: Internet search is vaguely defined Relevant and Popular. Enterprise search is “right” answered Specific documents.
- Social forces behind the creation of Internet and intranet contents: Internet is created for attracting and holding the attention of any specific group of users, and the collective voice of many authors who are free to publish content. Intranet is created for disseminating information. There is no incentive for content creation, and all users may not have permission to publish content.
- Security and copy rights: Internet includes public domain, and the user has copy rights. Enterprise: access privileges, and usually, the corporation has the copy rights.
- Deployment environments: Internet is controlled by one organization as a service. Enterprise: licensed to and deployed by a variety of organizations in diverse environments.

III. Security Requirements and Infrastructure of two searches

On the public World Wide Web, there's not much of a connection between security and search. You may have logins for certain web sites, and you may use https when buying something with your credit card, but that doesn't have much to do with a general search that you might do on Google or Yahoo. Yes, a few subscription sites limit what non-subscribers can see, or may require an account to read the full text of a document [7].

Enterprise security can be heavily tied to search. In larger organizations, your company login may be used to restrict which documents you can see in the search results list. Access control, usually in the form of an Access control list (ACL), is often required to restrict access to documents based on individual user identities. There are many types of access control mechanisms for different content sources making this a complex task to address comprehensively in an enterprise search environment.

Generally speaking, the security infrastructure and protocols tend to be different between the public web and the enterprise; and when security is needed in search, the search engine must integrate with the available security infrastructure and protocols. The Internet and Intranet both use SSL and HTTPS.

But on private networks, Single Sign On (SSO), LDAP and Active Directory are still the norm. A search application that cares about security will likely be using those protocols. A few enterprise apps still use the older "application level security"; search engine applications can do likewise. We haven't seen the distributed identity assurance model taking root in corporations yet, they still seem to prefer a tightly controlled central resource. One exception to this apathy is between different government agencies; they are starting to realize that government employees frequently need data from other agencies, and that distributed, cooperative security systems make this much more efficient, so that government employees don't have to keep creating new logins for every agency they visit.

IV. Characterizing Enterprise Search

The total is well known that there are various characteristics of the enterprises information. Such as text content in electronic form, diversity of content sources and formats, source access, structured and semi-structured search, federated search, content management, people and behaviors.

Beyond the difference in the kinds of materials being indexed, enterprise search systems also typically include functionality that is not associated with the mainstream web search engines [6]. These include:

- Adapters to index content from a variety of repositories, such as databases and content management systems.
- Federated search, which consists of (1) transforming a query and broadcasting it to a group of disparate databases or external content sources with the appropriate syntax, (2) merging the results collected from the databases, (3) presenting them in a succinct and unified format with minimal duplication, and (4) providing a means, performed either automatically or by the portal user, to sort the merged result set.
- Enterprise bookmarking, collaborative tagging systems for capturing knowledge about structured and semi-structured enterprise data.
- Entity extraction that seeks to locate and classify elements in text into predefined categories such as the names of persons, organizations, locations, expressions of times, quantities, monetary values, percentages, etc.
- Faceted search, a technique for accessing a collection of information represented using a faceted classification, allowing users to explore by filtering available information.
- Access control, usually in the form of an Access control list (ACL), is often required to restrict access to documents based on individual user identities. There are many types of access control mechanisms for different

content sources making this a complex task to address comprehensively in an enterprise search environment.

• Text clustering, which groups the top several hundred search results into topics that are computed on the fly from the search-results descriptions, typically titles, excerpts (snippets), and meta-data. This technique lets users navigate the content by topic rather than by the meta-data that is used in faceting. Clustering compensates for the problem of incompatible meta-data across multiple enterprise repositories, which hinders the usefulness of faceting.

• User interfaces, which in web search are deliberately kept simple in order not to distract the user from clicking on ads, which generates the revenue. Although the business model for enterprise search could include showing ads, in practice this is not done. To enhance end user productivity, enterprise vendors continually experiment with rich UI functionality which occupies significant screen space, which would be problematic for web search.

The factors that determine the relevance of search results within the context of an enterprise overlap with but are different from those that apply to web search. In general, enterprise search engines cannot take advantage of the rich link structure as is found on the web's hypertext content, however, a new breed of Enterprise search engines based on a bottom-up Web 2.0 technology are providing both a contributory approach and hyperlinking within the enterprise. Algorithms like PageRank exploit hyperlink structure to assign authority to documents, and then use that authority as a query-independent relevance factor. In contrast, enterprises typically have to use other query-independent factors, such as a document's recency or popularity, along with query-dependent factors traditionally associated with information retrieval algorithms. Also, the rich functionality of enterprise search UIs, such as clustering and faceting, diminish reliance on ranking as the means to direct the user's attention.

V. Oracle Ultra Search

Oracle Ultra Search allows you to reduce the time spent finding relevant documents on your company's information repositories. It crawls, indexes and makes searchable your corporate intranet through a Web-style search. It eliminates the need for coding against hard-to-use low level API's. It organizes and categorizes content from multiple repositories by extracting valuable metadata that can be used in portal applications [1].

A. Comprehensive Security Support

In Ultra Search, security has been comprehensively addressed:

• Searches return only documents that a particular user is allowed to see – based on group membership and user privileges.

• Integrated with LDAP. All data sources in your enterprise and Ultra Search share a common user .namespace.

• Query API and query samples allow for specifying a search user. Only users known to your corporate LDAP server can submit searches.

• All your resources. Web pages, rows in database tables, emails in IMAP accounts. Can now be protected by Access Control Lists (ACL).

• Ultra Search can obtain ACLs directly from your repositories.

• Search Administrators can specify authorization roles directly in the Ultra Search administrative interface by entering a grant. List of LDAP users, and groups, which are allowed to search documents. For example, all documents retrieved during a crawl can be marked searchable by anyone belonging to groups G1, G2 and G3. Or, all users U1, U2 and U3 can be granted permission to search all documents crawled.

• High performance implementation. ACLs are cached in memory and evaluated at run time during query invocation. Authorization automatically turned off if there are no secure documents in an Ultra Search instance.

B. Ultra Search Applications

Enterprises can benefit by using Ultra Search in many different types of applications:

• Portal Search--Ultra Search offers the most powerful search for Enterprise Portals developed with the Oracle Enterprise Portal Framework. For organizations who want to build their own portal from scratch, Ultra Search provides a canned, end-useroriented, web-style search over various corporate databases, HTML pages, IMAP email servers, or filesystem documents.

• Web Search for Oracle Text--Ultra Search is an application built on Oracle Text, Oracle's industry leading text retrieval engine.

• Content Management Search--Media organizations creating or publishing content in a collaborative manner need to search across content as it moves through multiple repositories in different stages of the content management life cycle: from the desktop file of the author to the staged version in a database.

VI. Sybase Enterprise Portal

Sybase Enterprise Portal provides a personalized, Web-based entry point to all your enterprise information. It delivers a continuously available, end-to-end secure environment that integrates existing business applications and databases, real-time data feeds, business events, and Web content into a coherent information interface.

Sybase Enterprise Portal is a set of services that you can use to address all aspects of the design, development, deployment, and management of a corporate portal. It provides a personalized Web site and the middleware that links data, applications, events, and content between the Web site and the rest of the enterprise.

A. Comprehensive Security Support

User populations for portals can be very large, creating an environment in which traditional user account management and access control cannot function well.

Enterprise Portal security provides a graphics-based administrative interface, relies on standards-based security mechanisms, and provides interoperability with existing Internet and enterprise security infrastructures such as secure socket layers (SSLs) and public-key infrastructures (PKIs) [2].

B. Enterprise Portal security architecture

Enterprise Portal security services include these elements:

- A security servlet on the Web server
- Enterprise Portal Application Server components
- The connection manager
- The object manager
- The SBO
- The Access Control Database (the security repository)

The Portal Security Officer manages Enterprise Portal security using the Enterprise Portal Security Manager, a graphics-based administration tool. The security architecture provides a uniform structure for user authentication and authorization for the entire Enterprise Portal—including both Sybase and non-Sybase applications and components.

VII. LOGCISQL-BASED ARCHIVE AND SEARCH SYSTEM

LogicSQL [8] is independently developed based on the Linux high-level security database system [3]. Ministry of Science and Technology "863" Database Group of the major projects comments LogicSQL database "It is the only hope that China's database products will be able to contend with America's mainstream database products in the next few years."

The core functions of the system as follows: Archiving that replaces the traditional backup functions is completed in the system; Information on the enterprise server and personal computers are protected and managed; The latest search technology is used to realize the rapid and accurate searching for the backup information, including keyword, semantics, the integrated use of intelligent search technology; Search results are consistent with the security level of users.

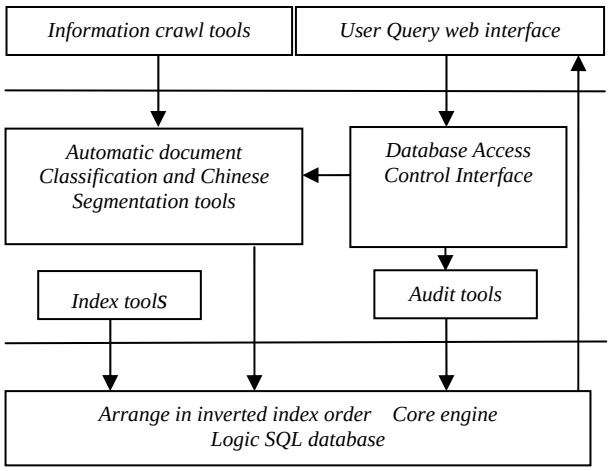


Figure 1. Web crawler architecture

The system architecture graph of the search system is shown as "Fig. 1".

In the search system, the following mandatory access control is adopted in the search process [9]. Access control label is used in specific designer, and it is divided into two parts: security classification and group sets.

Security level can be expressed as: L= (C, G), and C is security grade, and G is the group of set. Two security level: L1= (C1, G1) 、L2= (C2, G2) .If and only if C1 ≥ C2 and G1 include G2, L1 ≥ L2 is correct.

All operation of subject to object are abided by simple-security property and *_ property principle：

(1) When Ls≥ Lo, subject S can read object O, and it is called down-read.

(2) When Ls≤ Lo, subject S can write object O, and it is called up-write.

Audit function is enhanced in the search system based on database. All operation related events are audited, and they increase the security of the search system.

LogicSQL-based Enterprise search system achieves more accurate positioning information, and it realizes the value from the information across to the knowledge.

ACKNOWLEDGMENT

The authors wish to thank Wang Baohua. This work was supported in part by a grant from the Industrial Research Project of Guizhou Province under (Grant No. [2008]5014) and the Science and Technology Funds of Guizhou Province (Grant No.[2009]2125), the Stadholder Foundation of Guizhou Province (Grant No. 2009-111), and the Funds of Chongqing University of Arts and Sciences(Grant No. Y2009JS57, 68 and 69).

REFERENCES

[1] ORACLE DATA SHEET, Oracle Ultra Search, http://www.oracle.com/index.html, February 2010.

[2] Enterprise Portal Overview and User Guide, Sybase Enterprise Portal, LAST REVISED: May 2000.

[3] Yuan Liyan, ShiFang Enterprise Information Real-Time Archive and Search System. http://www.shifang. com.cn/products, July 2009.

[4] Internet Search, http://home.mcom.com/home/ internet-search.html, February 2010.

[5] Search Engines Guide, http://www.searchengineguide.com /searchengines.html, February 2010.

[6] http://en.wikipedia.org/wiki/Enterprise_search, February 2010.

[7] Differences Between Internet vs. Enterprise Search, Last Updated Feb 2009.http://www.ideaeng.com/tabId/98/itemId/154/20-Differences-Between-Internet-vs-Enterprise- Se. aspx#sec, February 2010.

[8] Yuan Liyan. The Documentation of Logic SQL [M]. Alberta University Press, Canada. 2005, pp.1-15

[9] Baohua Wang, Danning Li, Zhengyong Zhao, Xinqiang Ma,Dan Li. Research and Realization of Enterprise Security Search Application System Based on Database. Journal of Computational Information Systems. Vol.3(5).1841-1846. 2007.10.

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 242-245

Study of Electromyography Based on Informax ICA and BP Neural Network

Guangying Yang, and Shanxiao Yang
School of Physical & Electronics Engineering, Taizhou University, Taizhou City, China
ygy@tzc.edu.cn

***Abstract*— Surface electromyography (SEMG) signals decomposition algorithm based on Independent Component Analysis (ICA) are explored. The experiment shows that this method can decompose SEMG signal efficiently on the premise that different motor units are all independent. Then it can be concluded that ICA is a promising method of preprocessing for SEMG decomposition. After SEMG has been reconstructed, we create AR model with the original signal that was pretreated and take the coefficient as its eigenvector. Then, a three-layer BP neural network was designed to classify the muscle movement of forearm with AR model coefficient. The experiment indicates this measure can reduce workload and get the relatively good results.**

***Index Terms*—Independent Component Analysis (ICA); Surface Electromyography Signal; BP Neural Network; Pattern Recognition**

I. INTRODUCTION

Independent component analysis (ICA) is a way of finding a linear nonorthogonal coordinate system in any multivariate data. It is a new technique to separate blind sources, which has been used in some challenging fields of EMG, ECG, EEG processing. The directions of the axes of this coordinate system are determined by maximizing the statistical independence of the estimated components. Under a certain condition, we can separate independent part from source signal [1].

Electromyogram (EMG) is a signal obtained by measuring the electrical activity in a muscle has been widely used both in clinical practice and in the rehabilitation field [2]. Clinical analysis of the EMG is a powerful tool used to assist the diagnosis of neuromuscular disorders [3]. BP neural network (BPNN) is backpropagation algorithm in the medical field for the development of decision support systems [4]. In this paper, we discuss EMG signal by adopting the Informax ICA calculating way which has better effect to resolve the surface muscle telecommunication signal. And it can be a preprocessing mean to decompose the surface telecommunication signal.

II. RECOGNITION METHODS OF MUSCLE MOTION

A. Auto-R*egression*(AR) Model

Parametric model is an important method for analysis of electromyography signal, where the most typical is AR Mode. According to the AR model's theoretical analysis, we could see that the parameter selection is critical. Proper parameters are conductive to the recognition and parametrical evaluation of AR model. This paper adopts one of the evaluation way is direct evaluation which derives the model parameters directly from observed data or statistic characteristics of the data [5]. The recognition of mode [6] evaluates the model's parameters with a stable model structure and degrees by the means of auto covariance function and partial correlation function's character of truncation according to the information implied in a sample derived from the time sequence.

The stationary AR (p) model is stable if the roots λk(a), k = 1 . . . . . p, of the associated characteristic polynomial have moduli that are less than unity.

$$a(z) = 1 - \sum_{k=1}^{p} a_k z^{-k}, z \in C \qquad (1)$$

We define the stationary AR model as stable with margin 1 - p if all roots of the model's characteristic polynomial a (z) lie inside a circle of radius p < 1 in the complex plane. Correspondingly, a time-varying AR (p) model is stable if the roots of the corresponding time-varying characteristic polynomial.

$$a(z;t) = 1 - \sum_{k=1}^{p} a_k(t) z^{-k}, z \in C \qquad (2)$$

We define an AR model as hyperstabte if all roots of the model's characteristic polynomial lie inside a circle of radius p _< 1 in the complex plane. In this paper we present a method for estimating hyperstable AR models. Although the proposed method is applicable to other tranversal AR parameter estimation schemes, we discuss here only the nonwindowed least squares (LS) method.

The choice of model order p poses great problems. According to the past research and experiment [6], the experiment sets AR model to four because higher value of the order will not improve the performance but also will add burden of computation.

B. *Informax ICA Algorithm*

The idea behind using the Independent Component Analysis (ICA) is to reduce the redundancy of the original feature vector components. Assume that there is an N-dimensional zero-mean non-Gaussian source vector

This work is supported by education department Program of Zhejiang Province in University (2010) and Yong teacher Program of Taizhou University (09qn09)..

$S(t)=[S_1(t),...,S_N(t)]^T$, whose components are mutually independent. The vector s(t) corresponds to N independent scalar valued source signal si(t). An observed data vector $X(t)=[X_1(t),...,X_M(t)]^T$ is composed of linear combinations of sources si(t) at each time point t such that[7,8]:

$$x(t)=As(t)+n(t) \tag{3}$$

Where, X is the original feature vector, S is the underlying (independent) sources, and A is a mixing matrix. Only X is observed, and ICA algorithm estimates both S and A then trying to find the sources which are as independent as possible through a linear transformation W.

$$Y(t)=WX(t)=WAS(t) \tag{4}$$

The problem of ICA addresses the reconstruction of n independent source signals from m observed signals, possibly via estimation of the unknown mixing matrix. The overall structure of the ICA model is shown in Fig.1. This neural processor takes X as an input vector. The weight W is multiplied to the input X to give U and each component ui goes through a bounded invertible monotonic nonlinear function g(i) to match the cumulative distribution of the sources.

Infomax is the shortened form of the criterion of Infomax ICA [9,10]. It is introduce by researchers of Salk Institute calculate nerve biology laboratory firstly. The point of Infomax is introducing a nonlinear function $r_i=g_i(y_i)$ to replace the estimation of higher-order statistics. The block of Infomax is shows as fig.1.

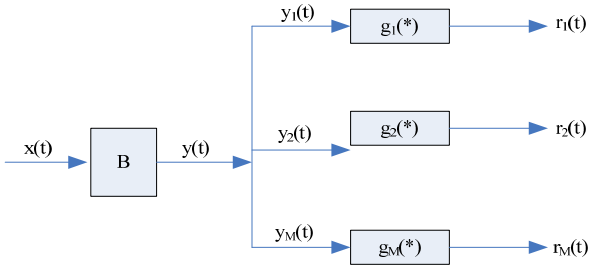


Figure 1. The block of Informax ICA

Assume that there is an M-dimensional zero-mean vector s(t)=[s1(t),…, sM(t)], such that the components si(t) are mutually independent. The vector s(t) corresponds to M independent scalar-valued source signals si(t). We can write the multivariate of the vector as the product of marginal independent distributions:

$$p(z)=\left|\det(w)\right| p(y) \tag{5}$$

Where, $p(z)$ is the hypothesized distribution of p(s). The log likelihood of equation (1) is as follows:

$$L(y,w)=\log\left|\det(w)\right|+\sum_{i=1}^{M}\log p_i(y_i) \tag{6}$$

Maximizing the log-likelihood with respect to W gives a learning algorithm for W as follows:

$$\Delta W \propto [(W^T)^{-1}-\phi(y)x^T] \tag{7}$$

We define function $f(w)=-w+\phi(y)y^Tw$, solve differential coefficient of w and obtain $Jf(w)$ as follows: :

$$Jf(w)=-I+\frac{\partial\phi(y)}{\partial y}xy^Tw+2\phi(y)x^Tw \tag{8}$$

Then we can get the learning algorithm as follows :

$$w^+=w-\frac{f(w)}{Jf(w)}=w-\left(-w+\phi(y)y^Tw\right)D(y) \tag{9}$$

Where,

$$D(y)=-\left(I-\frac{\partial\phi}{\partial y}xy^Tw-2\phi(y)y^T\right)^{-1}$$ is the weight be adjusted.

Suppose $\phi(y)=y+\tanh(y)$, we obtain the final steps of the algorithm :

1) The initial value $w_0=w$ is given , observe vector z ;

2) Calculate $y=wx$, $\phi(y)$, $D(y)$;

3) Calculate $w^+=w-\left(-w+\phi(y)y^Tw\right)D(y)$;

Repeat step (2) and (3) until it convergents to get w , then we use the formula $y=wx$ and obtain the independent components.

III. Experiment Research

The experimental system in this paper is composed of three parts, i.e., surface myoelectric signal acquiring instrument of Noraxon U.S.A. Inc., data gathering card NI-6024E and its software system Labview8.0 of National Instrument (NI) Company and MatlabR2008 system of U.S.A. A pair of surface electrodes is placed on the extensor carpi ulnaris and flexor carpi ulnaris of healthy testees respectively, each of which consists of 3 Ag-Agcl electrodes. Then select a muscle which is out of activity as the referenced point to ground.

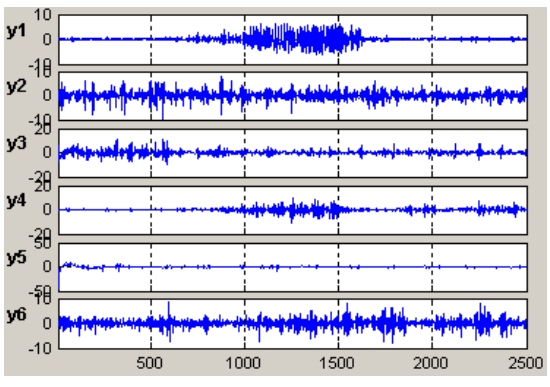


Figure 2. EMG disposed by Informax algorithm

One of the sample data shows as fig.2, it contains of six lead of EMG signals. After that EMG is disposed by Labview software and the motion pattern of prosthetic hand will be classified. We adopt 2500 points in the original source signal so it can provide enough data when using ICA to decompose it. The decomposing process of SEMG signal is that we decompose SEMG signal to each motor unit which constitutes motor electric potential component. Each motor unit can be regarded as independent component. If we suppose each source signal is statistic independently, and the process of creating SEMG can be regarded linearly. The Informax ICA algorithm is applied to decompose SEMG, after the experiment, we get the result in fig.3, which shows that the source signals have been separated and the basic motor unit action potentials are obtained.

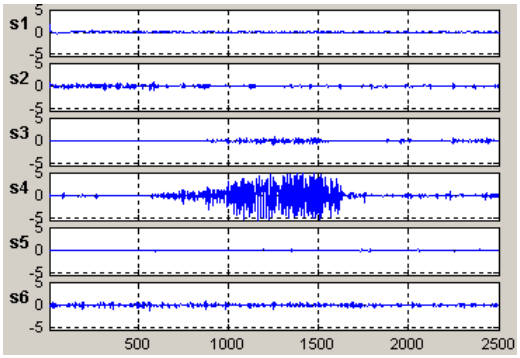


Figure 3. EMG disposed by Informax algorithm

After many experiments, we find motor unit shrinked by MUAPT s4 changing obviously when extensor carpi ulnaris transfer its motion pattern, so s4 will be used for the following analyzing procedure, s_1 - s_3 , s_5 - s_6 will be discarded as noise. Then we extract 4-order AR coefficient in Labview. Tanbe.1 shows typical 4-order AR coefficient derived from AR model of motions of extensor carpi ulnaris. The acquired average values of the 4-order AR coefficients, as the characteristic vector of the surface electromyography signals, are used in the recognition of motion mode.

TABLE I
THE 4-ORDER AR COEFFICIENTS OF ELECTROMYOGRAPHY SIGNALS OF EXTENSOR CARPI ULNARIS

	AR Coefficients			
Pattern	A01	A02	A03	A04
Laxation	-0.28	-0.05	0.02	-0.03
Hand opening	-0.59	-0.19	0.53	-0.16
Hand closing	-0.83	0.07	0.29	-0.21

In this paper, a three layer BP Neural Networks was applied to construct two models of ANN with Sigmoid function. The issue of modifying a typical BP neural network in a practical way to meet EMG pattern recognition is discussed. Then muscle's pattern recognition procedure is given by Neural Network in MATLAB which shows as fig.4. The structure of input

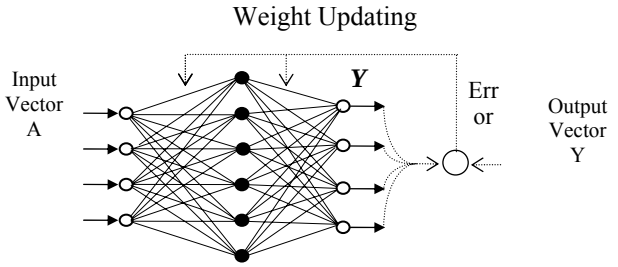


Figure 4. Three layer BP Neural Network

layer, the number of implicit layer units, the request of input, the study of velocity and the deliver of function, etc, are analyzed. With the ameliorations, the capability of the network is improved.

