Proceedings

The Third International Symposium Computer Science and Computational Technology (ISCSCT 2010)

14-15, August 2010

Jiaozuo, China

Edited by

Youfeng Zou
Fei Yu
Zongpu Jia
Zichen Li

Co-Sponsored by

Henan Polytechnic University, China
Peoples' Friendship University of Russia, Russia
Feng Chia University, Taiwan
Zhengzhou University, China
Fudan University, China
South China University of Technology, China
Nanchang HangKong University, China
Jiaxing University, China
Academy Publisher of Finland, Finland

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-10CN007

ISBN 978-952-5726-10-7

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

The 2010 International Symposium Computer Science and Computational Technology (ISCSCT 2010) was the third in the annual series that started in Shanghai,China, during 20-22 Dec. 2008 and 26 - 28,Dec. 2009,Huangshan ,China.

Welcome to ISCSCT 2010. Welcome to Jiaozuo, China. The 2010 International Symposium on Computer Science and Computational Technology (ISCSCT 2010) is Co-sponsored by Henan Polytechnic University, China;Peoples'Friendship University of Russia,Russia;Feng Chia University, Taiwan;Zhengzhou University, China;Fudan University, China;South China University of Technology, China;Nanchang HangKong University,China;Jiaxing University,China;Academy Publisher of Finland, Finland. Much work went into preparing a program of high quality. The conference received 350 paper submissions from 7 countries and regions; every paper was reviewed by 2 program committee members; 191 papers have been selected as regular papers, representing a 54% acceptance rate for regular papers. From these 191 research papers, through two rounds of reviewing, the guest editors selected 29 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 ISCSCT 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. Paul Werbos, IEEE Neural Networks Pioneer Award Receipient,Program Director of USA National Science Foundation from National Science Foundation,USA;IEEE Fellow Prof. Gary G. Yen, President of IEEE Computational Intelligence Society from Oklahoma State University,USA;IEEE Fellow Prof. Derong Liu, Editor-in-Chief of IEEE Trans. on Neural Networks from Chinese Academy of Sciences,China;IEEE Fellow Prof. Jun Wang from Chinese University of Hong Kong, Hong Kong; IEEE & IET Fellow Prof. Chin-Chen Chang from National Chung Hsing University, Taiwan; Prof. Wenchang Shi from Renmin University of China and Graduate University of Chinese Academy of Sciences, China; Prof. Qin Keyun from Southwest 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 ISCSCT 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 Jiaozuo, 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 ISCSCT 2011.

Youfeng Zou

President of Henan Polytechnic University

Third International Symposium Computer Science and Computational Technology

Organizing Committee

Honorary Chairs

Paul Werbos, *National Science Foundation,USA(IEEE Fellow,IEEE Neural Networks Pioneer Award Receipient,Program Director of National Science Foundation,USA)*

Gary G. Yen, *Oklahoma State University,