The BP neural network is trained to be able to simulate the pattern recognition procedure. We adopt the AR coefficients of different pattern as input of the BP neural network. Input layer have three nodes which names as Y_1 , Y_2 , Y_3 which corresponding to pattern Laxation, hand opening, hand closing respectively. To avoid the influence of gain in the procedure of signal processing, we normalize the AR coefficients in input layer.

IV. CONCLUTION

This paper introduced Informax ICA to analyze SEMG signal. Under the premise that each source signal is independent, the experiment shows that it can eliminate unmeasurabled noise jamming, and separate each source signal independently. At the same time, the experiment shows that the effect of Informax is well. Then we analyze surface electromyography signals and extract 4-order AR coefficient from AR model as characteristic vectors. The electromyography signals of pattern Laxation, hand opening, hand closing could be recognized by BPNN. The proposed method achieves performance in terms of categorization, recognition ratio and robustness. The experiment proved that it is viable and effective to extract from surface electromyography signals based on the AR mode and use BPNN as a categorizing tool to recognize movement in the application of recognizing wrist movement.

But the result also shows that the indeterminacy in amplitude and permutation has some trouble in signal disposal. Recently, a new blind signal separation method which names natural gradient ICA is presented to separate medical signals. It is said that the indeterminacy in amplitude and permutation can be eliminated by using this method. So the author's next plan is to confirm the validity of this new method by simulation experiment in the area of SEMG.

REFERENCES

[1] A. Hyvärinen and E. Oja. Independent Component Analysis: Algorithms and Applications[J]. Neural Networks, 2000, 13(4-5):411-430.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] Kumaravel, N., and Kavitha, V., Automatic diagnosis of neuromuscular disease using neural network. Biomed. Sci. Instrum. 90:245–250, 1994

[3] Butchal, F., An Introduction to EMG, Copenhagen, Gyldendal, 1957.

[4] Frize, M., Ennett, M., Stevenson, M., and Trigg, C. E., Clinical decision support for intensive care unit using ANN. Med. Eng. Phys. 23:217–222, 52001.

[5] Chen Liangzhou, Shi Wenkang. Time Sequence Analysis Random Sets Method[J].College Journal of Shanghai Communication University 2005,39(3),400-404.

[6] Liu Jiahai. Evaluation Research on Muscular Function Status based on SEMG Signal Non-linear Kinetic Analysis [M].Thesis of Doctor Degree Zhejiang University，2001.

[7] A.Hyvarinen, A Family of Fixed-point Algorithms for Independent Component Analysis[J]. In(ICASSP'97), IEEE Int. Conf. on Acoustics, Speech and Signal Processing Munich, Germany, 1997, 45(8) , 3917-3920.

[8] A.Hyvarinen, Fast and Component Analysis[J]. IEEE Robust Fixed-point Algorithms Traps. On Neural Networks, for Independent 1999, 25(10), 626-634.

[9] J.-F. Cardoso(1999). Higher order contrast for independent component analysis. Neural Computation. 11(1):157-193.

[10] V. Soon, et al. (1990). An extended fourth order blind identification algorithm in spatially correlated noise. Proc. IEEE ICASSP:1365-1367.

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 246-249

ECG Pattern Recognition Based on Wavelet Transform and BP Neural Network

Shanxiao Yang, and Guangying Yang
School of Physical & Electronics Engineering, Taizhou University, Taizhou, China
ysxtzc@126.com

***Abstract*—This paper introduces the Electrocardiogram (ECG) pattern recognition method based on wavelet transform and standard BP neural network classifier. Experiment analyzes wavelet transform of ECG to extract the maximum wavelet coefficients of multi-scale firstly. We then input them to BP classify for different kinds ECG. The experimental result shows that the standard BP neural network classifier's overall pattern recognition rate is well. The ECG in this paper was from MIT-BIH. Experimental result shows that feature vector extracted by the wavelet transform can characterize ECG patterns, and BP neural network classifier has a stronger ECG recognition effect.**

Index Terms—**Electrocardiogram (ECG); Pattern Recognition; BP neural network; Wavelet Transform; MIT-BIH**

I. INTRODUCTION

Since 1903 the Electrocardiogram(ECG) was introduced to clinical medicine, the techniques have been developed rapidly in the record, processing and diagnosis of the ECG whether it is in the biomedical area or in engineering and we accumulate considerable experience. ECG plays an important role in the clinical diagnosis of the heart disease. It provides an objective indicator for correct analysis, diagnosis, treatment and care of the heart disease. Because of its important social value and economic value, it has a very wide range of applications in the modern medicine. ECG is still a major research subject in the biomedical engineering [1].

ECG is the recording of the electrical activity of the heart, and has become one of the most important tools in the diagnosis of heart diseases [2]. ECG signal is shaped by P wave, QRS complex, and T wave. In the normal ECG beat, the main parameters including shape, duration, R-R interval and relationship between P wave, QRS complex, and T wave components are inspected. Any change in these parameters indicates an illness of the heart.

The study of ECG recognition has an important significance in understanding human hear in the role of human intelligence. Although many efforts have been taken recently to recognize ECG using different methods, current recognition systems are not yet advanced enough to be used in realistic applications.

This work is supported by education department Program of Zhejiang Province in University (2010).

II. THE SOURCE OF THE ECG

A. Data Acquisition from the Acquisition Card

Human ECG is a non-linear, non-stationary, weak signal and low signal to noise ratio. The voltage of the conventional ECG is just mv level and its frequency is 0.05-100HZ.

Due to equipment, human factors, the acquisition of ECG accompany with interference. Common ECG interference has three types ①Electrocardiogram (ECG) interference. It is caused by human activities and muscle tension. Its frequency is 5-2000HZ. ECG has wide spectrum and often mixed with the ECG's spectrum. So it is difficult to use the general filtering to separate them. ② The frequency interference of the power-line. It is fixed frequency interference. It is form of 50HZ and its harmonic components. It is caused by the space electromagnetic interference to human body from power-supply network and its equipment .In essence, the alternating current are non-stationary random process in the frequency and amplitude with a slow fluctuation. ③ Baseline drift. It is caused by the low-frequency interference, such as the movement of the measuring electric class, respiratory of the human. Its general frequency is less than 1HZ [3]. Therefore, the collected ECG is often accompanied with the above noise.

B. Recognition Methods --BP neural network

BP neural network is fully named as the Back-Propagation Network, that is, back-propagation network. It is a forward multi-layer network, which uses the error back-propagation algorithm to train the network. BP algorithm [4,5] was proposed by Rumelhart et al in 1986, and since then, due to simple structure, multi-adjustable parameters, much training algorithm and good operational performance, BP neural network got a wide range of practical application.

The network structure of the three-layer BP neural network is shown in Figure 1, from which we can see that, BP neural network contains an input layer, a middle layer (hidden layer) and an output layer. There is a full connectivity between the upper and lower layers and no connections between neurons in each layer. For the input signal, it needs to spread towards to hidden layer nodes and transformed by the function, then transmit the input signal of hidden layer nodes to the output layer nodes. Usually, the transfer function of BP neural network is Sigmoid Type differentiable function, which can achieve arbitrary non-linear mapping between the input and

output, so BP network has been widely applied in pattern recognition, function approximation and other areas [5].

The three nodes of the BP network is represented as: input node x_j, hidden node y_i, output node σ_j, network weight of input node w_{ij}, network weight of hidden node and output node T_{li}, the expectation output the output node t_l, Figure 1 shows the BP neural network structure.

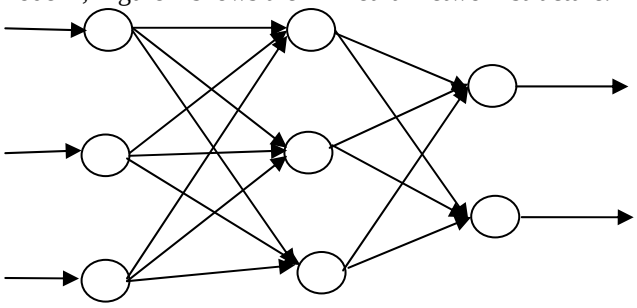


Figure 1. BP neural network structure

The basic learning algorithm of BP neural network:

l) Determine the various learning parameters on the basis of the known network structure, including input layer, hidden layer, output layer neuron number, learning rate and error parameters.

2) Initialize the network weights and thresholds.

3) Provide learning sample: input vector and target vector.

4) Start to learn, and do the following for each sample:

① Forward-calculation of the j unit in the l layer:

$$v_j^{(l)}(n)=\sum_{i=0}^{T} w_{ji}^{l}(n)y_i^{l-1}(n) \tag{1}$$

Equation (7) $y_i^{l-1}(n)$ is the signal transmitted from the i (i=0,set $y_0^{l-1}=-1$, $w_{j0}^{l}(n)=\theta_j^{l}(n)$) unit of the $(l-1)$ layer.

If the j unit activation function is sigmoid function, then

$$y_j^{(l)}(n)=\frac{1}{1+\exp\left(-v_j^{(l)}(n)\right)} \tag{2}$$

and

$$f'\left(v_j(n)\right)=\frac{\partial y_j^{(l)}(n)}{\partial v_j(n)}=y_j(n)\left[1-y_j(n)\right] \tag{3}$$

If the j unit belongs to the first hidden layer $(l=1)$, then

$$y_j^{(0)}=x_j(n) \tag{4}$$

If the j unit belongs to the output layer $(l=L)$, then

$$y_j^{(L)}(n)=O_j(n) \tag{5}$$

$$e_j(n)=d_j(n)-O_j(n) \tag{6}$$

② back-calculation of δ:

For the output units,

$$\delta_j^{(l)}(n)=e_j^{(L)}(n)O_j(n)\left[1-O_j(n)\right] \tag{7}$$

and for the hidden layer units,

$$\delta_j^{(l)}(n)=y_j^{(l)}(n)\left[1-y_j^{(l)}(n)\right]\sum_k \delta_k^{(l+1)}(n)w_{kj}^{(l+1)}(n) \tag{8}$$

③ Fix the right values according to the following:

$$w_{jk}^{(l)}(n+1)=w_{ji}^{(l)}(n)+\eta\delta_j^{(l)}(n)y_i^{(l-1)}(n) \tag{9}$$

5) Enter a new sample until it reaches the error requirement, and the input order of each cycle in training samples needs a re-random order.

The specific program flow chart of training network using BP algorithm shows as fig.2.

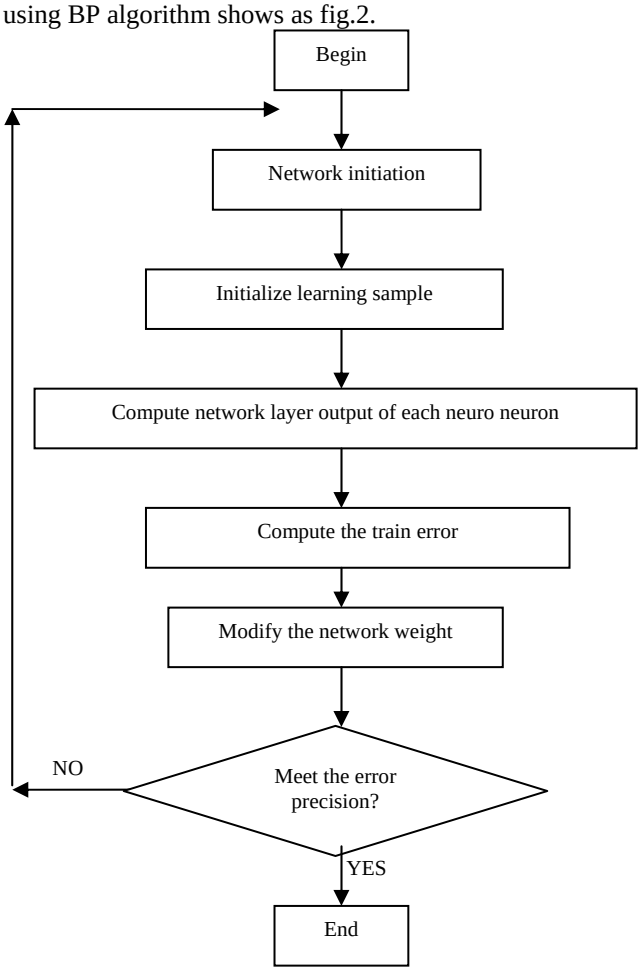


Figure 2. BP Algorithm Flow Chart

III. EXPERIMENT PROCESS

Experiment was carried out in the Matlab7.6 environment. Acquisition of a high-quality database of physiological signals is vital for the emotion recognition algorithm development. An important concern is the selection of signals that are to be used as input to the emotion recognition system. It is desirable that the

influence of emotion on the activity of the nervous system is effectively reflected in the physiological signals employed. Unlike the case of speech recognition or facial expression recognition, where knowledge of the correct class label of a given data point is self-evident, the acquisition of a high-quality physiological signal database with confidence in the underlying emotional status is an intricate task.

The physiological signal data of ECG is from the MIT-BIH, Standard MIT-BIH Arrhythmia database are collected by Beth Israel Hospital Arrhythmia Laboratory from 1975 to 1979. There are more than 4000 data Holtel sets [6]. Database has a total of 48 records, which are from 47 individuals (of which 201 and 202 are from the same individual).

In MIT-BIH database, each record contains data on two channels, which leads settings as follows: The first channel is the use of calibration limb leads Ⅱ; The second channel used correction V1 (occasionally have V2, V5 leads, there is another V4 leads). Data's sampling frequency is 360HZ and the sampling accuracy is 11 bits (sample data range between 0 ~ 2047).

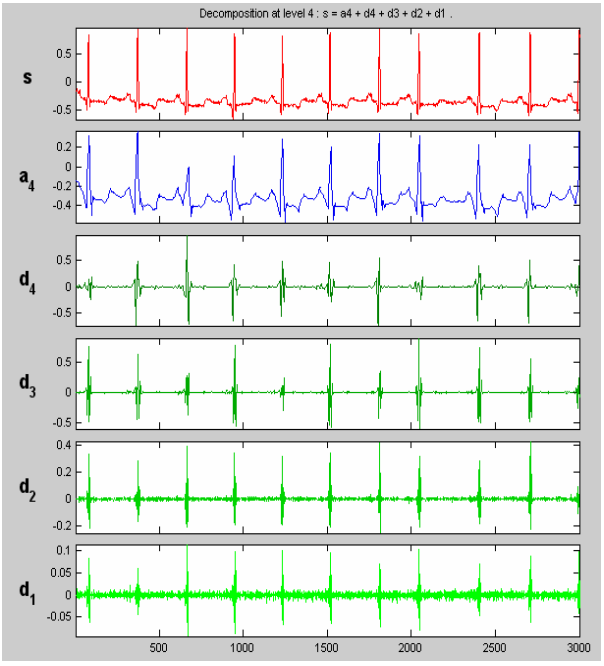


Figure 3 The ECG signals of sample 100 and wavelet transform in four scales

This paper uses a quadratural, compact Daubechies 2 wavelet as the base function for four-scale decomposition of the ECG physiological signal data each day. And extract the maximum values composition vector of each layer in wavelet decomposition as the feature vector of the ECG signal vector, constituting a 4-dimensional feature vector. The ECG signals Waveform of sample 100 and wavelet transform coefficients Wf (a, b) in different scales are shown in fig.3.

Then we extract wavelet coefficients of the signal's maximum value and minimum value. We classify and save four kinds of models of the wavelet coefficients. The features of the ECG can be extracted from the statistical analysis figure, and this paper extracts the maximum value as the feature vector of different pattern of ECG.

Wavelet Transform analyzed ECG parameters are statistically classified as emotional pattern joy, anger, sadness and pleasure. These parameters are then applied to ANN as training and testing data. Also, these parameters are considered as neurons in ANN. The neurons in a feedforward neural network are organized as a layered structure and connected in a strictly feedforward manner. The structure of a basic feedforward neural network is presented in Fig.1. The feedforward neural network is one of the most widely used ANNs. A great number of successful applications of this type of network have been reported [7].

So, after the four-scale decomposition procedure, we get wavelet coefficients of maximum value value of a typical ECG pattern. There are only four representative samples given. Each ECG classification pattern signal is composed by five coefficients. Then decompose the four-scale wavelet to get the 5-dimensional feature vector as the input feature vector of ECG for pattern recognition with method of BP neural network.

Beacuse each line has only one element of each column (under the painted lines are) which is much larger than other elements, which indicates that separation algorithm is very satisfactory.

For describing the separation effect quantitatively, we use the expression $\frac{1}{MN}\sum_{i=1}^{n}\frac{\|s_i - y_i\|_2}{\|s_i\|_2}$ to compute the average relative error of the source signals and separated signals, where N is the number of the source, M is the sampling points). The error curve is showed in Figure 4. We can know the system has large error when starting the train but the error slowly getting smaller and smaller, soon met the desired error and the convergence effect is also obvious in Figure 4.

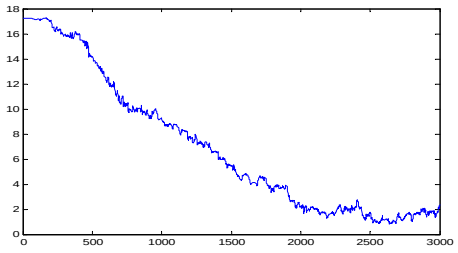


Figure 4. The graph of the error

IV. CONCLUSINO

In this study, a new method concluding wavelet transform and artificial neural network is proposed for classification of ECG arrhythmias. This method includes two phases, which are valued feature extraction and classification. In feature extraction phase of the proposed method, the feature values for each arrhythmia are extracted using discrete wavelet transform (DWT). In classification phase, obtained valued features are used as

input to the artificial neural network for classification of ECG arrhythmias.

Experiment shows that ANN is a valuable tool in the medical field for the development of decision support systems. The actual implementation of ANN analysis of ECG signals involves several stages of varying complexity.

ACKNOWLEDGMENT

We wish to thank our colleagues at Taizhou University for many suggestions and stimulating discussions.

REFERENCES

[1] Sunwen Duo. Twelve-lead ECG acquisition and analysis system [D]. Jilin, Jilin University, 2006：7-7.

[2] Yu, S. N., and Chou, K. T., Integration of independent component analysis and neural networks for ECG beat classification. Expert

Syst. Appl. 34:2841–2846, 2008.

[3] Li An,Li Dong. The Application of Wavelet Analysis inElectrocardiological Signal Digital Filter[J]. Computer Simulation,2001,18(6):70-73(In Chinese)

[4] F. Nasoz, K. Alvarez, C. L. Lisetti, N. Finkelstein: Emotion Recognition from Physiological Signals for Presence Technologies, International Journal of Cognition, Technology and Work, Special Issue on Presence, Vol 6(1), 2003

[5] Xinliang Zhang, Yonghong Tan. The adaptive control using BP neural networks for a nonlinear servo-motor [J]. Journal of Control Theory and Applications, 2008, 6(3): 273-276.

[6] Zhou Qunyi. ECG heart beat Identification Modeling and Analysis of [D]. Zhejiang, Hangzhou : Zhejiang University, 2004：1-1.

[7] K. H. Kim, S. W. Bang, S. R. Kim. Emotion recognition system using short-term monitoring of physiological signals [J]. Med. Biol. Eng. Comput., 2004, 42, 419–427

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 250-253

Property Analysis of Petri Net Reduction

Chuanliang Xia
School of Computer Science and Technology, Shandong Jianzhu University, Jinan, China
State Key Laboratory of Computer Science, Institute of Software, Academy of Sciences, Beijing, China
chuanliang_xia@126.com

***Abstract*—Petri net reduction can avoid the state exploration problem by guaranteeing the correctness. For system specified in Petri nets, this paper proposes reduction methods. A group of sufficient conditions or sufficient and necessary conditions of boundedness preservation and liveness preservation are proposed. These results are useful for studying dynamic properties of Petri nets.**

***Index Terms*—Petri nets, analysis, property preservation, boundedness, liveness**

I. INTRODUCTION

Petri nets are well known for their graphical and analytical capabilities in specification and verification, especially for concurrent systems. Research in reduction methods began with simple pattern modifications on Petri nets [1, 2]. Desel [2] showed that a live and safe FC (free choice) net without frozen tokens can be reduced either to a live and safe marked graph or to a live and safe state machine. A well-known result is the preservation of well-formedness and Commoner's property under the merge of places within a free choice net or an asymmetric-choice net [3]. In order to improve the Petri net based representation for embedded systems verification efficiency, Xia [4] proposed a set of reduction rules, these reduction rules preserve total-equivalence. Murata [5] presented six reduction rules to reduced ordinary Petri nets, these rules preserved livenss, safetyness and boundedness. Sloan [6] introduced a notion of equivalence among time Petri nets, and proved that their reduction rules yield equivalent net. This notion of equivalence guarantees that crucial timing and concurrency properties were preserved. Most reductions are quite specific, such as merging a few places or transitions [7, 8], reducing individual places or transitions [9] or very specific subnets. Esparza [10] provided reduction rules for LTL model-checking of 1-safe Petri nets. Huang [11] proposed some new rules to detect the existence of structural conflicts. Based on Delay Time Petri Net (DTPN), Jiang [12] transformed a Time Petri Net (TPN) component to DTPN model in order to preserve such properties as synchronization, conflict and concurrency during the reduction. In order to improve the analysis efficiency, Shen [13] reduced a large digital system to a simpler one by using three kinds of reduction rules.

This paper investigates one type of transformations and its property-preserving approach for verification. Two kinds of subnet reduction methods are proposed. Conditions of boundedness and liveness preservation of ordinary Petri reduction net are proposed.

This paper is organized as follows. Section Ⅱ presents basic definitions. Section Ⅲ investigates the first reduction method. Section Ⅳ studies the second reduction method. Section Ⅴ gives an example of manufacturing system verification. Section Ⅵ concludes this paper.

II. BASIC DEFINITIONS

A weighted net is denoted by $N=(P,T;F,W)$ where P is a non-empty finite set of places, T is a non-empty finite set of transitions with $P\cap T=\phi$, $F\subseteq(P\times T)\cup(T\times P)$ is a flow relation and W is a weight function defined on the arcs, i.e., $W:F\to\{1,2,3,...\}$. $N_1=(P_1,T_1;F_1,W_1)$ is called a subnet of N if $P_1\subset P$, $T_1=T$, $P_1\neq\phi$, $T_1\neq\phi$, $F_1=(P_1\times T_1)\cup(T_1\times P_1)$ and $W_1=W\mid F_1$, i.e., the restriction of W on F_1.

A Petri net is a couple (N,M_0), where N is a net and M_0 is the initial marking of N. A transition t is said to be live in (N,M_0) iff for any $M\in R(M_0)$, there exists $M'\in R(M)$ such that t can be fired at M'. (N,M_0) is said to be live iff every transition of N is live. A place p is said to be bounded in (N,M_0) iff there exists a constant k such that $M(p)\le k$ for all $M\in R(M_0)$. (N,M_0) is bounded iff every place of N is bounded.

III. THE FIRST REDCTION METHOD

In this section we present the first reduction operation. This operation preserves boundedness and liveness.

Definition 3.1 A net $N_0=(P_0,T_0;F_0,W_0)$ is said to be a p-subnet of $N=(P,T;F,W)$ iff,

(1) N_0 is a subnet of N ,

(2) ${}^\bullet T_0\cup T_0^\bullet\subseteq P_0$,

(3) N_0 is connected, $\{p_i,p_o\}\subseteq P_0$ and p_i is the only input place of N_0, p_o is the only output place of N_0.

Supposition 3.1 A p-subnet satisfies:

(1) p_i is the only place which can contain the initial marking (token(s)).

(2) In a process (tokens from outside flow into p_i, pass N_0 and then flow out from p_o), the number of tokens flowing into p_i is equal to the number of tokens flowing out from p_o.

Definition 3.2 p-subnet reduction operation**:** a reduced net $N'=(P',T';F',W')$ is obtained from original Petri net $N=(P,T;F,W)$ by using $\tilde{p}$ to replace a p-subnet $N_p=(P_p,T_p;F_p,W_p)$, where

(1) $P'=P\cup\{\tilde{p}\}-P_p$,

(2) $T'=T-T_p$,

(3) $F'=F\cup\{(t,\tilde{p})\mid t\in {}^{\bullet}P_i\}\cup\{(\tilde{p},t)\mid t\in P_o^{\bullet}\}$

$-\{(t,p_i)\mid t\in {}^{\bullet}p_i\}-\{(p_o,t)\mid t\in p_o{}^{\bullet}\}-F_p$.

Definition 3.3 (N',M_0') obtained from (N,M_0) by p-subnet reduction operation comprises net N' and marking M_0' where

$$M_0'=\begin{cases}[M_{(P\setminus P_p)0},0] & M_0(p_i)=0\\ [M_{(P\setminus P_p)0},n] & M_0(p_i)=n \quad (n\geq 1)\end{cases}$$

($M_{(P\setminus P_p)}$ is obtained from M by deleted the vector corresponding to P_p**).**

Definition 3.4 A net $(\overline{N}_p,\overline{M}_{po})$ is said to be a closed p-subnet if adding a transition set $T_{tp}=\{t_p\mid t_p$ corresponding to $t\in\tilde{p}^{\bullet}\}$ and arc set $\{(p_o,t_p),(t_p,p_i)\mid t\in T_{tp}\}$ to (N_p,M_{p0}), and preserving the marking of (N_p,M_{p0}).

Note that in this section, let

(N',M_0'): the original net; $N_p=(P_p,T_p;F_p,W_p)$: the p-subnet ; (N_p,M_{p0}) : the p-subnet system; $(\overline{N}_P,\overline{M}_{P0})$: the closed p-subnet system; (N,M_0): the reduced net.

Theorem 3.1 Suppose that (N,M_0) is obtained from (N',M_0') by p-subnet reduction operation. Then (N',M_0') is bounded iff (N,M_0) and $(\overline{N}_p,\overline{M}_{p0})$ are bounded.

Proof. (1) Since (N,M_0) is bound, then $\forall p\in P$, $\exists k_1>0$, such that $M(p)\leq k_1$, $\forall M\in R(M_0)$. Obviously, $\forall p\in P-\{\tilde{p}\}$, $M_{(P\setminus\tilde{p})}(p)\leq k_1$。 Since $(\overline{N}_p,\overline{M}_{p0})$ is bound, then $\forall p\in P_P$, $\exists k_2>0$, such that $M_P(p)\leq k_2$, $\forall M_P\in R(M_{P0})$. Let $k=k_1+k_2$, by Supposition 3.1, $\forall p\in P'$, $M'(p)=[M_{P\setminus\tilde{p}},M_P](p)\leq k$, $\forall M'\in R(M_0')$, so (N',M_0') is bound.

(2) Suppose that (N,M_0) is unbound, then $\exists p\in P$, $\forall k>0$, $\exists M\in R(M_0)$ and $M(p)>k$. By Supposition 3.1 and Definition 3.1-3.4, $\forall k>0$, $\exists M'\in R(M_0')$ and $M'(p)>k$. This contradicts with the fact that (N',M_0') is bounded.

Theorem 3.2 Suppose that (N,M_0) is obtained from (N',M_0') by p-subnet reduction operation. If (N',M_0') is live and $p_i\in\{p\mid(p\in P')\wedge(M_0'(p)>0)\}$, then (N,M_0) and $(\overline{N}_p,\overline{M}_{p0})$ are live.

Theorem 3.3 Suppose that (N,M_0) is obtained from (N',M_0') by p-subnet reduction operation. If (N,M_0) and $(\overline{N}_p,\overline{M}_{p0})$ are live, then (N',M_0') is live.

Ⅳ. The Second Redction Method

In this section we present the second reduction operation. This operation preserves boundedness and liveness.

Definition 4.1 A net $N_0=(P_0,T_0;F_0,W_0)$ is said to be a t-subnet of $N=(P,T;F,W)$ iff,

(1) ${}^{\bullet}P_0\cup P_0^{\bullet}\subseteq T_0$,

(2) N_0 is connected, $\{t_i,t_o\}\subseteq T_0$ and t_i is the only input transition of N_0, t_o is the only output transition of N_0.

Supposition 4.1 A t-subnet system (N_0,M_{t0}) contains t-subnet N_0 and initial marking M_{t0}, satisfy

(1) In a process (tokens from outside flow into t_i, pass N_0 and then flow out from t_o), t_o is fired, iff t_i is fired.

(2)Before t_i is fired and after t_o is fired, $\forall t\in T_0-\{t_i,t_o\}$, t can not be enabled.

(3) If P_0 dose not contain token in initial state, P_0 dose not contain token after a process; If P_0 contains token(s) in initial state, the token(s) will come back to the initial state after a process.

Definition 4.2 t-subnet reduction operation**:** a reduced net $N'=(P',T';F',W')$ is obtained from original Petri net $N=(P,T;F,W)$ by using $\tilde{t}$ to replace a t-subnet $N_t=(P_t,T_t;F_t,W_t)$, where

(1) $P'=P-P_t$;

(2) $T' = T \cup \{\tilde{t}\} - T_t$;

(3) $F' = F \cup \{(p,\tilde{t}) \mid p \in {}^{\bullet} t_i\} \cup \{(\tilde{t}, p) \mid p \in t_o{}^{\bullet}\}$

$$-\{(p,t_i) \mid p \in {}^{\bullet} t_i\} - \{(t_o, p) \mid p \in t_o{}^{\bullet}\} - F_t.$$

Definition 4.3 (N', M_0') obtained from (N, M_0) by t-subnet reduction operation comprises net N' and marking M_0', where $M_0' = M_{(P \setminus P_t)0}$ (where $M_{(P \setminus P_t)}$ is obtained from M by deleted the vector corresponding to P_t).

Definition 4.4 A net $(\overline{N}_t, \overline{M}_{t0})$ is said to be a t-closed net if we add a transition t_t and arcs $\{(p,t_t) \mid p \in t_o{}^{\bullet}\} \cup \{(t_t, p) \mid p \in {}^{\bullet} t_i\}$ to t-subnet (N_t, M_{t0}), and the marking of (N_t, M_{t0}) is preserved. Note that in this section, let

(N', M_0'): the original net; $N_t = (P_t, T_t; F_t, W_t)$: the p-subnet ; (N_t, M_{t0}): the p-subnet system; $(\overline{N}_t, \overline{M}_{t0})$: the closed t-subnet system; (N, M_0): the reduced net.

Theorem 4.1 Suppose that (N, M_0) is obtained from (N', M_0') by t-subnet reduction operation. Then (N', M_0') is bounded iff (N, M_0) and $(\overline{N}_t, \overline{M}_{t0})$ are bounded.

Proof. (1) Since (N, M_0) is bounded, then $\forall p \in P$, $\exists k_1 > 0$ such that $\forall M \in R(M_0)$, $M(p) \le k_1$. Since $(\overline{N}_t, \overline{M}_{t0})$ is bounded, then $\forall p \in P_t$, $\exists k_2 > 0$ such that $\forall M_t \in R(M_{t0})$, $M_t(p) \le k_2$. Let $k = k_1 + k_2$, $\forall p \in P'$, $\forall M' \in R(M_0')$ such that $M'(p) \le k$. So, (N', M_0') is bounded.

(2) Suppose that (N, M_0) is unbounded, then $\exists p \in P$, $\forall k > 0$, $\exists M \in R(M_0)$ such that $M(p) > k$. That is $\forall k > 0$, $\exists M' \in R(M_0')$ such that $M'(p) > k$. This contradicts with the fact that (N', M_0') is bounded.

Theorem 4.2 Suppose that (N, M_0) is obtained from (N', M_0') by t-subnet reduction operation. If (N', M_0') is live and ${}^{\bullet} t_i \subseteq \{p \mid p \in P' \wedge (M'(p) > 0)\}$, then (N, M_0) and $(\overline{N}_t, \overline{M}_{t0})$ are live.

Theorem 4.3 Suppose that (N, M_0) is obtained from (N', M_0') by t-subnet reduction operation. If (N, M_0) and $(\overline{N}_t, \overline{M}_{t0})$ are live, then (N', M_0') is live.

Ⅴ. APPLICATIONS

In this section we apply results of Section Ⅲ and Section Ⅳ to reduce a flexible manufacturing system.

The manufacturing system consists of one workstation WS for assembly work and two machining centers for machining. Machining center_1 and WS share robot R_1. Machining center_2 and WS share robot R_2. The system runs as follows:

In the machining center_1, the intermediate parts are machined by machine M_1. Each part is fixtured to a pallet and loaded into the machine M_1 by robot R_1. After processing, robot R_1 unloads the final product, defixtures it and returns the fixture to M_1.

In the machining center_2, parts are machined first by machine M_3 and then by machine M_4. Each part is automatically fixtured to a pallet and loaded into the machine. After processing, robot R_2 unloads the intermediate part from M_3 into buffer B. At machining station M_4, intermediate parts are automatically loaded into M_4 and processed. When M_4 finishes processing a part, the robot R_2 unloads the final product, defixtures it and returns the fixture to M_3.

When workstation WS is ready to execute the assembly task, it requests robot R_1, robot R_2 and machine M_2 and acquires them if they are available. When the workstation starts an assembly task, it cannot be interrupted until it is completed. When WS completes, it releases the robot R_1 and robot R_2.

Firstly, we give the Petri-net based model of the manufacturing system. Secondly, a reduced net system is obtained by p-subnet reduction method and t-subnet reduction method. Thirdly, we will analysis property preservation of the reduced net system.

The Petri-net based model (N, M_0) of the original manufacturing system is illustrated in Fig.5.1.

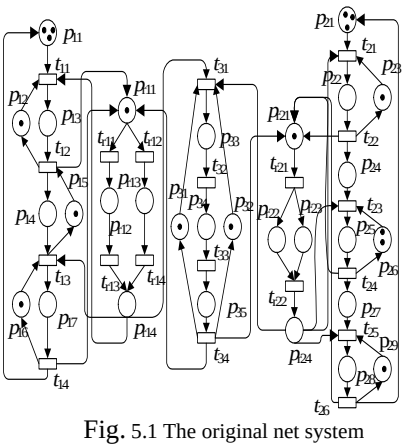


Fig. 5.1 The original net system

The reduction process consists of the following three steps.

Step 1: (N_1, M_1) (Fig. 5.2) is obtained by p-subnet reduction method. By Theorem 3.1-3.3, (N, M_0) is bounded and live, iff (N_1, M_1) is bounded and live.

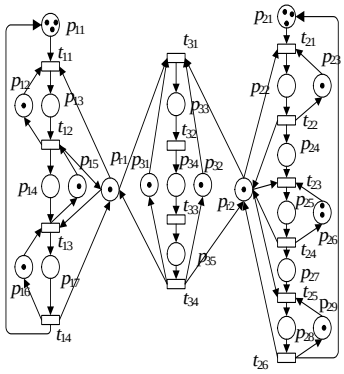


Fig. 5.2 The reduced net (N_1, M_1)

Step 2: (N_2, M_2) (Fig. 5.3) is obtained by t-subnet reduction method. By Theorem 4.1-4.3, (N_1, M_1) is bounded and live, iff (N_2, M_2) is bounded and live.

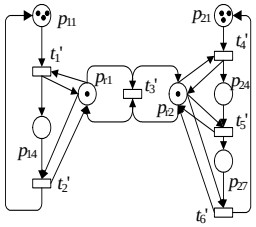


Fig. 5.3 The reduced net (N_2, M_2)

Step 3: (N_3, M_3) (Fig. 5.4) is obtained by deleted places p_{r1} and p_{r2}. Since ${}^\bullet p_{r1} = p_{r1}{}^\bullet$, $M_2(p_{r1}) > 0$ and ${}^\bullet p_{r2} = p_{r2}{}^\bullet$, $M_2(p_{r2}) > 0$, then (N_2, M_2) is bounded and live, iff (N_3, M_3) is bounded and live.

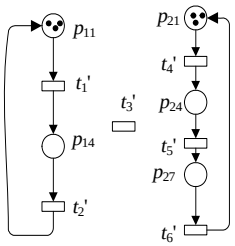


Fig. 5.4 The reduced net (N_3, M_3)

It is easy to see that (N_3, M_3) is bounded, live ([5]). By Theorem 3.1-3.3 and Theorem 4.1-4.3, the original net system (N, M_0) is bounded and live.

VI. Applications

In this paper we investigate property preservations of Petri reduction net. Two Petri net reduction methods are proposed, which are the key methods to ensure the reduced net preserving well behaved properties.

Acknowledgment

This work was financially supported by the National Natural Science Foundation of China under Grant No. 60573012, 60672180 and 60721061, the National Grand Fundamental Research 973 Program of China under Grant No. 2002cb312200, and CAS Key Laboratory of Computer Sciences (SYSKF0903).

References

[1] I. Suzuki, T. Murata, A method for stepwise refinement and abstraction of Petri nets. J. Comput. System Sci. 27 (1983) 51-76.

[2] J. Desel, Reduction and design of well-behaved concurrent systems, Lecture Notes in Comput. Sci. 458 (1990) 166-181.

[3] L. Jiao, T. Y. Cheung, W. M. Lu, On liveness and boundedness of asymmetric choice nets, Theoretical Computer Science 311 (2004) 165-197.

[4] C. Xia, Reduction rules for Petri net based representation for embedded systems, 2008 ATPN's workshop: Protocols Engineering Based on Petri Nets and Modeling of Concurrent Systems, (2008) 16-33.

[5] T. Murata. Petri nets: properties, analysis and applications, Proceedings of the IEEE 77(4):541-580, 1989.

[6] R. H. Sloan, U. Bug, Reduction rules for time Petri nets, Acta Informatica 33 (1996), 687-706.

[7] M. D. Jeng, F. DeCesare, Synthesis using resource control nets for modeling shared-resource systems, IEEE Trans, Robotics Automat. 11(3) (1995) 317-327.

[8] W. M. Mak. Verifying property preservation for component-based software systems (a Petri-net based methodology), Ph. D. Thesis, Department of Computer Science, City University of Hong Kong, June 2001.

[9] J. Esparza. Reduction and synthesis of live and bounded free choice Petri nets, Inform. Comput. 114(1) (1994) 50-87.

[10] J. Esparza, C. Schröter, Net reductions for LTL model-checking, CHARME 2001, LNCS 2144, pp. 310-324.

[11] Y. Huang, H. Wang, P. Yu, Y. Xia, Property-transition-net-based workflow process modeling and verification, Electronic Notes in Theoretical Computer Science 159 (2006) 155-170.

[12] J. Jiang, X. Zhou, Y. Sun, Component-level reduction rules for time Petri nets based on DTPN, Journal of Information and Computing Science, Vol.1 No.1, 2006, pp.37-46.

[13] Victor R. L. Shen, A PN-based approach to the high-level synthesis of digital systems, INTEGRATION, the VLSI journal 39 (2006) 182-204.

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 254-257

An Improved Volumetric Compression Algorithm Based on Histogram Information

Liping Zhao, and Yu Xiang
School of Mathematics and Information Engineering, Jiaxing University, Jiaxing 314001, China
Email: zhaoliping_jian@126.com

***Abstract*—Considering the histogram information can reflect the statistical characteristics of volume data ,an improved volume data compression algorithm based on classified VQ is presented. Firstly, statistic the characteristics of the data itself using histogram technique, and classify the blocks which divided from total volume data into two groups, the blocks with meaningless information as one group(also called empty blocks), and those with meaningful information as the other group(also called object blocks). Secondly object blocks are then compressed by vector quantized. When applying this algorithm to the volume data, all experimental results demonstrate the proposed algorithm can improve the compression rate significantly in the premise of the good quality of reconstruction image.**

***Index Terms*—Vector quantization, Classify, Object blocks, Volume compression**

I. INTRODUCTION

Volume rendering [1] is the process of projecting a volumetric three-dimensional dataset onto a two-dimensional image plane in order to gain some meaningful information about the dataset. For a large volumetric dataset, its real-time volume rendering using hardware acceleration largely depends on and often gets limited by the capacity of graphics memory. Accordingly, CVR (Compressed Volume Rendering) [2-5], with the principle of coupling decompression with rendering, has been proposed and has shown to be an effective approach for solving the just mentioned problem.

Vector quantization [2-5], hereafter abbreviated to VQ, is an ideal choice as an asymmetric coding scheme for CVR. Although the encoding of VQ is complex, its decoding is simple because it is essentially a single table look-up procedure. VQ has first been applied in volume rendering for compression purpose by [2]. A hierarchical VQ method has been described in [3]; it can lead to good fidelity, but it has lower compression rate by about three times compared with the classified VQ method in [2]. In [4], a volumetric compression algorithm has been developed based on transform coding by exploiting Karhunen-Loeve-Transformations and classified VQ, where each block is classified by the contribution of regions in the transform domain to the fidelity of the reconstructed block. However, such an algorithm is rather complicated and time-consuming. We have presented an efficient volumetric compression algorithm called FCHVQ (Flag based Classified Hierarchical VQ) in [5]. FCHVQ adopts a subtle classification strategy and sets flags for each block during the data pre-processing stage. Note that for most volumetric datasets, they are not chaotic, and usually they have a certain percentage of empty regions. Moreover, when the block size is not too large, data in one block are usually highly correlated. As a result, applying FCHVQ to volumetric compression can get good performance.

In the context of CVR, compression rate and decoding speed are the key concerns. Taking into account the fact that histogram information reflects statistics of a volumetric dataset, an improved volumetric compression algorithm is presented in this paper. With the objective of obtaining higher compression rate while at the same time resulting in satisfactory final image, our algorithm classifies the blocks into two groups based on histogram information of the volumetric dataset: empty blocks and object blocks. Most experiments show that more than 60% of blocks are empty blocks. Object blocks are then vector quantized to achieve higher compression rate.

II. RELATED WORKS

Regarding image compression, it should be pointed out that some of the blocks contribute more to the visual quality of the final image, while the others make little contributions. To address this issue, all the blocks need to be classified into different groups, with each group being quantized separately, thereby guaranteeing their presence in the code vector population. Classification of blocks is more important in the case of volumetric compression due to arbitrary nonlinear mapping of the transfer function. In addition, classification of blocks allows the encoder to adapt to a volumetric dataset containing various perceptually important features (e.g., surfaces, boundaries, etc.). Therefore, most literature works [2-5] adopt the idea of classification for the goal of obtaining better performance. In the following we will provide a summary of the relevant background work.

A. Classified VQ

Classified VQ was introduced in [2]. The motivation was to speed up volume rendering of scalar fields through ray tracing. Different from the simplest material model, a shading model sensitive to surface normal was used. A tuple $\{f, n_x, n_y, n_z\}$ was encoded at each voxel, where f stands for the scalar field value, and (n_x, n_y, n_z) is determined by the direction of the field gradient. The classification scheme simply defined two groups of blocks: uniform blocks and active blocks. Uniform blocks, with the size of each being $2\times2\times2$, are those in which

AP-PROC-CS-10CN006

all eight normal vectors are identically zero; all the other blocks thus are active blocks. Separate codewords were designed for either group of blocks and the final codebook was the union of these codewords.

B. Flag Based Classified Hierarchical VQ

FCHVQ (Flag based Classified Hierarchical VQ) [5] aimed at obtaining higher compression rate, faster decoding, and better fidelity. Specifically, a volumetric dataset was first divided into smaller regular blocks and each block was classified according to whether its average gradient value is zero or not. Below we explain the classification strategy of FCHVQ algorithm. Please refer to Fig. I for an illustration of the strategy.

$$\text{Total Blocks} \begin{cases} B_{g=0}, & \text{if } AvH(B_i) = 0 \\ B_{g\neq 0}, & \text{if } AvH(B_i) \neq 0 \end{cases}$$

Figure I. The classification strategy of FCHVQ algorithm

Here $AvH(B_i)$ is the average gradient value of each block, $B_{g=0}$ is the group of blocks that hold zero average gradient value, and $B_{g\neq 0}$ is the group of blocks that hold non-zero average gradient value. Assume that (x, y, z) are the three-dimensional coordinates of a data point, we use $H(x, y, z)$ to represent the gradient value of that point, and $f(x, y, z)$ to denote the original value at that point. We have

$$\begin{aligned} H(x,y,z) &= \left| f(x+1,y,z) - f(x,y,z) \right| \\ &+ \left| f(x,y+1,z) - f(x,y,z) \right| + \left| f(x,y,z+1) - f(x,y,z) \right| \end{aligned} \quad (1)$$

Based on Eq. (1), $AvH(B_i)$ can be defined as in Eq. (2):

$$AvH(B_i) = \frac{1}{n \times n \times n} \sum_{x=x_i}^{x_i+n} \sum_{y=y_i}^{y_i+n} \sum_{z=z_i}^{z_i+n} H(x,y,z) \quad (2)$$

Clearly, if the block with data that they have the same values, it will belongs to the zero average gradient class. While the block with data that they have much different value, it will belongs to the non-zero average gradient class. Because we only save the indices of non-zero blocks, so just $2\times B_{g!=0}$ would be occupied. That's why FCHVQ can reach higher compression rate.

Many experiments show that the gradient values of non-zero blocks always occupy a much more proportion of the total blocks. In the context of this fact, different compression methods to different classes were introduced in order to make FCHVQ algorithm efficient. However, noticing that the blocks that hold the zero gradient value have two situation: one is the blocks that hold the same value *a(a!=0)*, the other is the blocks that hold the same value of 0. Higher compression rate can be reached by using more efficient classify strategy. What's more, the classify strategy presented in FCHVQ would not be coupled to the acceleration techniques of rendering In GPU because of its SIMD architecture.

III. IMPROVED CLASSICAL HIERARCHICAL VQ

In CVR domain , the compression rate and the decode speed are the key concern while applying VQ in volume compression in the premise of the good fidelity. Ning and Hesselink [2] argued that the decompression should allow fast, direct, random access to voxels, and then Fout and Ma[4] recognized two more desired traits: one is compact, separable decompression; the other is uniform decompression. These demands only focus on the rendering speed, for the decompression is coupled to rendering in CVR domain. However, one more important idea should keep in mind is that volume rendering is a method of extracting meaningful information from volumetric data[1].In addition to the decompression restrict ,the characteristics of the volumeric data should be coupled to the compression and rendering. This is even more important when we take the classified VQ for a solution of compression as classification can identify the characteristics of the volumetric data. So, we recognize two more desired traits that specifically target classified VQ.

1. Leave more codewords to express the detail of the image: Classified VQ, generally speaking, can be divided into two classes of methods based on whether each class take the same solution of compression. With the efficiently of today's GPU, bits allocation can be more flexibility, different classes can held their own compression scheme so that leave more codewords to express the detail of the reconstruction image.
2. Classification scheme should be coupled to the acceleration techniques of rendering: The final goal of CVR are faster rendering speed and better image fidelity. To get that goal, Classification scheme should be considered with the acceleration techniques of rendering, such as empty space leaping.

The constraints associated with these objectives put more emphasis on the relations among the volumetric data characteristics 、compression sheme and rendering, not only the relations between compression and rendering. That's to say, before compression, we may extract the meaningful information using certain scheme. By this way, we can get higher compression rate for that the blocks with useless information will be discarded and better reconstruction image quality for that total codewords will leave to save the detail of the image.

Under this background, this paper presents an improved volume data compression algorithm based on classified hierarchical VQ. The procedure of the algorithm is as follows: During the data pre-process stage, statistic the characteristics of the data itself using histogram technique; From the histogram info , classify the blocks which divided from total volume data into two groups, the blocks with meaningless information as one group(also called empty blocks), and those with meaningful information as the other(also called object blocks);Only object blocks are decomposed into a three hierarchical representation manner and vector quantized in order to leave more details of the reconstructed image [3] [5]. The overall algorithm is illustrated in Fig.II.

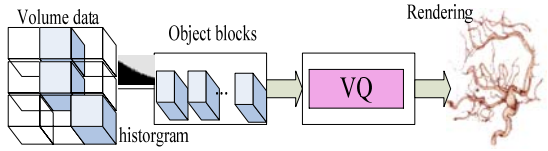


Figure II. Improved classical hierarchical VQ

A. Volume data Pre-processing

A volumetric data set is typically a set V of samples (x, y, z, v) ,also called voxels, representing the value v of the some property of the data, at a 3D location (x, y, z) .If the value is simply a 0 or an integer i within a set I ,with a value of 0 indicating background and the value of i indicating the presence of an object o_i ,then the data is referred to as binary data[1]. So, before using classification scheme, to investigate the characteristic of the volume data is needed. Historgram Info, a simple but useful way , to be used by us to get the distribution of the volume data. Taking the large size of the volume data into account, we firstly resample the volume data in order to reduce the algorithm complexity. Fig III shows the pseudo-code of getting the histogram info.

```
/*!
raw _data:the volume data reading from the raw file
resx,resy,resy: respectively the volume data size in
x,y,z axis
sample_no: resample numble
historgram[i]:the number of data that equals i of the
resample data*/

getHistorgram(raw_data, resx, resy, resz, sample_no,
histogram[256])
begin
   total_sno=sample_no*sample_no*sample_no;
    xstep=resx/sample_no;
    ystep=resy/sample_no;
    zstep=resz/sample_no;
     for   z=0 to resz
         for y=0 to resy
            for x=0 to resx
            begin
                    getDataOfSample(x,y,z);
                    x+=xstep;
                    y+=ystep;
                    z+=zstep;
                end
       for i=0 to total_sno
          begin
                    temp=getDataFromSample(i);
               histogram[temp]++;
          end
```

Figure III. Historgram info of the volume data algorithm

Specifically, the resample number can be user defined based on different volume size, and then uniformly sampling is used on the normalization volume data(0~255), finally get the histogram info of the resample volume data.

From the histogram info of the resample volume data, we focus on the raw volume size、user defined resample volume size、the background size(the total number of data whose value is 0) and the background proportion (equals background size/ resample size).The experimental results lists in table I, where background proportion is shorted for b_proportion.

An important conclusion of the volume data characteristic obtained from table I is that the useful

TABLE I.
CHARACTERISTIC OF THE VOLUME DATA

size	Volume data		
	bonsai	foot	aneurism
Volume size	16777216	16777216	16777216
resample size	262144	262144	262144
background size	207393	1855550	259467
b_proportion	79.11%	70.78%	99%

object information just occupy a little proportion of the volume data while the background proportion takes more proportion. It is not unique, [6] presented that the volumetric data set can be divided into two classes: one class with the empty blocks and the other class with the object blocks. These blocks classify into two classes: one class is empty blocks that the mean value of the block is 0,and the rest blocks is the other class that also called object blocks, So our Classification scheme show as following.

$$\text{Total Blocks} \begin{cases} \text{Empty blocks}, & \text{if } Mean(B_i) = 0 \\ \text{Object blocks}, & \text{if } Mean(B_i) \neq 0 \end{cases}$$

Figure IV.the classification strategy of our proposed algorithm

B. Encoding

The compression scheme of our proposed algorithm is to take different ways to different blocks. Specifically, only object blocks are decomposed into a three hierarchical representation manner and should be vector quantized while the empty blocks are ignored. The hierarchical strategy is the same as the way used in [3,5]. For compassion convenience, the vector quantization of the object blocks also take the splitting scheme based on a principal component analysis[3,5]. One difference should be pointed is that only the small part of object blocks are trained in VQ. Thus, we do not only save a large amount of computation, but also can get better reconstruction image quantity because of leaving the whole code words to the blocks that has the meaningful information of the volume data.

With this compression scheme of our proposed algorithm, we can get higher compression rate. Considering a volume of $N \times M \times K$ data points, each point holds B bytes, the block size of is $n \times n \times n$ and the down-sample block size of hierarchical representation is $(n/2) \times (n/2) \times n/2$. Suppose the number of object blocks

is B_{object}, and the number of empty blocks is B_{empty}, the compression rate can be computed by (3).

$$Rate_{comp} = \frac{N \times M \times K \times B}{C_{flags} + 3 \times B_{object} + C_{codebook}} \quad (3)$$

Here, C_{flags} is the capacity of the flags. The flags of each block which are used for identifying the different class should be stored in capacity of $N \times M \times K / (n \times n \times n \times 8)$ .And $C_{codebook}$ presents the capacity of the codebook.

C. Decoding

The main idea of decoding algorithm of our algorithm is to reconstruct the whole data in each block according to the saved information like FCHVQ. Different from that of FCHVQ, for empty blocks , we just skip that blocks ,while in FCHVQ for those blocks whose average gradient values are zero, we need replace their whole block data with their mean values. Evidently, our method is faster than FCHVQ. What's more, when decompress in GPU, for empty blocks, we just discard that blocks for that these blocks make no contributes to the final reconstructed image. So, acceleration techniques for GPU-based volume rendering [7],for example, empty space leaping can be well used.

D. Results and Comparison

In order to provide a context for the evaluation of our work, we compare our approach with analogous implementations of FCHVQ.

The performance of VQ is measured by the compression rate(Original Data Size/Compressed Data Size) and the reconstructed image quality. The reconstructed image quality is evaluated by the peak signal to noise ratio (PSNR) [8]. Here, the number of codeword in the codebook is 256. The size of volume data bonsai, aneurism and foot is 256×256×256×8 bits. The comparison of the compression rate among different volume data illustrates in Fig.V. ICVQ presents our proposed algorithm.

From Fig.V , we can see that our proposed algorithm can get much higher compression rate than FCHVQ. Especially for aneurism volume data , the compression rate of our proposed algorithm is almost three times more than that of FCHVQ .For the same volume data, we use different codebook size (64,128,256), our proposed algorithm can also get higher compression rate. The PSNR obtained from our proposed algorithm is about 0.1~0.2 higher than that of FCHVQ. Take volume data bonsai for example, the PSNR of the our proposed algorithm and FCHVQ is respectively 36.51 db and 36.36 db. But, our proposed algorithm is a little more time-consuming than that of FCHVQ.

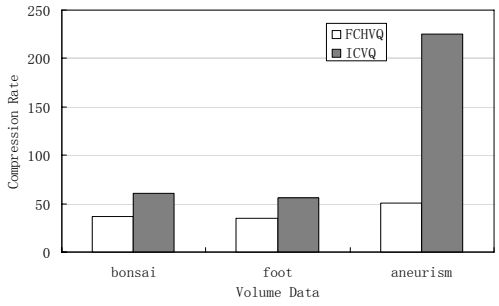


Figure V. Compression rate comparison among volume data

IV CONCLUSION AND FUTURE WORK

Classified Vector quantization has been proved to be an efficient solution for CVR. Noticing that classification scheme should be coupled to the acceleration techniques of rendering because of its SIMD architecture. This paper presents an improved efficient large-scale data compression algorithm, the key to our algorithm is to give full consideration of the characteristics of the volume data by histogram technique and make the classfication sheme more nature. While applying the proposed algorithm to the testing data sets, the experimental results show that our algorithm can not only obtain a better image reconstruction quality, but also increase the compression rate significantly. What's more, our proposed algorithm can be more easier decompress and do rendering on GPU. In the future, we will investigate how to apply our algorithm to the unstructured volume data.

REFERENCES

[1] A. Kaufman and K. Mueller, "Overview of Volume Rendering," chapter for *The Visualization Handbook*, C. Johnson and C. Hansen, Eds., Burlington, MA: Academic Press, 2005

[2] P. Ning and L. Hesselink, "Fast Volume Rendering of Compressed Data," *IEEE Conference on Visualization*, San Jose, CA, 1993

[3] J. Schneider and R. Westermann, "Compression Domain Volume Rendering," IEEE Conference on Visualization, Seattle, WA, 2003

[4] N. Fout and K. L. Ma, "Transform Coding for Hardware-Accelerated Volume Rendering," *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 6, pp. 1600-1607, 2007

[5] L. P. Zhao, D. G. Xiao, and K. L. Li, "An Efficient Algorithm for Large-Scale Volume Data Compression and Its Application in Seismic Data Processing," in Chinese, *Journal of Computer-Aided Design and Computer Graphics*, vol. 21, no. 11, 2009

[6] Tong Xin, Tang Zesheng. 3D texture hardware assisted volume rendering with space leaping. [J]. Chinese Journal of Computers, 1998, 21(9): 807-812

[7] Krüger J, Westermann R. Acceleration techniques for GPU-based volume rendering[C] Proc. of the 14th IEEE Visualization Conf. 2003: 38~43

[8] Linde Y, Buzo A, Gray R M. An algorithm for vector quantizer design[J].IEEE Transactions on Communications,1980,28(1):84-95

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 258-261

An Object Extraction Method Based on Oscillatory Network

Min Li[1], Lizhong Xu[1, 2], and Min Tang[1, 2]
1. College of Computer and Information, Hohai University, Nanjing, China
Email: lm_0711@hhu.edu.cn
2. Institute of Communication and Information System Engineering, Hohai University, Nanjing, China
Email: lzhxu@hhu.edu.cn, hhice@126.com

***Abstract*—Based on locally excitatory globally inhibitory oscillator networks (LEGION), an object extraction method was proposed. The oscillation network can segment image and extracted objects by the scene and object feature. In this paper, the normalized difference water index (NDWI) was used as the feature of water body to segment remote sensing image. Because of the good coherence of spectrum property of water body, the NDWI of the water body would have strong couple weight on NDWI. The special spectrum property of water was also considered as a feature in this paper. Because the proposed method can extract object with their features instead of the pixel intensity, it was robust to noise and advantage to keep the detail information of object. The results of experiment validate the effective of the proposed method.**

***Index Terms*—image segmentation; object extraction; neural network**

I. INTRODUCTION

Image segmentation and object extraction had important significant for Remote sensing. The remote sensing data can be used to analyze, detect and monitor earth surface globally. Especially for the water area, remote sensing data supplied for the macroscopically and nondestructively descriptions. There have been plentiful achievements for remote sensing technology on water quality retrievals, water body extraction, coastline changes detection, the area of water body change detection and so on.

The mostly used methods for extraction of water body were threshold method and multi-bands method. The threshold method was based on the ground object spectra characteristic of every single band, chose an appropriate threshold to extraction the water body. Multi-band method was to construct an index to enlarge the distance between water and other ground objects. Such as spectrum property method [1], normalized difference vegetation index (NDVI) [2], normalized difference water index (NDWI) [3, 4] and improved indices (MNDWI) [4, 5]. Beside above methods, such as decision tree, step iterative method [6-8]and so on. Beside those, the spectrum value of objects is affect by the surrounding environment and noise [9].

Project number: National Natural Science Foundation of China No. 60774092 and No.60901003
Corresponding author: Min Li
Email:lm_0711@hhu.edu.cn

There are mainly four disadvantages of the pixel-wise methods: Firstly this kind of method is depended on the resolution, the luminance condition and the content of the image; secondly the water bodies extracted by the pixel-wise methods are lack of integrity; thirdly it is easy to be disturbed by noise and other ground object; fourthly the result of these methods can't supply the available input to the higher process.

Differing from the pixel-wise method, the object-wise method is based on the macroscopic analysis of the image. At 2000, Baatz M. et al [10] had proposed an local mutual best fitting area increasing strategy which was an object-oriented optimum segmentation method of the remote sensing image. This method had been used in a remote sensing classification software eCongnition. And Gamanya R. et al [11] improved the method of Baatz. M. and proposed an auto-segmentation method. Both of these methods cut the object body from object-level.

This paper would construct the couple weight based on NDWI instead of the luminance of pixels to extract the water body by oscillation network. The couple weight of water body areas should be larger. In order to inhibit other ground objects which have the same spectrum characters of water, this paper also brings the special spectrum property into the oscillation network. Combining these two methods offset the shortages of each other.

II. THE INTRODUCTION OF LEGION

Inspired by biological visual system and brain, Wang and Terman et.al [12] proposed the locally exciting globally inhibition oscillation network (LEGION). This neural model is originated from correlation theory of Von Der Malsburg et al [13-15] proposed at 1994. They assert that the brain can bind the similar features of objects by firing activities the scattered cells from different cortex areas. Won Der Malsburg [13] considered this integrated perceive function of brain as the temporal correlation theory. At 1989 Eckhorn [16] and Gray [17] point out that oscillation correlation is the special form of temporal correlation. The model of LEGION is described as follows:

$$x'_{i,j} = 3 \times x_{i,j} - x^3_{i,j} + 2 - y_{i,j} + I_{i,j} + S_{i,j} + \rho \qquad (1)$$

$$y'_{i,j} = \varepsilon(\alpha(1+\tanh(\frac{x_{i,j}}{\beta})) - y_{i,j}) \quad (2)$$

$$S_{i,j} = \sum_{k,l \in N(i,j)} W_{ij,kl} H(x_{k,l}) + W_p H(p_{i,j} - \theta_p) - W_z H(z - \theta_z) \quad (3)$$

$$P'_{i,j} = (1 - P_{i,j}) H\left[\sum_{k,l \in N_P(i,j)} H(x_{k,l}) - \theta_p\right] - \varepsilon P_{i,j} \quad (4)$$

$$z' = H\left[\sum_{k,l} H(x_{k,l}) - 1\right] - z \quad (5)$$

where $I_{i,j}$ is the stimulate input of pixel (i, j). $x_{i,j}$ and $y_{i,j}$ are two oscillators on different temporal scale. And ρ is the variance of Gaussian noise. $S_{i,j}$ was the couple equation. $W_{ij,kl}$ is the couple weight of pixel (i,j) and pixel (k,l). H is the step function, W_p, W_z are the weight coefficient correspond to $P_{i,j}$, z is the globally inhibition. $P_{i,j}$ is the later potential of pixel (i, j), and θ_p, θ_z are the threshold of $P_{i,j}$ and z.

Equation (1), (2), (3) and (4) are differential equations. The physical significance of these equations had been analyzed particularly and comprehensively by Terman and Wang in [12]. They had proposed a simplified LEGION method in [18]. The couple weight coefficient $W_{ij,kl}$ was the important factor, it decided what feature will be chosen to cut the image.

III. THE LEGION BASE ON NDWI

A. The model of NDWI

The NDWI was defined as follow:

$$NDWI = \frac{Green - NIR}{Green + NIR} \quad (6)$$

Based on the NDWI, the distance between water and other ground objects was enlarged. But because of the pollution and the magnitude of sandiness, the coherence on the water surface will be reduced and the NDWI fluctuated above zero.

Fig.1 shows the spectrum change trend of water area, vegetation area and building area around the river. Like paper [5], the spectrum of building on the forth band (TM4) and the second band (TM2) had the similarity change trend, but with the different mean. Fifteen sample points of NDWI were plotted in Fig. 2. The variance within the water body and the building body were smaller, and the variance between water and building was larger.

B. The couple weight based on NDWI

From the above analysis, construct one couple weight based on NDWI to describe the relationship of pixels:

$$W_{NDWI}(i, j; k, l) = \frac{N(i,j) + N(k,l)}{|N(i,j) - N(k,l)| + \varepsilon} \quad (7)$$

Where $N(i,j)$ was the NDWI value of pixel $I(i, j)$. The equation (2) could be rewritten as follow:

$$S_{i,j} = \sum_{k,l \in N(i,j)} W_{NDWI}(i,j;k,l) H(N_{i,j}) + W_p H(p_{i,j} - 0.5) - W_z H(z - 0.5) \quad (8)$$

▪ *C. LEGION based on NDWI*

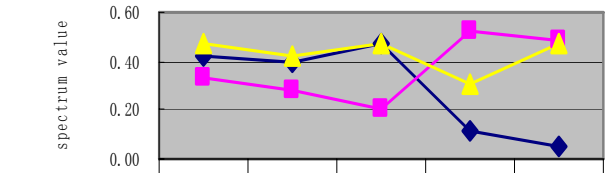
The special spectrum property was introduced into the proposed extraction method which was found by Zhou. The special spectrum property of water is:

$$B_3(green) > B_4(SW) \ \& \ B_2(red) > B_1(IR) \quad (9)$$

So the equation (8) could be rewritten as follow:

$$S_{i,j} = \sum_{k,l \in N(i,j)} W_{NDWI}(i,j;k,l) H(N(i,j)) H((I_3(k,l) + I_2(k,l)) > (I_4(k,l) + I_1(k,l))) + W_p H(p_{i,j} - 0.5) - W_z H(z - 0.5) \quad (10)$$

Where $I_1(k,l)$ represent the spectrum luminance of pixel $I(k,l)$ of the first band. $H(x)$ was the step function, and $H(N(i,j))$ compared the NDWI value of pixel $I(i, j)$ with zero. Adding $H(N(i, j))$ onto the second term was to make the object to be remained which was with the NDWI larger than zero.



	TM1	TM2	TM3	TM4	TM5
water	0.43	0.40	0.47	0.12	0.05
vegetation	0.33	0.28	0.21	0.52	0.48
building	0.48	0.42	0.47	0.31	0.47

Figure.1 The spectrum change trend of water area, vegetation area and building area around the river

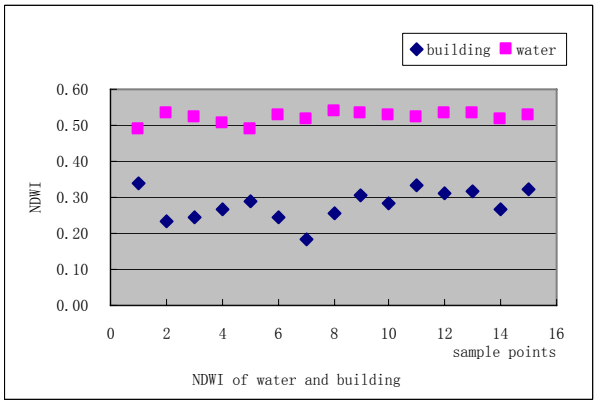


Figure.2 NDWI of water body and building body

IV. EXPERIMENTS AND ANALYSIS

A. Experiment Data

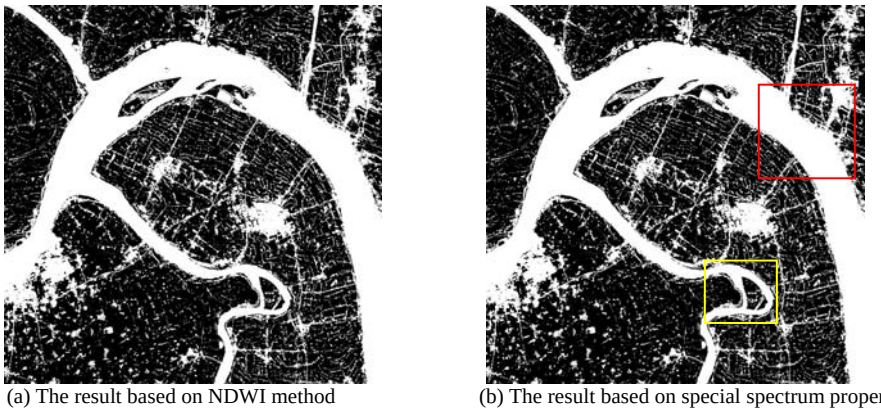
In this paper, we chose the Landsat 7 TM image date to complete the experiments. This remote image contains the area of Tai lake and the part of Chang Jiang River which was cut on the north of the lake.

This cut part contains Yangzhou city and other small towns. The size of image was 800×810 pixels.

Fig. 3 showed the water body extraction results based on NDWI method and the spectrum property. Fig. 4 was the two enlarged parts of original image. This paper used the enlarged part of fifth band (TM5) to show the detail. Fig. 4(a) was the enlarged yellow block. And the Fig. 4(b) was the enlarged red block. Fig.5 shows the results of the LEGION method. Fig.5 could extract the water body with vivid edge on the transition part and the integrity bridge body. At the same time the edge and coherency of Fig. 5(b) was better than Fig. 5(a).

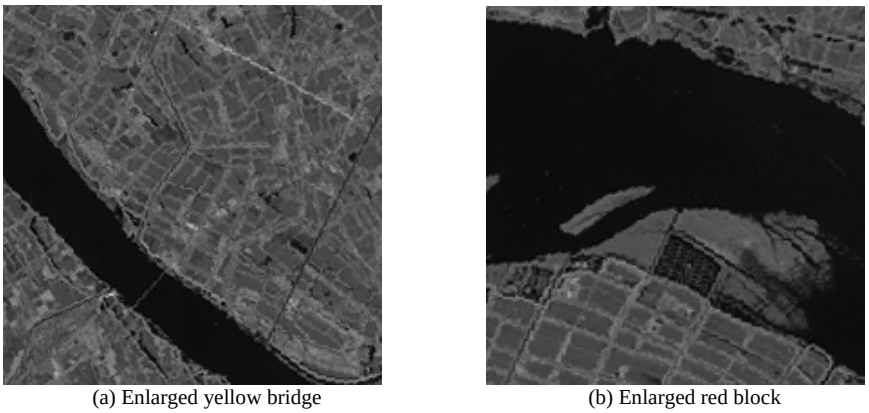
B. The convergence of the proposed Method

The areas of water body were enlarged as the LEGION iterated. This section talked about the convergence of LEGION method based on the river body. Fig. 6 showed the change of increase magnitude between two adjacent iteration times, red line represented the area increase of LEGION method and the blue line represented the area increase of LEGION method with spectrum property. From this figure, 40 times were an optimum iterative time and the areas of water body stay on a stable value..



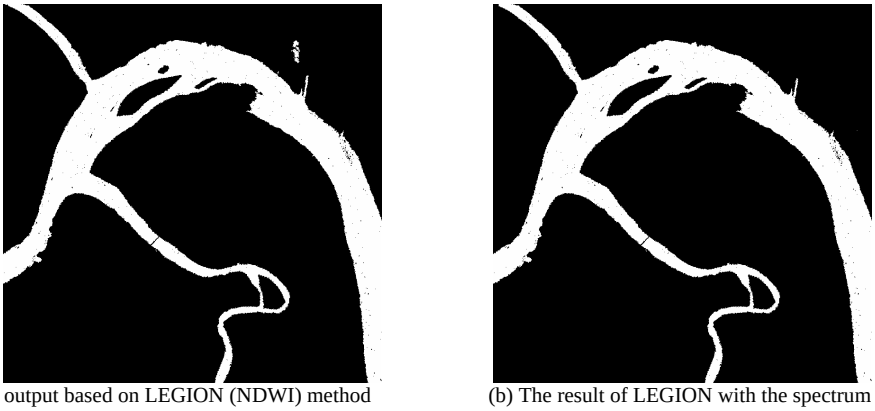
(a) The result based on NDWI method (b) The result based on special spectrum properties

Figure.3 The results based on NDWI method and the spectrum properties $w_z = 30; \theta_p = 1100$, and the iterative time was 80.



(a) Enlarged yellow bridge (b) Enlarged red block

Figure.4 Figure. 4 The enlarged parts of original image



(a) The output based on LEGION (NDWI) method (b) The result of LEGION with the spectrum property

Figure.5 The results of LEGION with NDWI

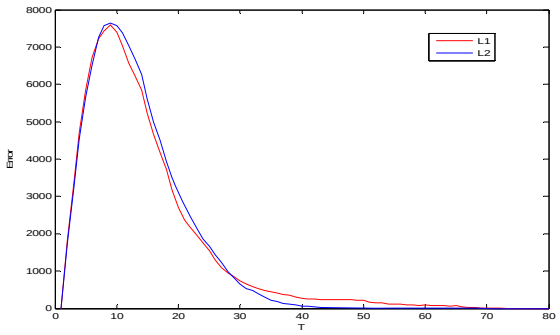


Figure.6The change of areas of every two iterative times

V. CONCLUSION

This paper presents a water body extraction method based on LEGION (NDWI). The experiment results indicated that the proposed method could extract the coherent water body with smooth and clear edge accurately.

Because the oscillation network was dependent on the feature function, it was important to construct an appropriate feature function which can bind the object tightly.

ACKNOWLEDGMENT

This paper was supported by National Natural Science Foundation of China (No. 60774092 and No.60901003).

REFERENCES

[1] Chenhu Zhou and Jiancheng Luo, "*Geomantic understanding and analysis of remote sensing image*," 1999.

[2] Lidong Ding and Hao Wu, "*Study of the water body extracting from MODIS images based on spectrum-photometric method*," Geomatics & Spatial Information Technology, vol. 29, no.006, p.25-27, 2006.

[3] Mcfeeters S. K. "*The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features*," International journal of remote sensing(Print), vol. 17, no. 7, pp.1425-1432, 1996.

[4] Ronglong Cao and Cunjun Li, "*Extracting Miyun reservoir's water area and monitoring its change based on a revised normalized different water index*," Science of Surveying and Mapping, vol.33, no.002, pp.158-160, 2008.

[5] Jinsong Deng and Ke Wang, "*Study on the automatic extraction of water body information from SPOT-5 images using decision tree method*," Journal of Zhejiang University(Agric.& Life Sci.), vol. 31, no.002, pp.171-174, 2005.

[6] Jianchen Luo and Yongwei Sheng, "*High precision auto-extraction of water information of multi-band remote image by step iterative method*," Journal of Remote Sensing, vol.13, no.4, pp. 610-615, 2009.

[7] Jinkang Du and Yongsheng Huang, "*Study on water bodies extraction and classification from SPOT Image*," Journal of Remote Sensing, vol.5, no.003, pp.214-219, 2001.

[8] Townshend J. and Justice C., Li W., "*Global land cover classification by remote sensing-Present capabilities and future possibilities*," Remote Sensing of Environment, vol.35, pp.243-255, Feb.-Mar. 1991.

[9] Baatz M. and Sch P. A., "*Multi-resolution segmentation an optimization approach for high quality multi-scale image segmentation*," pp.12–23, 2000.

[10] Gamanya R. and De Maeyer P., "*An automated satellite image classification design using object-oriented segmentation methods: A move towards standardization*," Expert Systems with Applications, vol.32, no.2, pp.616-624, 2007.

[11] Terman D. and Wang D. L., "*Global competition and local cooperation in a network of neural oscillators*," Physica D: Nonlinear Phenomena, vol.81, no.1-2, pp.148-176,1995.

[12] Von Der Malsburg C., "*The correlation theory of brain function Models of Neural Networks II: Temporal Aspects of Coding and Information*," Processing in Biological Systems, pp.95-119, 1994.

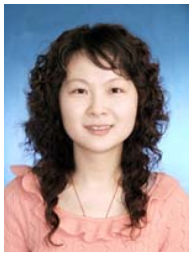
[13] Von Der Malsburg C., "*The what and why of binding: review the modeler's perspective*," Neuron, vol.24, pp.95-104, 1999.

[14] Eggermont J. J., "*The correlative brain: Theory and experiment in neural interaction,* " Springer-Verlag, 1990.

[15] Eckhorn R. and Bauer R., "*Coherent oscillations: A mechanism of feature linking in the visual cortex*" Biological Cybernetics, vol.60, no.2, pp.121-130, 1988.

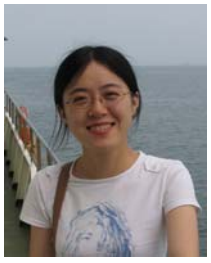
[16] Gray C. and Nig P. M., "*Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties*," Nature, vol.338, no.6213, p.334-337, 1989.

[17] Wang D. L. and Terman D., "*Image segmentation based on oscillatory correlation*," Neural Computation, vol. 9, no. 4, pp. 805-836, 1997.



Min Li is a PhD student of Hohai University, and received the B.S. degree in Communication Engineering from Hohai University, Nanjing, P. R. China in 2005 and 2007 respectively. Her main research interest is remote sensing images processing, bionic vision computation and analysis.

Lizhong Xu is a Professor at College of Computer and Information, as well as the Director of Institute of Communication and Information System Engineering, Hohai University, Nanjing, P. R. China. He received the Ph.D. degree in Control Science and Engineering from China University of Mining and Technology, Xuzhou, P. R. China in 1997. He is a Senior Member of IEEE, Chinese Institute of Electronic, and China Computer Federation. His current research areas include signal processing in remote sensing and remote control, information processing system and its applications, system modeling and system simulation. He is the corresponding author of this paper, his email: lzhxu@hhu.edu.cn, hhice@126.com



Min Tang received PhD degrees from Nanjing University of Science and Technology in 2006. Currently, she is a lecturer at the Hohai University. Her research interests include computer vision and computer graphics.

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 262-266

CNC System of MKS1632 External Cylindrical Grinding Machines Based on Embedded

Xuyan Zhou[1,2], Bin Tan[1], Bing Zhu[1,2], XuanGe Peng[1,2], and Yonghua Pei[3]
[1] School of Information Science and Communication, Jinggangshan University,Jiangxi,Ji'an 343009
zhouxuyan@jgsu.edu.cn
[2] School of Information Engineering, East China Jiaotong University,Jiangxi,Nanchang 330033
Jazhubing@126.com
[3] Jiangxi Jack Machine Co.Jiangxi,Ji'an 343000

***Abstract*—The traditional CNC system based on the IPC is complex structure, high cost, hardware redundancy and can not be cut, high power consumption and low reliability, this paper proposed and designed a CNC System of External Cylindrical Grinding Machines, and gives the software and hardware architecture design.The CNC system operation and management of ARM as the master chip module and PLC as the master motion control module chip, through the I / O modules and servo drives and machine I / O connected to complete the controlling of grinding machine motion .Design and Implementation of Embedded CNC edge Cylindrical Grinder grinding system to overcome the current shortcomings of the system, with versatility, high reliability, system flexibility, cost-effective high merits in line with an open CNC system development needs.**

***Index Term*-embeddedsystem; MMI;ceutter radius compensation**

I. INTRODUCTION

MKS1632 CNC cylindrical grinding machine is a high-end ARM, as well as programmable logic controller as the core of the numerical control system, using three-axis CNC two-axis can be realized into a sports show, radial feed movement precise digital control.Processing, through the man-machine interface, data input processing, use the control panel to issue a variety of operation commands. Programmable logic controller inverter spindle realize stepless speed regulation, in order to meet the requirements of processing technology. To meet the processing technology, designed the right cut, left-cut, continuous processing method, using a radial cut into the descending spiral of feeding method.In order to improve machining accuracy, positioning accuracy, backlash compensation is studied.

DESIGN OF CNC SYSTEM

We designed the CNC system belongs to outline numerical control system, using semi-closed-loop control mode.The numerical control system functional modules shown in Figure 1.

Primarily with the display and manual data entry keyboard, NC panel, as a central display and input unit; machine operator panel for manual machine tools, including processing operation mode, speed, magnification and other settings as well as various operational functions; the central control unit (NC device) used in various components of the installation and connection, the internal module is equipped with numerical control, human-computer control module, programmable logic controller (PLC), network communication module, monitoring circuit.

All the features of the CNC system consists of two microprocessors CNC-CPU and the PLC-CPU and the serial communication interface-sharing, communication between the various parts of the completion by the internal bus.CNC-CPU for each channel of data preprocessing and interpolation calculations. Here, each one channel is a stand-alone system, with multiple axes and independent of the CNC machining procedure that requires a corresponding pre-processing procedures and the interpolation processing, also asked the various channels can be parallel processing of NC processing program, each channel by the PLC is responsible for the coordination. PLC-CPU instruction by receiving T and M switch command, a variety of electrical machine control, machine control panel is also responsible for handling input.

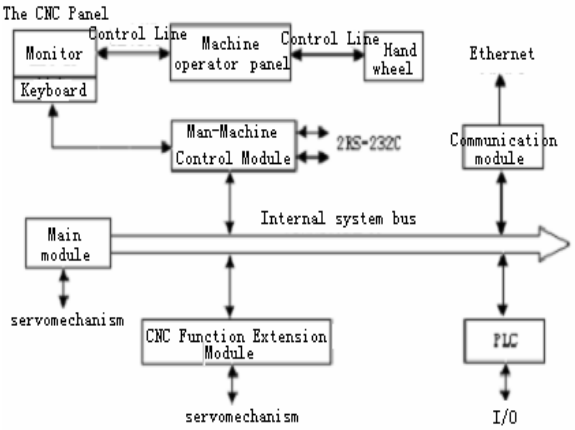


Figure1. DIAGRAM OF NUMERICAL CONTROL SYSTEM LOGIC BLOCK

A. ARM Chip

We use ARM9 S3C2410A microprocessor operation and management of master-chip modules. S3C2410A microprocessor is a launch by Samsung's 16/32-bit RISC processor, it uses the ARM920T core, 0.18um standard CMOS process of macro-cell and storage unit, it also uses a technique called Advanced Microcontroller Bus

Architecture (AMBA) a new bus architecture, and its low power consumption, streamlining and excellent all-static design is particularly suitable for cost-and power-sensitive applications. In order to reduce the total cost of the entire system and reduce external components, S3C2410A provides the following extensive internal equipment: separate 16KB Instruction Cache and 16KB of data Cache, MMU virtual memory management, LCD controller (supports STN & TFT), NAND Flash controller, System Manager (chip select logic and SDRAM controller), 3-channel UART, 4-channel DMA, 4-channel PWM timer, general-purpose I / O ports, RTC, 8-channel 10-bit ADC and touch screen interface, IIC-BUS interface, IIS -BUS interface, USB host, USB device, SD master card & MMC card interface, 2-channel SPI and the internal PLL clock multiplier[3].

Programmable Logic Controller

Programmable Logic Controller PC (Programmable Controller) also known as programmable logic controller, PLC (Programmable Logic Controller), is a digital computing operation of an electronic system designed for industrial environment applications.

PLC model of a wide range, which is mainly by the central processing unit CPU, memory, input, output and other components. We use the Mitsubishi FX2N series produced in micro-programmable logic controller, according to the input and output points, the number of selected FX2N-64MT, as shown in Figure 2. When an external contact is closed, + 24V voltage is added to the receiver circuit, after filtering and level conversion processing, the output to the PLC internal, the internal electronic circuit can receive and process signals.Receiving circuit coupled with photoelectric conversion circuit, the switch input signal is machine-side switches, buttons, relay contacts, testing sensors acquisition closed / disconnected state of the signal. The state signal is subject to photoelectric conversion circuit coupled processing, in order to become a PLC can receive signals. When using an active switch period (such as non-contact switch, Hall switches, etc.) must be used DC24V specifications, test components, NPN / PNP model must be consistent PLC interface electrical specifications.PLC output circuit is generally relay output, transistor output, thyristor output, maximum output current is typically 100 mA or so, can drive light-emitting diodes, a small relay coils, etc., in addition to be able, through relays, thyristor and so enlarged re-exported to the external interface. Relay and thyristor outputs a larger load capacity can reach 2 A above, to drive solenoid valve and AC contactor coil and so on.

Figure 2. FX2N-64MT Appearance

Touch Screen

Man-machine interface is in operation personnel and equipment for two-way communication between the bridge, the user can freely combination of text, buttons, graphics, digital, etc. to deal with or monitor the management and change information at any time to cope with multi-function display screen. The use of man-machine interface allows the machine wiring standardization, simplification, but also can reduce the required PLC controller, I / O points. Reduce production costs at the same time, due to miniaturization of the control panel and high-performance, increase the added value of the entire device. As a new touch-screen human-machine interface, from the appearance of a concern, it's easy to use, powerful features and excellent stability make it ideally suited for industrial environments, and even can be used in daily life, the application a very broad [8]. Figure 3 for the MT500 series touch screen.

Figure3. MT500 SERIES TOUCH SCREEN

MT500 series touch screen is designed for PLC applications, it is different from a few simple instrument, or some other simple control of PLC equipment, its powerful, easy to use, suitable for modern industry is growing workload and functional requirements [9].MT500 series touch screen COM1 and COM2 can be connected with the PLC, if it is to use COM2 connection, wiring shown in Figure 4.

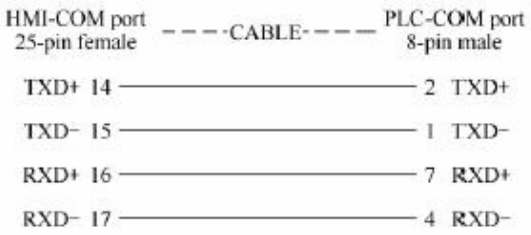


Figure4. COM2 and PLC wiring diagram

Spindle driving and control system

CNC machine tool spindle drive system has a DC spindle drive system and the AC spindle drive system. With the micro-electronics technology, communication theory and speed the development of high-power semiconductor technology, communication-driven into the use phase.In this paper, Cage AC Motor Frequency transform configuration vector spindle drive system. This is because the cage AC motors DC motors do not have mechanical trouble and change to bring in high-speed, high-power to the restrictions imposed, while the other hand, the performance of AC drive has reached the level

of DC-driven, and the AC motor with a small size, light weight, full-enclosed work and so on. Here we focus on talking about AC spindle drive unit with the spindle axis associated with the feed control.

1) AC spindle drive unit

This feed AC servo motor using permanent magnet synchronous motor, but the spindle cage AC motor uses AC motors, it is because the CNC machine tool numerical control machine tool spindle drive system without the feed drive system, as did need such a high dynamic performance and transfer speed range. We use the vector control algorithm, the basic idea is through a complex coordinate transformation to an equivalent AC motor into a DC motor and controlled. Using this algorithm, AC motors and DC motors very similar to the mathematical model, which can be equally good speed performance[4,9].

AC spindle drive unit from the grid-side inverter, control regulator, the load-side inverter, microprocessor controller, such as the composition of components.Shown as Figure 5.

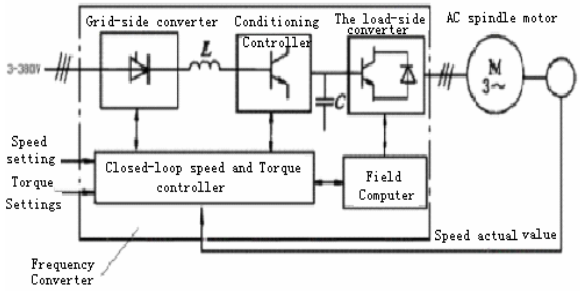


Figure5. AC SPINDLE DRIVE UNIT STRUCTURE

Spindle motor operating characteristics as shown in Figure 6.

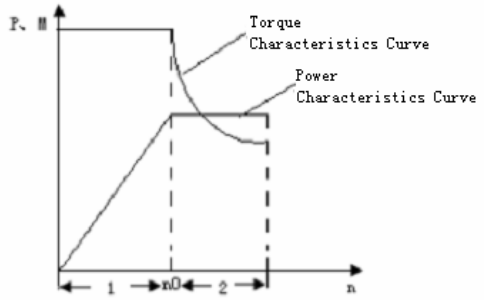


Figure6. SPINDLE MOTOR OPERATING CHARACTERISTIC CURVE

Figure 6, M is torque, P is power, n to the base speed. The following belong in the base speed n0 constant speed governor, by changing the armature voltage methods. Its speed formula is:

$$n = \frac{U - I_a R}{C_e \Phi} \quad (1)$$

$$\Phi = K I_f \quad (2)$$

Which, n_0 following excitation current I_f the same, changing the armature voltage U speed, the output depends on the maximum torque Mmax maximum armature current Imax, the maximum current time due to a constant, and therefore the maximum torque output is constant , but the output power will increase as the speed increases, so n_0 hereinafter referred to as constant torque speed.

N_0 above the base speed up adoption of a weak magnetic method of speed governor, that is, by adjusting the I_f the method, the output of the maximum torque Mmax as follows:

$$M_{\max} = C_M \Phi I_{\max} \quad (3)$$

Speed in the weak magnetic l, $I_{f\text{-fold}}$ reduction of K, rotational speed will be a corresponding increase in K times the output of the maximum torque motor is because the magnetic flux Reduction in K times the maximum power output remains unchanged, thereby achieving the constant power speed.

2) Spindle axis associated with the feed control

a) SPC

In the spindle and feed axis control to be associated with the use of the pulse encoder, commonly used can be divided into incremental and absolute type, contact and contactless card types. We use incremental pulse encoder because the encoder with high precision, simple structure, reliable.

Incremental pulse encoder is generally 90 degrees out of phase two of the A, B signal and a zero-bit C signal, in which A, B signal can either be used to calculate the size of angular displacement at the same time use them to lead or lag phase They can also identify the relative direction of rotation, then the B signal ahead of A signals that fat transfer; C signal for each transponder can serve as a zero signal, the signal for the A or B signals, several technical benchmarks[4].

In addition, incremental pulse can also be used for the preparation of spindle speed measurement, but the measured maximum speed n max single pulse width limited by its corresponding formula is as follows:

$$n_{\max} = \frac{60}{NT_0} \quad (4)$$

Where, n_{max} is the maximum measured pulse encoder speed (r / min); N for pulse encoder pulses per revolution generated by the number of (r); T0 for a single pulse width (s).

b) spindle rotation and axial feed of the Association Control

We have installed on the spindle to detect the spindle encoder pulses corner, phase, zero-signal, the spindle rotation process, linked with the pulse encoder continuously send pulses sent to CNC devices, according to the results of interpolation calculations, the control into the to the axis servo system that allows feed rate and spindle speed to maintain the ratio between required.

c) spindle rotation and radial feed of the Association Control

In order to ensure the smooth processing of end-formation, it is necessary to make the surface of the surface roughness values of less than or equal to a set value, which must ensure that the workpiece with the wheel contact point at the speed of a constant speed. As the grinding face, we should constantly be feeding action so that grinding diameter gradually reduced. Therefore, the NC device control software must be designed to complete the corresponding spindle speed adjustment.

II. SOFTWARE DESIGN

Software architecture with front and back structural model.Foreground program is an interrupt service routine, it is almost assumed, the entire real-time functions (interpolation, position control, etc.), while the scheduling management, and human-computer interaction in the background process to complete[6]. Background process is a loop run the program, in its operation process, the prospects continue to insert real-time interrupt program worked together to accomplish the task CNC machining. Their operational tasks, including system management and system control two parts. System Management include communications, display, syntax checking, parameter setting, fault diagnosis, procedures, coding and other tasks, such procedures are less demanding real-time; system control is mainly motion control, including, tool compensation, speed control, trajectory interpolation , position control, limit and emergency stop and other tasks, such procedures are completed very strong real-time control tasks. The overall structure of the software system as shown in Figure 7.

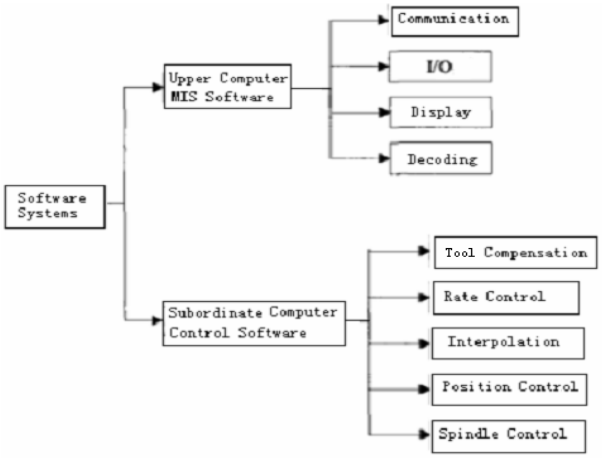


Figure7. BEFORE AND AFTER THE TAIWAN-BASED SYSTEM SOFTWARE ARCHITECTURE

III. INTERPOLATION ALGORITHM

Parts of the contour shape is from a variety of shapes (such as linear, circular, spiral lines, parabola, the free curve, etc.) of different circles. Foremost among these is a straight line and arc. Users in parts processing, in general, provides only necessary to describe the linear parameters, such as a straight line, to provide its starting point and end point; pairs of arcs, and to provide a starting point and end point, clockwise or counter-clockwise circular arc and the center of a circle relative at the start position. Therefore, in order to achieve the process in motion trajectory control must be calculated in real-time linear and the feed rate to meet the requirements of a number of intermediate points (between the start and end). This is the CNC interpolation techniques (Inierpolation) concept. Accordingly, interpolation is defined as follows: The so-called interpolation is based on a given feed rate and a given linear outline the requirements of the outline between the known points to determine the number of mid-point method, this method is called interpolation fill methods or interpolation theory. For each method (principles) can be used on a different calculation method to achieve such a specific method of calculation is called interpolation algorithm.

Interpolation algorithm can be divided into pulse incremental interpolation algorithms and data sampling interpolation algorithm for two. Pulse incremental method is generally used with stepper motor-driven devices, the open-loop numerical control system, it is common pulse incremental method with point by point comparison method and the digital integration method. Data sampling method has been applied relatively high speed and accuracy requirements of NC system, the common data sampling time-division method to expand DDA method and so on[8].

The system uses digital integration method of circular interpolation algorithm interpolation method.

digital integration method for the basic principles

From the geometric point of view, is to find the function of integral operator Y = f (t) curve and the horizontal axis of the enclosed area. As shown in Figure 8.

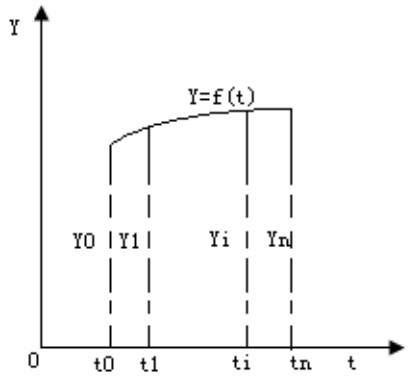


Figure8. GEOMETRIC DESCRIPTION OF THE FUNCTION POINTS

We can see from Figure 8, from the time t0 to tn the function Y = f (t) the integral value is expressed as:

$$S = \int_{t0}^{t_n} Ydt = \int_{t_0}^{t_n} f(t)dt \tag{5}$$

In geometry is to use a series of small rectangular area, and to approximate that of the function f (t) the following integral area. 5 for further processing of the formula, take Δ t as the basic unit of one, then the formula is converted to digital integrator formula:

$$S = \sum_{i=0}^{n-1} Y_i \tag{6}$$

If we set accumulator capacity of a unit area value in the accumulator during the more than one area of the value of accumulator generates a pulse, then the number of pulses generated by the demand shall be the total area.

digital integration method for circular interpolation

To I-quadrant inverse round NR1 explain DDA circular interpolation of the basic principles of law, as shown in Figure 9.

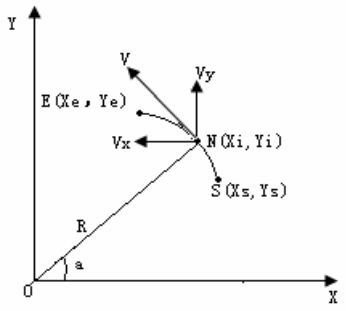


Figure9. DDA CIRCULAR INTERPOLATION THEORY

When the cutting tool along the arc SE , arc radius R, cutting speed of v, the speed of the two axis components of Vx and Vy, fixed point and N (Xi, Yi), based on the similarity relations between triangles can be very :

$$\frac{v}{R} = \frac{v_x}{Y_i} = \frac{v_y}{X_i} = K(\text{constant}) \tag{7}$$

Where, v = KR; Vx = KY i; Vy = KXi.

As the radius R is a constant, if tangential velocity v is uniform, then K is a constant, then the fixed point in two coordinate axis velocity component will vary and change the value of its coordinates.

When given a time increment Δ t, the fixed point in X, Y axis displacement increments are as follows:

$$\Delta X_i = -v_X \Delta t = -KY_i \Delta t \tag{8}$$

$$\Delta Y_i = v_Y \Delta t = KX_i \Delta t \tag{9}$$

Due to against the X-axis coordinate values corresponding to a gradual decrease in Quadrant I, so the formula 8, Δ X i expression to take a negative sign, that is, v X and v Y are taking the absolute value. 8 and 9 according to the formula I obtained the first quadrant inverse circular arc DDA Interpolation formula is:

$$X = \sum_{i=1}^{n} \Delta x_i = -\sum_{i=1}^{n} ky_i \Delta t_i \tag{10}$$

$$Y = \sum_{i=1}^{n} \Delta y_i = -\sum_{i=1}^{n} kx_i \Delta t_i \tag{11}$$

IV. SUMMARY

This system combined CNC technology and embedded system, designed and implemented the ARM + PLC embedded face cylindrical grinding machine numerical control system hardware&software systems, embedded numerical control system has proven superior performance of the system resources can be configured by software and hardware reduction, access to non-redundant structure, highly-integrated numerical control system.Based on this system face cylindrical grinding machine control system, with versatility, high performance, reliability, system flexibility, cost-effective high advantages, greatly improving the face cylindrical grinding machine automation and efficiency.

REFERENCES

[1] ZHOU Zude,LONG Yihong,LIU Quan. Embedded-Based network numerical control technology and system [J]. CHINESE JOURNAL OF MECHANICAL ENGINEERING，2007.43(5):1-6

[2] ShibataT，YanoM， Kodama F. EmPirical analysis of evolution of Product architecture Fanue numerical controllers from1962 to1997[J]. Research Poliey. 2005. l.34:13-31

[3] WANG Dianhong,LI Weizhong,LIU Bing . Design of Control System Based on ARM Processor [J]. MANUFACTURING TECHNOLOGY & MACHINE TOOL，2007(12):31 一33

[4] ZhangJinHuan . Structure of embedded digital control system and research on optimization of PID parameters of its servo. WuHan :School of Information Engineering, Wuhan University of Technology，2006

[5] Gloria AlessandroDe.MieroProeessor design for embedded system[J]. Journal of Systems Arehiteetufe ， 1999.45:1139-1149

[6] Jean.J.Labrosse. Embedded Real-Time Operating System uc/os -Ⅱ [M]. ShaoBeibei. BeiJing :Beijing University of Aeronautics and Astronautics Press.2003:283-316.

[7] LiFang，LiuKai，Wang TongTong.Implementation of Chinese Characters Display Module in Embedded CNC System Based on ARM[J].MANUFACTURING INFORMATION ENGINEERING OF CHINA ， 2008(9):53-55

[8] Liao Xiao Guo.Zhu Qi Shu.Computer numerical control.WuHan : Huazhong University of Science.1999:78-95.

[9] Liu BaoYan, ChengShuKang .Stepper motor and its drive control system.Harbin: Harbin Institute of Technology Press.1997:202-216.

[10] Stylo， AW/Diana， GA low cost， high performance PC based integrated Real-time motion control development system.IEEE International Symposium on Industrial Eleetronies.Proeeedings.ISIE， 98.1998.Vol.1.PP127-130.

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 267-270

Parabola Interpolation With Adaptive Error Compensation

Guangming Yang[1], and Fengqi Yu[2]
[1, 2] Department of Integrated Electronics，Shenzhen Institute of Advanced Technology, CAS
Shenzhen, China, 518067
[1]Email: mark111yang@hotmail.com
[2]Email: fq.yu@siat.ac.cn

***Abstract*—This paper proposes a novel scheme to interpolate images based on Lagrange Interpolation Theory. Parabola polynomial with an adaptive Lagrange error compensation is adopted to determine 1-D pixels. In order to predict 2-D pixels more accurately, the direction perpendicular to gradient vector is determined. To simplify the interpolation process, we give a range of the direction. Simulation results show that the proposed method can get clearer and sharper image than traditional methods. From objective point of view, average gradient and mean structural similarity (MSSIM) are calculated to demonstrate its superior to other methods.**

***Index Terms*-parabola interpolation, Lagrange interpolation error, Sobel operator, gradient direction , MSSIM.**

I. INTRODUCTION

Image interpolation is a method to improve resolution of an image. It is widely applied in outer space image, medical image, and images in consumer electronics, etc. There are many image interpolation techniques, e.g. bilinear, cubic convolution [1], and bicubic B-spline interpolators. However, These techniques suffer from artifacts such as zigzag, ringing and blurriness. Especially the blurriness artifact blurs image detail so badly that we can't do any further image analysis.

Efforts have been made by many people to improve the image quality, Lei Zhang [6] proposed an interpolation algorithm based on directional filtering and data fusion, which can preserve the edge information to a certain extent. Schultz proposed MAP (maximum a posteriori) algorithm to expand image [4], which places the interpolation problem into a statistical framework. For the purpose of reducing the interpolation error, EASE method was proposed by Cha [2]. In his scheme, based on bilinear method, Lagrange interpolation error theory is used to compensate the pixel error in 1-D signal, which makes the pixel values more accurate. However, Zhang's algorithm and MAP require more computations. Although EASE applies error compensation to interpolation successfully, the quality of the interpolated image cannot meet some requirements because it is based on biline method in 1-D.

To further improve image resolution, we propose a parabola interpolation method which interpolates the 1-D direction pixels with a parabola interpolation polynomial, where an error compensation technique is considered. Then based on bilinear method, 2-D pixels along the direction perpendicular to gradient direction are interpolated. The gradient direction is calculated by vertical and horizontal Sobel operators.

This paper is organized as follows: Section Ⅱ briefly describes the parabola interpolation and the error estimation theory. Section Ⅲ presents the proposed algorithm of 1-D and 2-D interpolation. Experiment results and performance analysis are presented in section Ⅳ, where average gradient and MSSIM quality assessment methods are adopted from objective point of view. A brief conclusion is presented in Section Ⅴ.

II. REVIEW OF PARABOLA INTERPOLATION AND ERROR ESTIMATION

As a special part of Lagrange interpolation, parabola interpolation polynomial $L_2(x)$ can be formulated as [5]:

$$L_2(x)=y_{k-1}\frac{(x-x_k)(x-x_{k+1})}{(x_{k-1}-x_k)(x_{k-1}-x_{k+1})}+y_k\frac{(x-x_{k-1})(x-x_{k+1})}{(x_k-x_{k-1})(x_k-x_{k+1})}+y_{k+1}\frac{(x-x_{k-1})(x-x_k)}{(x_{k+1}-x_{k-1})(x_{k+1}-x_k)} \quad (1)$$

where each variable is interpreted in figure 1.

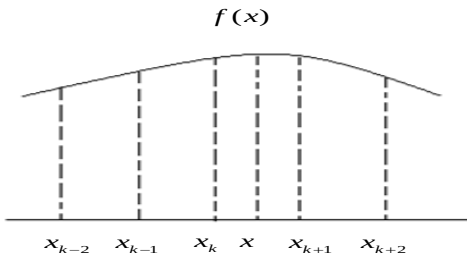


Figure 1. Interpolation sketch map

The value of x is estimated by the surrounding elements with corresponding weights. The interpolation error is discussed in ref. [5].

Assume $f^{(n)}(x)$ is continuous in the range $x\in[x_0, x_1]$, and $f^{(n+1)}(x)$ exists when $x\in(a, b)$, $L_n(x)$is the interpolation polynomial. The interpolation error is:

$$R_n(x) = f(x) - L_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}\omega_{n+1}(x) \quad (2)$$

where $\xi \in [a, b]$, and

$$\omega_{n+1}(x) = (x - x_0)(x - x_1)...(x - x_n) \quad (3)$$

In fact, the *n+1* order derivative of image function *f*(x) is unavailable and it has to be replaced. It is well known that adjacent pixels always have similar property and errors are closely related to each other. Based on this observation, we propose an error estimation which will be discussed in section III.

III. THE PROPOSED INTERPOLATION ALGORITHM

For a typical image expansion, for example the magnification factor is *2x2*, an image of *NxM* pixels is expanded to high resolution *2Nx2M*. We can get arbitrary high resolution image by changing polynomial coefficient in (1). Two steps are processed to implement 1-D and 2-D pixels interpolation respectively. As shown in figure 2, the □ denotes pixels of original low resolution image, O and ⊗ are unknown pixels which are called 1-D and 2-D signal respectively.

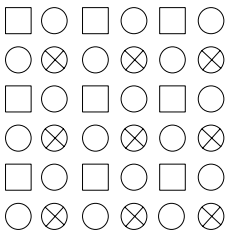


Figure 2. Distribution of 1-D and 2-D pixels

A. 1-D pixels interpolation

First, the interpolation of the O position in figure. 2 can be considered as 1-D interpolation of the pixel between x_k and x_{k+1}, which is denoted as *x*, as shown in figure 1. Because the magnification factor is *2x2*, the interpolation problem is converted to figure out the value of $x_{k+1/2}$. According to the values of x_k and x_{k+1}, the interpolation method is as follows. If $y_k = y_{k+1}$, we directly let $y_{k+1/2} = y_k$, without considering error compensation. Otherwise we use parabola interpolation introduced in previous section and an error is compensation which will be discussed shortly.

For 1-D interpolation, using (1), $L_2(x_{k+1/2})$ can be obtained

$$L_2(x_{k+1/2}) = y_{k-1} * a + y_k * b + y_{k+1} * c \quad (4)$$

According to the weight term in (1), we can determine the coefficients *a*=-(1/8), *b*=3/4, *c*=3/8. These coefficients take the influence of each adjacent pixel into consideration. Using equation (2), the interpolation error can be expressed as

$$error\_x_{k+1/2} = \frac{f^3(\xi)}{6}\omega_3(x_{k+1/2}) \quad (5)$$

From (3) and (5), we have:

$$error\_x_{k+1/2} = \frac{f^3(\xi)}{6}(-\frac{3}{8}) \quad (6)$$

As said in section II, we must find another way to get $error\_x_{k+1/2}$ since $f^{(3)}(\xi)$ cannot be found.

Using (1), the estimated error of x_k can be expressed as :

$$error\_forward = f(x_k) - L_2(x_k)$$
$$= f(x_k) - y_{k-2} * d + y_{k-1} * e + y_{k+1} * f \quad (7)$$

where *d*=-(1/3), *e*=1, *f*=1/3. From (2) we can also get another error expression

$$error\_forward = \frac{f^3(\xi 1)}{6}\omega_3(x_k) \quad (8)$$

When $\xi \in [x_{k-2}, x_{k+2}]$, we can assume that $f^3(\xi)$ varies little. Then:

$$error\_forward = \frac{f^3(\xi)}{6}(-2) \quad (9)$$

So far, an obvious relation between $error\_x_{k+1/2}$ and *error_forward* is available from (6) and (9):

$$error\_x_{k+1/2} = (3/16) * error\_forward \quad (10)$$

The same method can be applied to get the error of y_{k+1}. In addition we have the following equation:

$$error\_x_{k+1/2} = (3/16) * error\_back \quad (11)$$

We have two error evaluations which are adaptively selected. The reason is the following. If $y_{k+1}-y_k \geqslant y_k-y_{k-1}$, it indicates that the varying rate tends to get bigger, $y_{k+1/2}$ is more likely to approach to y_k. So the error of y_k is adopted, and vice versa.

The pseudo code is listed as follow to interpret this scheme clearly:

```
If (y_{k+1}-y_k ⩾ y_k-y_{k-1})
        error_x_{k+1/2}=(3/16)*error_forward;
else
        error_x_{k+1/2}=(3/16)*error_back;
end
```

With the above equations, we can predict each 1-D pixel at horizontal and vertical directions

$$L_2(x_{k+1/2}) = y_{k-1} * a + y_k * b + y_{k+1} * c + error\_x_{k+1/2} \quad (12)$$

B. 2-D pixels interpolation

In the last subsection, 1-D pixels interpolation was discussed. In this subsection we will introduce the method for interpolating the ⊗ shown in figure 2, called 2-D pixels. We find that the pixels along the direction of local edge normally have similar values. Therefore, a precise prediction can be done if we predict the pixels using its neighboring pixels that are in the same direction of the edge.

In this paper, we adopt two Sobel operators that correspond to difference in vertical and horizontal direction to calculate edge direction. For example, if we evaluate p5, each surrounding pixel is located as figure 3.

p1	p2	p3
p4	p5	p6
p7	p8	p9

Figure 3. Template for Sobel operator.

For p5, Sobel operator is applied and obtain:

$$g(x) = (p7 + 2*p8 + p9) - (p1 + 2*p2 + p3) \quad (13)$$

$$g(y) = (p3 + 2*p6 + p9) - (p1 + 2*p4 + p7) \quad (14)$$

where *g(x)* and *g(y)* represent the difference between adjacent pixels in vertical and horizontal directions, respectively. The gradient direction angle is defined as:

$$\alpha = \arctan(g(y)/g(x)) \quad \alpha \in (-\pi/2, \pi/2) \quad (15)$$

It means that the pixels vary fastest along angle α. In order to predict pixel p5, the edge direction β, along which pixel value varies most smoothly, has to be found. According to the definition of gradient, when direction β satisfy $\beta \perp \alpha$, p5 varies most smoothly. So p5 is calculated by the neighboring pixels along β. Here, a simple average value of two adjacent pixels is adopted as p5.

Let's look at an example. If α=0, it indicates g(x) >> g(y), the values at vertical direction vary more fast than the horizontal direction. So we use the horizontal pixels to predict. The pseudo code for p5 prediction is

```
If abs(α)≤(π/8)
    p5=(p4+p6)/2;
else if abs(α) >(π/8 )&& abs(α)≤(3*π/8)
   if g(y)*g(x) ≥0
      p5=(p3+p7)/2;
   else
      p5=(p1+p9)/2;
else
   p5=(p2+p8)/2;
```

Using our technique, we can interpolate all pixels in the image. Parabola interpolation polynomial is used to interpolate 1-D signal with an adaptive error being considered, improving interpolation precision. Gradient-based method is adopted to get 2-D signal value.

IV. EXPERIMENT RESULTS

The performance of the proposed method is evaluated in this section. The test sequences are Hall, Mobile, News, Container, and Foreman. Each sequence is in qcif and magnification factor is *2x2*. Four methods, Bilinear, cubic convolution, EASE, and the proposed method, are simulated and compared. Two metrics are adopted and they are average gradient and mean structural similarity (MSSIM) [8]. Average gradient is defined as:

$$T = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\sqrt{\frac{\Delta I_x^{\,2} + \Delta I_y^{\,2}}{2}} \quad (16)$$

where $\Delta I_x = f(i+1, j) - f(i, j)$

$\Delta I_y = f(i, j+1) - f(i, j)$

It represents the contrast of an image. The bigger the T is, the sharper the image is. In table I, the proposed method shows its superior to other methods. The bilinear method have the smallest average gradient, so it is the smoothest. It shows the interpolated image looks more blurry.

TABLE I. average gradient results of the interpolated images by four algorithms

Method / Image	Bilinear	Cubic	EASE	Proposed
Hall	4.4011	4.9118	4.8379	5.3374
Mobile	8.5801	10.0720	9.8354	11.0250
News	4.6758	5.2719	5.1695	5.5733
Container	4.7412	5.4123	5.3276	5.8723
Foreman	3.8324	4.1950	4.1317	4.3801

TABLE II. MSSIM RESULTS OF THE INTERPOLATED IMAGES BY FOUR ALGORITHMS

Method / Image	Bilinear	Cubic	EASE	Proposed
Hall	0.8805	0.8927	0.8749	0.8813
Mobile	0.6061	0.6291	0.5399	0.5604
News	0.8941	0.9062	0.8896	0.8988
Container	0.8174	0.8357	0.8137	0.8270
Foreman	0.9013	0.9084	0.9124	0.9157

Mean structural similarity describes the difference between an image and its distorted version. The bigger the MSSIM is, the interpolated image is more closer to the original image. From table II, we can see the cubic method has the best performance. The proposed method is a little worse than it. EASE is the worst. Combining the two criteria, we can conclude that the proposed method has the best performance.

From above tables, we can also conclude that if the image has more details, the average gradient is bigger; when interpolating, the MSSIM is smaller because the edge is smoothed, the Mobile shows this property.

For objective point of view, we interpolate standard test image Lena. The local area of each interpolated image is shown in figure 4. Figure 4(a) is obviously blurry. The

interpolated image using the proposed method has the best performance. It has higher contrast, little zigzag artifact, and little ringing noise.

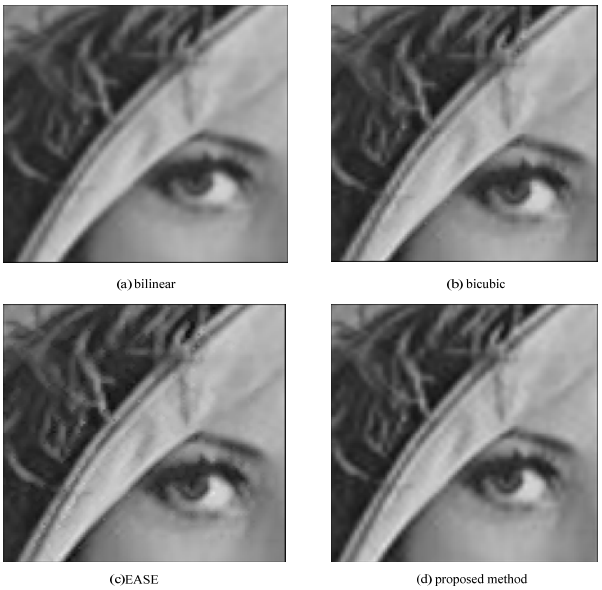


Figure 4. The local area of interpolated images

V. CONCLUSION

An interpolation method based on parabola interpolation with an adaptive error compensation has been proposed. It has better performance than EASE and bilinear interpolation. An appropriate gradient direction is discussed to predict 2-D pixels. It is simpler than EASE and introduces little artifacts on the 2-D images. Simulation results show the superior performance of the proposed method in terms of artifact and contrast.

REFERENCES

[1] R.G. Keys, "Cubic convolution interpolation for digital image processing," IEEE Trans. acoustics, speech, and signal processing, vol. assp-29, NO.6, pp:1153-1160, Dec. 1981.

[2] Y. Cha and S. Kim, "The error-amended sharp edge (EASE) scheme for image zooming", IEEE Trans. Image processing, vol.16, NO.6,pp:1496-1505,Jun.2007

[3] F. Pan, X. Lin, Susanto Ragardja , "Fast mode decision algorithm for intraprediction in H.264/AVC video coding," IEEE Trans. Circuits and systems for video technology, vol.15, NO.7,pp:813-822 Jul.2005.

[4] R. R. Schultz and R.L. Stevenson, "A Bayesian approach to image expansion for improved definition," IEEE Trans. Image processing, vol.3, NO.3, pp:233-241, May.1994.

[5] Q.Y. LI, N.C. Wang, and D.Y. Yi, "Numeric analysis" fourth edition, Tsinghua University Press,pp:23-30. （in Chinese）

[6] L. Zhang and X.L. Wu, "An edge-guided image interpolation algorithm via directional filtering and data fusion," IEEE Trans. Image processing, vol.15, NO.8,pp:2226-2238, Aug,2006.

[7] H .W. Park and Y.L. Lee, "A postprocessing method for reducing quantization effects in low bit-rate moving picture coding," IEEE Trans. Circuits and systems for video technology, vol.9, NO.1,pp:161-171,Feb,1999.

[8] Z. Wang, A.C. Bovik and H. R. Sheikh, "Image quanlity assessment : from error visibility to structural similarity,"IEEE trans. Image processing,vol.13,NO.4,pp:600-612,Apr,2004

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 271-274

Cloud Trust Model in E-Commerce

Min Wu
Jiaxing College, Jia Xing, China
Email: swjtu_wm@163.com

***Abstract*—Trust is one of the most important concerns in the E-commerce'development. In view of the significance of trust management in E-commerce, a model of trust in Electronic commerce based on the cloud computing is prospered in the paper. The Cloud Trust Model (CTM) offers cloud trust services to solve the issues of trust management in E-commerce,such as the computing of . social network analysis (SNA).**

***Index Terms*—Electronic Commerce , Trust Model , Cloud Computing , SNA**

I. INTRODUCTION

During the Internet age, online environment is now part of everyday life. As a result, Web applications have become a major tool for connecting consumers, partners, and employees to company applications using the Internet as an open and global network.E-commerce is a popular and growing Web application which enables customers to achieve a variety of purposes and services. Electronic commerce lets companies integrate internal and external business processes through information and communication technologies. Companies conduct these business processes over intranets, extranets, and the Internet. E-commerce lets businesses reduce costs, attain greater market reach, and develop closer partner relationships. Numerous e-commerce companies have created very profitable businesses since pioneering e-commerce traders (such as Amazon.com) emerged. However, using the Internet as the underlying backbone network has led to new risks and concerns. Often, industry analysts cite trust and security as the main hurdles in growing e-commerce.[1][2]

In the abstract sense, trust is the extent to which one party measures the other party's willingness and ability to act in the measuring party's interest.[3] The notion of trust varies in different contexts. Broadly speaking, there are two classes of trust computing: security-oriented trust computing and non-security-oriented trust computing. [4] In consequence, security can be regarded as one of the trust issues.

In security-oriented trust computing, trust provides a mechanism for enhancing security, covering issues of authentication, authorization, access control, and privacy. [4]Trust is the degree by which a target object (such as software, a device, a server, or any data they deliver) is considered secure.

Reputation-based trust evaluation correlates to non-security-oriented trust computing. In general, a service gains a good reputation after it has accumulated good quality services over a long time period. The evaluation is usually based on customer ratings. However, to compute the final reputation value correctly, studies on relationships among raters and ratees are necessary, and might help reduce the rating noise (which we discuss further later on) and obtain more objective trust results.[5]

We can view security-oriented trust and non-security-oriented trust as the trust in EC respectively from a technological perspective and a systemic perspective. In this paper, we will discuss mainly the latter.

II. CATEGORY OF TRUST EVALUATION

Three are three types of trust management in EC according to trust management architecture.

A. Centralized Trust Management

Centralized trust management system puts up a centralized management server, in which service clients or buyers report ratings to a trust authority after transactions. The server manages service providers' and clients' portfolio data as well as service providers' trust data. Some well-known e-commerce websites have taken Centralized trust management system to build up trust, such as eBay and Taobao set forth internal credit rating mechanisms.

For purposes of E-Commerce, we can define a trusted third party as a widely accepted, reliable, independent, and highly secure entity that generates trust through attestation or certification.[2]Compared with the legal mechanism, credit mechanism is a kind of lower-cost mechanism, which can also apply to E-Commerce environment. [6] To guarantee security and reliability of e-commerce, dozens of third-party online authorities has been fostered out of the legal framework, among which the online authentication institution through guarantee or implementation supervision by a third party to prevent deceptive acts is the most effective. One of the most typical third-party payment method is 'Zhi-Fu-Bao', which has applied by many web sites. At Taobao, after each transaction, a buyer can give feedback to the system about the seller's service quality that can be positive, neutral, or negative. Taobao stores this rating at a centralized management location. It calculates the feedback score via $S = P - N$, where P is the number of positive ratings left by members (customers) and N is the number of negative ratings. Taobao displays the S value on the seller's Web page. Another value $R = (P - N)/(P + N)$ ($1 \geqslant R \geqslant 0$) is the positive feedback rate.

B. Decentralized Trust Management

A decentralized (such as P2P) architecture trust management, which also has its benefits.[7]The typic example of the decentraliazed trust management of

E-commerce occurs in the P2P(peer-to-peer) E-commerce. The peer-to-peer (P2P) systems and applications employ distributed resources to perform critical functions in a decentralized manner. Peer-to-peer computing is the sharing of computer resources and services through direct communication between systems. a P2P-based architecture doesn't require extra costs to set up separate servers, but once a requesting peer needs to know a service provider's trust status, in general, it must broadcast a request to other peers. Thus, this architecture is costly in terms of network communication.

Person-to-person online auction sites such as eBay and many business-to-business (B2B) services such as supply-chain-management networks are examples of P2P communities built on top of a client-server architecture. In E-Commerce settings, P2P communities are often established dynamically with peers（transction parties） that are unrelated and unknown to each other. Peers have to manage the risk involved with the transactions without prior experience and knowledge about each other's reputation. One way to address this uncertainty problem is to develop strategies for establishing trust and develop systems that can assist peers in assessing the level of trust they should place on an E-Commerce transaction. [8]

C. Distributed Trust Management

Different from either of these architectures is a distributed architecture, which comprises a set of trust management brokers that partition the data among themselves. This method also helps partition the trust computation workload and provides a more reliable environment because it can ensure a relatively complete data set. However, the collaboration among brokers and the cost to set them up might be concerns.

Researchers have also actively studied trust issues in multiagent environments. A software agent is autonomous and self-interested, expected to complete the tasks its owner or other agents specified. In addition to evaluating trust in agent interactions (such as transactions in an e-commerce context or services in an SOC one), studies looking at multiagent environments must consider other issues, such as agents' motivations and the influence and dependency relationships among them.[9]

III. CLOUD TRUST MODEL IN EC

Recently, cloud computing has emerged as an important technology that has received attention from both the research community and service industry. Cloud computing is a recent trend in IT that moves computing and data away from desktop and portable PCs into large data centers. Within the cloud,Consumers purchase such services in the form of infrastructure-as-a-service (IaaS), platform-as-a-service (PaaS), or software-as-a-service (SaaS) and sell value-added services(such as utility services) to users. We can apply cloud computing technology to the trust management in E-commerce.And then,a Cloud Trust Model (CTM) is prospersed in the article concerning the applications of cloud computing at persent.

A. Applications of Clouds

We can distinguish two different architectural models for clouds: the first one is designed to scale out by providing additional computing instances on demand. Clouds can use these instances to supply services in the form of SaaS and PaaS. The second architectural model is designed to provide data and compute-intensive applications via scaling capacity.[10]

B. the Framework of Cloud Trust Model in E-commerce

As showed by the figure 1,the CTM comprises

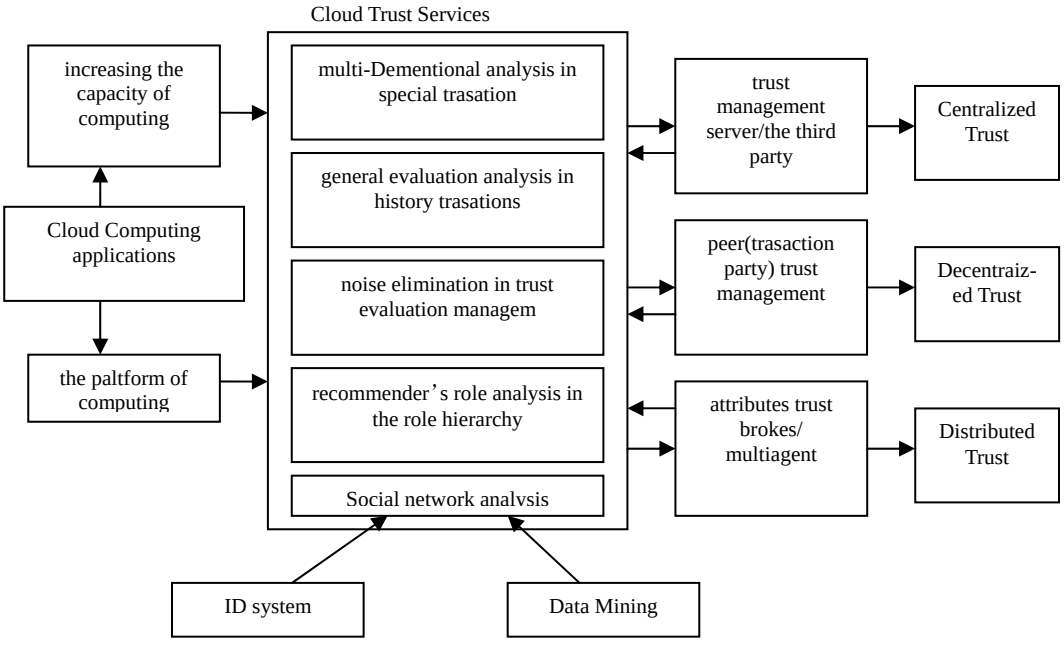


Figure 1. the Framework of Cloud Trust Model in E-commerce

four layers.The far right of the graph is the applicationg layer,and the middle right is the agent layer,and the middle is the service layer which is the heart of the model,and the the far left layer is the cloud layer.In the service layer of CTM,social network analysis (SNA) service can refer to the ID system in physical world and the data mining technology. general evaluation analysis of history trasations,but also multi-Dementional analysis in special trasation.

C. Cloud Trust Services in CTM

a) multi-Dementional analysis and general evaluation analysis

In most existing trust models, a given seller's trust is computed as the general trust based on ratings from all previous transactions with the seller in a recent period. This trust value might not indicate the exact trust level a new transaction might have; this is a consumer's real concern, particularly when the seller is unknown.

Researchers have conducted more studies that bind the trust evaluation to other properties of a new transaction.[8][11]

In Cloud Trust Services,there are not only the general evaluation analysis of history trasations,but also multi-Dementional analysis in special trasation.

b) noise elimination

One major impediment to obtaining objective trust results is rating noise, which can occur when a friendship or competitor relationship exists between a rater and ratee, leading to low accuracy ratings. For example, if a rater is a friend of the ratee, his or her rating might be overly high. On the other hand, if the rater is a competitor (or a friend of a competitor) of the ratee, the rating might be overly low.

An improved method of credit counting was introduced, which used the interval distribution of commodity prices to rate success trades and used different deducting coefficients to rate failed trades according to different credit rating.[11]

c) Recommender's role analysis

Some studies assume that transaction trust is conceptually equivalent to recommendation trust[12] or aim to analyze raters' credibility.[13] This might help reduce the bias in trust computation, but these studies calculate credibility from the requesting party's viewpoint, using its own experience, which yields local and subjective result. Thus, it's not global and might not be valuable to other service customers. We must develop new approaches to analyzing a rating's trustworthiness.

One study proposes a role-based recommendation and trust-evaluation model,[14] which uses a recommender's role to evaluate the recommendations he or she gives. This role includes recommenders' social position, title, or rank, reflecting the expertise level, as well as the recommendations' impact level in the target domain. Typical scenarios from real applications include job hunting, in which a referee for the job seeker plays an important role in terms of how his or her recommendation influences the potential employer. However, when using this realistic rolebased framework, the issue of how to build up or describe the role hierarchy remains challenging.

d) Social network analysis

Traditional studies focus more on analyzing the trust rating data itself, trying to identify the noise statistically. We think the analysis should consider some additional relationships as well. We can adopt data mining and social network analysis (SNA) techniques for this purpose, based on graph theory and relational algebra, for analyzing social relationships between individuals. [15] Massive quantities of data are available via blogs, e-commerce sites, social networking sites, newsgroups, chat rooms, and so on. These networks typically have tens of thousands to millions of nodes and contain sufficient information for assembling into analysis models.[16] By applying SNA, we can analyze various parties'relationships and enhance our abilities to trace colluding attacks in trust evaluations. The establishment of social credit system in combination with the ID system.[6]

Ⅳ. CONCLUTION

The could trust model effectively solves the some problems in trust management.Of the three trust services we've discussed, some utilize transaction parameters to inspect only relevant ratings, whereas the others improve rating data's accuracy. However further study on these topics will require the development of new and extended theoretical models to manage the complex issues behind trust.

REFERENCES

[1]Radwan M. AL-Dwairi, Mumtaz A. Kamala。An Integrated Trust Model for Business-to-Consumer (B2C) E-Commerce:Integrating Trust with the Technology Acceptance Model. 2009 International Conference on CyberWorlds.p352

[2]Vijay Ahuja.Building Trust in Electronic Commerce.IT Pro. May x June 2000.pp.61-62

[3]D.H. Knight and N.L. Chervany.*The Meaning of Trust*,tech. report WP9604, Univ. of Minnesota, 1996.

[4]N. Griffiths. Trust: Challenges and Opportunities.*AgentLink News*, no. 19, 2005, pp. 9–11.

[5]Yan Wang and Kwei-Jay Lin.Reputation-Oriented Trustworthy Computing in E-Commerce Environments.

[6]Lihua Ruan, Ding Tian.A Research of Trust Based on E-Commerce. International Symposium on Electronic Commerce and Secu 2008 IEEE.pp.776-779

[7]L.-H. Vu, M. Hauswirth, and K. Aberer, "QoS-Based Service Selection and Ranking with Trust and Reputation Management," *Proc. Int'l Conf. Cooperative Information Systems* (CoopIS 05), LNCS 3760/3761,Springer-Verlag, 2005, pp. 466–483.

[8]Li Xiong and Ling Liu, PeerTrust: Supporting Reputation-Based Trust for Peer-to-Peer Electronic Communities,*IEEE Trans. Knowledge and Data Eng.*, vol. 16, no. 7, 2004, pp. 843–857.

[9] T. Huynh, N. Jennings, and N. Shadbolt, An Integrated Trust and Reputation Model for Open Multi-Agent Systems. Autonomous Agents and Multiagent Systems, vol. 13, no. 2, 2006, pp. 119 – 154.

[10]Marios D. Dikaiakos and George Pallis Dimitrios Katsaros Pankaj Mehra Athena Vakali . Cloud Computing :Distributed Internet Computing for IT and Scientific Research. IEEE Internet Computing.Serptmeber/October 2009.pp.10-13.

[11]WANG Liang, GUO Ya-jun1. Trust Modeling and Evaluation in E-commerce System. Computer Engineering, Vol.35 No.10, May 2009,pp.129-131.

[12]S.D. Kamvar, M.T. Schlosser, and H. Garcia-Molina,"The EigenTrust Algorithm for Reputation Management in P2P Networks," *Proc. World Wide Web Conf.* (WWW03), ACM Press, 2003, pp. 640–651.

[13]B. Yu, M.P. Singh, and K. Sycara, "Developing Trust in Large-Scale Peer-to-Peer Systems," *Proc. IEEE 1st Symp. Multi-Agent Security and Survivability*, IEEE CS Press, 2004, pp. 1–10.

[14]Y. Wang and V. Varadharajan, "Role-Based Recommendation and Trust Evaluation," *Proc. IEEE Joint Conf. E-Commerce Technology* (CEC 07) *and Enterprise Computing, E-Commerce, and E-Services* (IEEE 07), IEEE CS Press, 2007, pp. 278–295.

[15]F. Martino and A. Spoto, "Social Network Analysis: A Brief Theoretical Review and Further Perspectives in the Study of Information Technology," *PsychNology J.*,vol. 4, no. 1, 2006, pp. 53–86.

[16]S. Staab et al., "Social Networks Applied," *IEEE Intelligent Systems*, vol. 20, no. 1, 2005, pp. 80–93.

ISBN 978-952-5726-09-1 (Print)
Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10)
Jinggangshan, P. R. China, 2-4, April. 2010, pp. 275-281

Database Auditing Design on Historical Data

Narongrit Waraporn
School of Information Technology, King Mongkut's University of Technology Thonburi, Bangkok, Thailand
Email: narongrit@sit.kmutt.ac.th

***Abstract*—Database Auditing is one of the major issues in information security. Lack of data auditing leads the business applications to the lost trail of business processes. To cope with auditing, historical data or temporal database is needed in order to track operations and actors of the operation with the time. Valid and transaction times are two major timestamps in temporal database. In this paper, we demonstrate the techniques to handle database auditing in business transaction by considering operations, valid times, and actors of the operations. These techniques are divided in two sets; using relational databases, and using semi-structure data.**

***Index Terms*— Database Auditing, Historical Data, Temporal Database, Object-Relational Database, XML**

I. INTRODUCTION

Information System, IS, and Information Retrieval, IR, have been defined and adopted into the practices since the beginning of digital revolution. While information has been used for business processes, information security emerged in term of authentication and authorization. Those concepts can protect information but do not provide help in investigation. Then, the log files were proposed to track the trail of access to databases and systems. However, the main purpose of database log is to recovery the transaction. To be able to investigate the transaction, the database audit was introduced.

Many commercial databases suggest database auditing by using the log files. DB2 offers the creation of "audit policy" to apply to certain roles of users. However, they cannot cover every aspect of business processes when considering historical data. We propose techniques of database auditing based on historical data for proprietary systems.

In some transactional systems, database design includes auditing data into operational table. However, the auditing data is not used often. We propose the separation between auditing data and business operation data.

This paper is organized into three main parts. First, the fundamental and reviews of auditing and historical data are in section II and section III respectively. Second, the design of database auditing is presented in various approaches. Using well-established relational databases to implement is suggested with three approaches; row-based, column-based and log-table auditing, in section IV. While semi-structured database serves the complex information, we suggest the use of two alternatives; Object-Relational Database and XML in section V. At last, we conclude and discuss other issues of database auditing in section VI.

II. AUDITING

Auditing a change to a management system such as accounting, performance, and software, has been well defined and proctored to improve the quality. One of the major sources of auditing is data where auditor can find the errors and misuses. Lately, data, itself, is audited for its movement to identify malicious behavior, maintain data quality, and improve system performance in [7], and [13]. [7] proposed a framework for auditing the changes to a database system with the retention policies. The declarative rules for expressing retentions, semantics for audit query answers, and performance of audit query under retention restrictions were demonstrated. To detect anomaly behavior, [13] suggested a grid-based clustering method on a transaction in an audit data stream containing a set of user activities at a particular time.

Since auditing data is likely to be abundant, searching for an auditing answer requires data mining techniques such as classification and regression. [8] presented an auditing data tool based on data mining algorithms. They evaluated their QUality Information System (QUIS) database in different alternatives, instance based classifiers, naive Bayes classifiers, classification rule inducers, and decision trees.

To extend auditing concept into communities, [16] developed a hand-on experience on database security and auditing on relational table containing an audit trail of all operations that change data on the target table. But, the paper did not specify the design of audit trail table while our paper suggests the design alternatives with their pros and cons.

III. HISTORICAL DATA

One key success of auditing is to be able to track the change on the trail of who made the changed, what operation of the modification, and when it happen. The first two trails can be manipulated in relational model while the last trail requires temporal characteristics to handle.

Two different types of time are considered when recording the history of information; valid time and transaction time. [1] suggested that valid and transaction times should guarantee no information loss in any circumstantial transactions. [2] suggested that the transaction time is needed to satisfy the data mining. However, in this paper, we demonstrate only the valid time. Transaction time can be stamped when the "user" column saves the username whom modifies the record.

However, temporal databases have not been widely adopted by commercial databases and IT communities

AP-PROC-CS-10CN006

due to the overhead of temporal computing. A different approach on temporal data is to use XML document to manage historical data. [3] fragments transaction information into event data to retrieve some data from XML document. This is due to transmitting large transaction but only the context change based on historical data is sent. An extension of XQuery language is also demonstrated in the paper. Alternatively, implementation techniques based on temporal characteristics to audit data in plain-relational and XML-enabled databases are suggested in this paper.

IV. Historical Design for Auditing on Relational Databases

Historical data can be modeled in Relational Database, RDB, in several techniques such as separated tables for history records, transaction logs, and multi-dimension data using XML. Creating a separated table for historical data of each relational table is a straight forward solution with minimum design complexities while the transaction logs must cross-check data from the data dictionary. Both solutions require no change to the original relational tables. Without additional tables, XML columns can be added to the original table with the background of XPATH, XQUERY, and SQL/XML for the data retrieval.

To maintain historical for auditing, four suggested techniques can be implemented; row-based auditing, column-based auditing, log-table auditing, and semi-structure-based auditing which can be implemented by using object-relational types or XML. The first three types are implemented in the relational databases while the semi-structure-based auditing must rely on the extension of database engines on object-relational and XML technologies such as in IBM DB2 9.5, Oracle 10g, and MS SQL Server 2007, but not in MySQL 5.1 where XML data is still maintain in varchar column [10]. We explained the semi-structure-based auditing the section V.

A. Row-based Auditing

This technique creates a separated table for each relational table to maintain historical data. The operational table remains the same as in non-auditing system.

An EMPLOYEE table of operational tables, as shown in Table I maintains only the current value of each employee for business operations. The table, also, includes both static data and historical data. The static data remains unchanged such as data of birth or rare to change such as name. For the historical data, only the last-updated values are maintained in the operational table. The non-historical query, which is regularly requested, still remains unchanged in order to retrieve the information from the EMPLOYEE table.

The auditing table contains every column of the operational table as shown in Table II. To diminish the join-operation query, static data is included in the auditing table. Two timestamps are needed for the valid times; start time and end time to maintain the lifespan of the data. Two additional attributes; operation type and username should be considered. Operation type is recorded to reduce the overhead of comparison among histories of the same data. Username is stored for the reason of responsibility of the sensitive data.

TABLE I.
Operational Employee Table Referenced by Employee-History Table for Row-Based Auditing

ID	Name	DOB	Address	Hired Date	Salary
101	Tom	1980-01-01	New York	2008-01-01	55000
102	Ann	1985-06-16	London	2009-01-01	60000

Table II shows the history table of the EMPLOYEE table. The history table shows that Jack has worked since 2008/01/01. He moved from New York to Hong Kong on 2008/06/15. On 2009/01/01, his wage was increased to 70000 authorized by Mel. At the end of May 2009, he left the company and it is recorded by Matt.

The records are the same between records in the operational table and the auditing table whose end time is null. The redundancy of these records is for the historical query so that there is no need to retrieve data by using union operation between them.

The row-based auditing has pros and cons.

Pros

- It simplifies the implementation of auditing. When a DML statement such as INSERT, UPDATE, and DELETE, is executed on the operational table, the application can simply copy every value in the record into the auditing table. At the same time, the end time column of the previous history must be updated with the operated time. However, this could be done by database trigger if the system does not handle the temporal issues at the application level. [16], also, used this approach.
- Retrieving static data and historical data are independence. There is no need to union the data from two tables together in any query because the auditing table maintains the static data.

Cons

- Redundancy of current record in both tables. However the historical query would be less complex.
- Retrieving historical data requires the comparison between two records by using recursive query. For example, SQL statement for the auditing of salary of Jack whose ID is 103 is

```
SELECT E1.SALARY, MINS, MAXE,
E1.USER, OPERATION
FROM EMPLOYEE_HISTORY_R E1,
    ( SELECT E2.SALARY, MIN(E2.STARTTIME)
        MINS,   MAX(E2.ENDTIME) MAXE
      FROM EMPLOYEE_HISTORY_R E2
      WHERE ID = 103  GROUP BY SALARY) E3
WHERE E1.SALARY = E3.SALARY
```

TABLE II.
OPERATIONAL EMPLOYEE TABLE REFERENCED BY EMPLOYEE-HISTORY TABLE FOR ROW-BASED AUDITING

PK	ID	Name	DOB	Address	HiredDate	Salary	StartTime	EndTime	Operation	User
1	101	Tom	1980-01-01	New York	2008-01-01	50000	2008-01-01	2009-01-01	I	Mike
2	101	Tom	1980-01-01	New York	2008-01-01	55000	2009-01-01		U	Mel
3	102	Ann	1985-06-16	London	2009-01-01	60000	2009-01-01		I	Mike
4	103	Jack	1975-01-01	New York	2008-01-01	60000	2008-01-01	2008-06-15	I	Mike
5	103	Jack	1975-01-01	Hong Kong	2008-01-01	60000	2008-06-15	2009-01-01	U	Mike
6	103	Jack	1975-01-01	Hong Kong	2008-01-01	70000	2009-01-01	2009-05-31	U	Mel
7	103	Jack	1975-01-01	Hong Kong	2008-01-01	70000	2009-05-31	2008-05-31	D	Matt

B. Column-Based Auditing

The column-based auditing solves the redundancy of the row-based auditing. This auditing table does not include the static columns such as date of birth and hired date of an employee. Data in historical column of auditing table are stored only the changed value except the primary key, such as ID, which is used to reference its operational table.

Employee history in Table III stores only the changed data. It is clear to see that data in Table II is less redundant than data in Table I. Under the same ID, Jack moved from New York to Hong Kong on Jun/15/08 and get raise from 60000 to 70000 on Jan/01/09. Selecting not-null value on a particular auditing column in SELECT statement would display only the actual change. For example,

```
SELECT SALARY, STARTTIME, ENDTIME,
USER, OPERATION
FROM EMPLOYEE_HISTORY_C
WHERE ID = 103 AND SALARY IS NOT NULL
```

The query displays the auditing of Jack's salary. Comparing with row-based auditing on the same query, the SELECT statement is much less complex.

Each record in column-based auditing table cannot contain more than one value of historical data because of the uncertainty of end time of each auditing data.

TABLE III.
EMPLOYEE_HISTORY_C TABLE USING COLUMN-BASED AUDITING

PK	ID	Address	Salary	StartTime	EndTime	Opera-tion	User
1	101	New York		2008-01-01		I	Mike
2	101		50000	2008-01-01	2009-01-01	I	Mike
3	101		55000	2009-01-01		U	Mel
4	102	London		2009-01-01		I	Mike
5	102		60000	2009-01-01		I	Mike
6	103	New York		2008-01-01	2008-06-15	I	Mike
7	103		60000	2008-01-01	2009-01-01	I	Mike
8	103	Hong Kong		2008-06-15		U	Mike
9	103		70000	2009-01-01		U	Mel
10	103			2009-05-31		D	Matt

Pros

- The historical query on auditing is less complex than row-based auditing. So, it is likely to execute faster.
- Disk space would be used less than the row-based auditing if the historical column is varchar.

Cons

- Maintaining many NULL values in the table would lead to other problems especially when writing a query without considering the semantic of NULL cautiously [15].
- If most operation on the table is INSERT, but not UPDATE, the number of records would be n times bigger where n is the number of historical columns; two times in our examples because of ADDRESS, and SALARY.
- SELECT statement requires a recursive join in order to query two or more auditing columns. Therefore, it is higher complexity than the case of row-based auditing for this query.

C. Log-Table Auditing

Log tables have been used for transaction management in the relational database for a long time. Due to the nature of transaction that needs to know operation, data, and time of execution, log tables can be utilized for the auditing purpose too. For example, DB2 audit facility allows DBA to maintain an audit trail for a series of predefined database events and saved it in an audit log file [6]. Most major databases such as Oracle and SQL Server use a similar method [11] and [9] while MySQL is partially support [10]. Postgres recommends the use of trigger for audit trail [12].

However, the log tables of commercial databases are not intended for business process but rather for database administrator purpose. Additionally, log tables do not serve some auditing aspects such as end time to application. We suggest two developing approaches to handle auditing data using log mechanism.

1) Multiple Audit Log Column Tables for Transaction Logs

To separate auditing log data from the operational data, we create extra table for each auditing column. For example, if ADDRESS and SALARY columns in the EMPLOYEE table are auditing columns, we create ADDRESS and SALARY tables for auditing purposes as shown in Table IV and Table V.

Pros

- It reduces the size of an auditing record.
- It simplifies the design of auditing tables.

Con

- The number of tables which is depending on the number of auditing columns for each table could be high.

TABLE IV.
AUDIT LOG TABLE FOR ADDRESS

PK	ID	Address	StartTime	EndTime	O	User
1	101	New York	2008-01-01		I	Mike
2	103	New York	2008-01-01	2008-06-15	I	Mike
3	103	Hong Kong	2008-06-15	2009-05-31	U	Mike
4	102	London	2009-01-01		I	Mike
5	103		2009-05-31		D	Mel

TABLE V.
AUDIT LOG TABLE FOR SALARY

PK	ID	Value	StartTime	EndTime	O	User
1	101	50000	2008-01-01	2009-01-01	I	Mike
2	103	60000	2008-01-01	2009-01-01	I	Mike
3	101	55000	2009-01-01		U	Mel
4	103	70000	2009-01-01	2009-05-31	U	Mel
5	102	60000	2009-01-01		I	Mike
6	103		2009-05-31		D	Mel

2) Single Audit Log Table for Transaction Logs

To combine audit data into one place, we integrate every auditing column from all operational tables into one single auditing log table. The audit log table composes of name of table and column, primary key of the record in the operational table, new value, valid start time, operation that causes the change and name of user who manipulates this information.

Example of single audit log table of the database containing EMPLOYEE and DEPARTMENT tables is shown in the Table VI. Every transaction occurring to the operational table on the sensitive column is written into the single audit log table. A single insertion of employee number 101 into EMPLOYEE table activates the insertion into audit log table two times; one log record for ADDRESS and another for SALARY if EMPLOYEE table has two auditing columns. Updating on an auditing attribute will insert an auditing record into the log table. Similar to insertion, deletion of a record will be logged twice into audit log table in case of two auditing columns such as deletion of employee 103 in Table VI.

TABLE VI.
SINGLE AUDIT LOG TABLE FOR EVERY TABLE; EMPLOYEE AND DEPARTMENT, IN DATABASE

PK	Table	Column	ID	Value	StartTime	EndTime	O	User
1	Employee	Address	101	New York	2008-01-01		I	Mike
2	Employee	Salary	101	50000	2008-01-01	2009-01-01	I	Mike
3	Employee	Address	103	New York	2008-01-01	2008-06-15	I	Mike
4	Employee	Salary	103	60000	2008-01-01	2009-01-01	I	Mike
5	Employee	Address	103	Hong Kong	2008-06-15	2009-05-31	U	Mike
6	Employee	Salary	101	55000	2009-01-01		U	Mel
7	Employee	Salary	103	70000	2009-01-01	2009-05-31	U	Mel
8	Employee	Address	102	London	2009-01-01		I	Mike
9	Employee	Salary	102	60000	2009-01-01		I	Mike
10	Employee	Address	103		2009-05-31		D	Mel
11	Employee	Salary	103		2009-05-31		D	Mel
12	Department	Manager	D1	103	2008-01-01		I	Mel
13	Department	Manager	D1	101	2009-05-31		U	Mel

Pro

• It minimizes the number of tables. There is no need to scatter the auditing data into many tables.

Cons

• Query requires extra predicates to check the names of table and column.

• The data type of VALUE column must be general type such as VARCHAR due to the various possibility of data type of auditing column e.g. CHAR for ADDRESS, but DECIMAL for SALARY. Data type must be casted back to the original type before comparing it with the original column.

For example, an SQL statement to find name of last user who made a change (UPDATE or INSERT) the SALARY column is

```
SELECT SL.USER, E.name
FROM EMPLOYEE E, SINGLELOG SL
WHERE E.ID = SL.ID AND
  SL.TABLE = 'Employee'  AND SL.COLUMN =
   'Salary' AND E.SALARY = INT(SL.VALUE)
```

• Audit log table is very large if there are many auditing columns from different tables, especially, on the system with lot of transactions. Dividing data into subsystems and having a single audit log table for each subsystem are recommended.

Both approaches require extra processing for every transaction occurring to the databases, especially, the auditing data. By default, database engines have already manipulated log tables. With this extra processing, the overall system will be degraded. If the auditing is not the main purpose or major use of the application, the system log handling by the database engine is recommended. The supports from database administrators and extra learning of programmers are required.

V. SEMI-STRUCTURED-BASED AUDITING

Since the emerging of objection-oriented paradigm, structured data has been redefined into complex information. Objected Oriented Databases, OODB, had been researched and proposed to communities for many years. However, the IT industry, especially in database software such IBM DB2, Oracle DB and others, has not completely adopted OODB. In fact, Object-Relational database, ORDB, was added into relational database for semi-structured information. Oracle DB includes user-defined types, array of a type, nested tables, and object methods to serve the inheritances, polymorphism, and encapsulation [5].

In contrast, IBM DB2 has proposed the semi-structured information into another technology, XML which claims to be pure XML [4]. It means that data is neither stored as Character Large Object, CLOB, nor shredded into multiple relational tables. It stores data in tree-structure as the nature of XML.

Therefore, we suggest two semi-structured types to manage auditing data; object-relational and XML types.

A. Object-relational Type

To audit data in ORDB, we create new types for auditing such as StringAudit type to store auditing data of character columns, and NumberAudit of numeric columns.

```
CREATE TYPE  StringAudit AS(
    VALUE VARCHAR(20),
    STARTDATE DATE,
    ENDDATE DATE,
    OPERATION CHAR(1),
    USER CHAR(10));

CREATE TYPE  NumberAudit AS(
    VALUE DECIMAL(10,2),
    STARTDATE DATE,
    ENDDATE DATE,
    OPERATION CHAR(1),
    USER CHAR(10));
```

Auditing data such as start and end dates of such information, operation and user that manipulate the information are included in the new auditing types. However, the auditing type is used to store only one single record of transaction. To handle the historical data as temporal database, we create an array type for each auditing type.

```
CREATE TYPE StringAuditArray as
        VARRAY(100) OF StringAudit;

CREATE TYPE NumberAuditArray as
        ARRAY(100) OF NumberAudit;
```

However, DB2 array is based only on one of the built-in data type. It means creating array of user-defined type is currently not allowed in DB2. In contrast, Oracle allows array of user-defined type in a special type called VARRAY.

Each element in the array represents the single record of history. Therefore, when creating a table for the auditing purposing, we declare the auditing array type for columns needed to be audited.

```
CREATE TABLE EMPLOYEE_AUDIT_ORDB
 ( ID CHAR(10) NOT NULL,
   NAME CHAR(15),
   DOB DATE,
   ADDRESS VARCHAR(20),
   ADDRESS_AUDIT STRINGAUDITARRAY,
   HIREDDATE DATE,
   SALARY DECIMAL(10,2),
   SALARY_AUDIT NUMBERAUDITARRAY,
  PRIMARY KEY (ID) );
```

Since most transactions need the current value instead of its history, the current-valued columns such as ADDRESS, and SALARY of auditing columns are created separately from the auditing array for faster transaction of most query.

B. XML Type

Due to the semi-structure of XML, auditing data can be defined in the nested elements of XML without creating a new data type as in ORDB. XML also allows repeating elements under the same parent node. So, the historical records of an auditing column such as ADDRESS can be maintained as many as records.

Most commercial databases enable XML data type to maintain semi-structure infomation as it is. We suggest two auditing approaches using XML;

- column-based XML to audit each field of the record
- row-based XML to audit each record

1) Column-based XML Auditing

Column-based XML auditing maintains history of each column of the record separately. The XML tree is at minimum. For example, Employee table in Table I is added with two additional XML columns, ADDRESS_AUDIT and SALARY_AUDIT. XML tree of SALARY_AUDIT stores only of salary history of one employee only. Tom's salary has increased once. His auditing salaries in XML tree are in two child-nodes as shown in Table VII. XML schema for the salary auditing is shown in Fig. 1.

TABLE VII.
COLUMN-BASED XML TABLE TO AUDIT ADDRESS AND SALARY

ID	Name	DOB	Address	Address_Audit	Hired Date	Salary	Salary_Audit
101	Tom	1980-01-01	New York		2008-01-01	55000	
102	Ann	1985-06-16	London		2009-01-01	60000	
103	Jack	1975-01-01			2008-01-01		

Pros

- There is no extra auditing table.
- Auditing data is separated in different columns but under the same record. This is the same with ORDB approach.
- Size of table is as small as the number of records except the deleted records which is maintained for the auditing purpose. We can create view to retrieve only current employee by using XML/SQL to choose only XML node whose "operationType" element is "D". Setting the salary of deleted employee to an invalid value such as -1 would simplify the query to SQL only.

Cons

- It requires extra knowledge of XQUERY of developers.
- Joining two or more XML tree may be needed for a query that needs to access two or more auditing columns.

```
<schema >
 <element name="auditing">
  <complexType>
   <sequence>
    <element minOccurs="1" name="history" type="historyType"/>
   </sequence>
  </complexType>
</element>
<complexType name="salaryHistoryType">
   <sequence>
     <element minOccurs="1" name="value" type="salaryType"/>
     <element minOccurs="1" name="startdate" type="Date"/>
     <element minOccurs="0" name="enddate" type="Date"/>
    <element minOccurs="1" name="operation" type="operationType"/>
     <element minOccurs="1" name="user" type="string"/>
   </sequence>
   <attribute name='id' type='string'/>
</complexType>
<simpleType name="operationType">
    <restriction base="string">
      <enumeration value="I"/>
      <enumeration value="U"/>
      <enumeration value="D"/>
     </restriction>
  </simpleType>
 <simpleType name="salaryType">
    <union memberTypes="salary0  salaryRange"/>
 </simpleType>
 <simpleType name="salary0">
     <restriction base="decimal">
       <enumeration value="0"/>
     </restriction>
  </simpleType>
   <simpleType name="salaryRange">
     <restriction base="decimal">
        <minInclusive value="10000"/>
        <maxInclusive value="1000000"/>
     </restriction>
  </simpleType>
</schema>
```

Figure 1. XML Schema for column-based XML auditing of Salary_Audit column in Employee table

2) Row-based XML Auditing

To combine auditing information of entity into a single XML tree, row-based XML auditing merges child nodes of every XML trees of the same record of column-based XML auditing into one tree under the same column as shown in Table VIII. An attribute named "column" of "history" element in XML Schema in Fig. 2 identifies the auditing column.

TABLE VIII.
ROW-BASED XML TABLE FOR AUDITING

ID	Name	DOB	Address	Hired Date	Salary	Audit
101	Tom	1980-01-01	New York	2008-01-01	55000	
102	Ann	1985-06-16	London	2009-01-01	60000	
103	Jack	1975-01-01		2008-01-01		

Pros

• Auditing data is not retrieved often. Separating it in a column can reduce the overhead of business operational functions.

• XML data still requires larger spaces comparing to relational data. Commercial data engines tend to store them separately from the relational data. Therefore, it does not increase the number of data blocks that contain only relational data.

• Physically, combining all XML data from different logical columns into a single column would simplify data retrieval. Another word, an Xquery expression can retrieve information from a single source.

Cons

• Design of auditing data does not divide information logically.

• Xquery on row-based XML auditing will be more complex than of column-based.

```
<schema >
 <element name="auditing">
  <complexType>
   <sequence>
    <element minOccurs="1" name="history" type="historyType"/>
   </sequence>
  </complexType>
 </element>
  <complexType name="historyType">
    <sequence>
     <element minOccurs="1" name="value" type="string"/>
     <element minOccurs="1" name="startdate" type="Date"/>
     <element minOccurs="0" name="enddate" type="Date"/>
     <element minOccurs="1" name="operation" type="operationType"/>
      <element minOccurs="1" name="user" type="string"/>
    </sequence>
    <attribute name='id' type='string'/>
    <attribute name='column' type='columnType'/>
  </complexType>
  <simpleType name="operationType">
     <restriction base="string">
       <enumeration value="I"/>
       <enumeration value="U"/>
       <enumeration value="D"/>
     </restriction>
  </simpleType>
   <simpleType name="columnType">
    <restriction base="string">
      <enumeration value="Address"/>
      <enumeration value="Salary"/>
     </restriction>
   </simpleType>
</schema>
```

Figure 2. XML Schema for row-based XML auditing of Employee table

Our XML approach on temporal data suggests the use of XML elements to demonstrate the auditing data primarily. On the other hand, [3] suggests the use of XML attributes for temporal data. Their data stream management framework for historical XML data proposed XCQL, which is a XQuery Language for Continuous Queries for the temporal extension. Our approach could be simply adapted to their scheme when it is adopted by commercial databases.

VI. CONCLUSION AND DISCUSSIONS

Table space (or disk space) of auditing table should be separated from the operational table. Database engine could run an auditing query faster than running it against one large table containing both operational and auditing data. Even though, we partition one large table into two table spaces. Overhead of checking which partition will be used against the query is added

to execution time. Also, DBA would manage the DBMS easier.

To retrieve auditing information, XQuery such as FLWER expression, in SQL/XML provides us access to elements and attributes in XML data.

XML indexing is adding to commercial databases while semantic searching on XML is on horizon [14]. These issues and others will provide support to semi-structured approaches.

We suggest a various number of solutions of database auditing. Some solutions are suitable for databases that do not efficiently support semi-structured data. While semi-structure data are emerging into some commercial databases.

Database auditing is a major concern in some systems. Choosing the right design would prevent the performance deterioration, reduce data redundancy, save the storage, and simplify auditing queries.

REFERENCES

[1] G. Bhargava, S. K. Gadia, "Relational Database Systems with Zero Information Loss" IEEE Transactions on Knowledge and Data Engineering, Vol. 5, No. 1, pp. 76-87, February 1993

[2] X. Chen, I. Petrounias, "A Framework for Temporal Data Mining" Database and Expert Systems Applications, Springer Berlin / Heidelberg pp. 796-805 Vol. 1460/1998

[3] S. Bose, L. Fegaras, "Data Stream Management for Historical XML Data", SIGMOD 2004 Paris, France, pp. 239-250, June 13-18, 2004

[4] R. F. Chong, X. Wang, M. Dang, D. R. Snow, "Understanding DB2 Learning Visually with Examples", 2nd edition, 2007

[5] "Oracle VARRAY Examples" http://www.dba-oracle.com/tips_oracle_varray.htm

[6] "Introduction to the DB2 Universal Database (DB2 UDB) audit facility" http://publib.boulder.ibm.com/infocenter/db2luw/v8/index.jsp?topic=/com.ibm.db2.udb.doc/admin/ c0005483.htm

[7] W. Lu, G. Miklau, "Auditing a Database Under Retention Restrictions", IEEE Inter. Conf. on Data Eng., ICDE 2009, pp. 42-53

[8] D. Luebbers, U. Grimmer, M. Jarke1, "Systematic Development of Data Mining-Based Data Quality Tools", proc. of the 29th VLDB Conference, pp. 548 - 559, Berlin, Germany, 2003

[9] "Understanding SQL Server Audit" http://msdn.microsoft.com/en-us/library/cc280386.aspx

[10] "Simple Data Auditing" http://forge.mysql.com/wiki/SimpleDataAuditing

[11] "Logging, Auditing, and Monitoring the Directory" http://download.oracle.com/docs/cd/B14099_19/idmanage.1012/b14082/logging.htm#i126963

[12] The PostgreSQL Global Development Group, "Postgre SQL 8.4.0 Documentation" http://www.postgresql.org/files/documentation/pdf/8.4/postgresql-8.4.0-A4.pdf

[13] N. H. Park, W. S. Lee, "Anomaly Detection over Clustering Multi-dimensional Transactional Audit Streams", IEEE International Workshop on Semantic Computing and Applications IWSCE 2008, pp. 78-80

[14] U. Supasitthimethee, T. Shimizu, M. Yoshikawa, K. Porkaew, "Meaningful Interrelated Object Tree for XML Keyword Search", 2nd International Conference on Computer and Automation Engineering, Singapore, pp. 339-344, January 26 – 28, 2010

[15] N. Waraporn, K. Porkaew, "Null Semantics for Subqueries and Atomic Predicates" IAENG International Journal of Computer Science, Vol. 35, Issue 3, pp. 305-313, September 2008

[16] Li Yang, "Teaching Database Security and Auditing", SIGCSE'09, pp. 241-245, March 4–7, 2009

Author Index

Aihan Yin 218
Aiyun Zhan 218
Benzheng Wei 174
Bin Tan 19,35,140,262
Bin Xiao 149
Bing Zhu 35,140,262
Bo Wang 43
Bo Yuan 35
Bo Yuan Bin Tan 133
Chao Huang 1,153
Chao Luo 43
Chuanliang Xia 250
Chunxiao Ye 210
Congfeng Jiang 197
Cui Zhang 124
Daguo Jiang 47
Danning Li 238
Dong Xu 161
Dongliang Jiao 214
Fan Guo 279
Fang Mei 144
Fawen Yu 210
Fei Zhu 157
Feng Luo 19
Fengchen Huang 186
Fengfan Zou 230
Fengjing Shao 234
Fengqi Yu 267
Florent Autrusseau 124
Fuyou Mao 31
Genghui Dai 108
Guangming Yang 267
Guangxue Yue 39,58,62,69,81,89,233
GuoSun Zeng 120
Guangying Yang 242,246
Guojia Yu 54
Haibo Hu 210
Haifeng Wu 58
Hong Xiang 210
Hongjun Diao 157
Hongmei Deng 5
Hongmei Xiang 66
Hua Ye 108
Hua Zhu 206
Hui Kang 144
Hui Liu 161
Huizhong Qiu 54,182
Jian Wan 197
Jianbin Zhao 206
Jianchao Zeng 85
Jianxun Liu 69
Jie Ma 93
Jie Zhao 193
Jigui Sun 112
Jilin Zhang 197
Jincheng Zhang 136
Jinfang Han 9
Jing Song 186
Jining Li 124
Jiqin Peng 9
Jun'e Liu 170
Junyu Dong 124
Kaiquan Wang 116
Kede Qin 81
Kun Hou 112
Lei Qian 31
Li Li 120
Li Wang 178
Lianzhong Liu 214
Libo Yang 223
Lifeng He 23
Lin Tian 54
Lin Xin 81
Linquan Xie 89

Liping Chen 69,89
Liping Zhao 254
Lizhong Xu 258
Meilian Qiu 97
Meiyu Zhang 77,230
Min Li 258
Min Tang 258
Min Wu 271
Ming Cheng 105
Ming Zhong 101
Narongrit Waraporn 283
Peiquan Jin 193
Peisong Ye 39
Qi Hu 77
Qiang Wang 27
Qin Wan 73
Qingmiao Zhang 218
Qingsheng Zhu 165
Qiushi Cong 136
Qiuwen Zhang 101
Rencheng Sun 234
Rong Wang 170
Rui Zhong 62
Ruilong Yang 165
Sanjun Liu 12
Shanxiao Yang 242,246
Shilong Ma 214
Shuang Yang 73
ShuangShuang Huang 50
Shuangshuang Zhang 144
Shuo Wan 182
Sijun Ma 112
Su Yan 275
Suping Peng 129
Tao Liu 129
Tieli Sun 112
Tingjian Zhong 97
Tinsen Huang 210
Wei Huang 157,186
Wei Yang 112
Weidong Zhao 58
Weifeng Du 12,206
Weihui Dai 58,108
Wenjuan Liang 105
Wenjuan Wang 58
Wenli Peng 201
Wenni Zhang 201
Wenping Guo 15,190
Xianghua Xu 197
Xiaoming Zhao 15,190
Xiaoni Wang 214
Xiaoning Song 234
Xiaoyu Wang 116
Xin Peng 174
Xingfu Wang 31
Xingyun Dai 108
Xinlian Zhou 149
Xinqiang Ma 238
Xixiang Zhang 69,81
Xizhi Zhu 227
Xu Zhou 50
XuanGe Peng 262
Xueying Liu 1,153
XuGuang Min 27
Xuyan Zhou 262
Xuzheng Liu 43
Yanheng Liu 144
Yanhui Zhou 27
Yi Huang 238
Ying Chen 15,190
Ying Wang 89,112
Yinghui Zhao 197
Yong Wei 73
Yonghua Pei 262
Youru Yu 5
Yu Li 77
Yu Xiang 254
Yueli Jiao 218
Yun Liu 1,153
Yunni Xia 165
Zaitao Liu 174
Zhan Guo 19

Zhan Guo 140
Zhangang Wang 129
Zhaosheng Zhu 31
Zhenghao Shi 23
Zhengjiang Wu 12
Zhengjiang Yi 12
Zhibin Xue 85
Zhihong Zhao 223
Zhipeng Xiao 5
Zhitang Li 93
Zhongbao Qin 201
ZhongYue Xiao 27
Zuming Xiao 19,140
Zuoming Huang 136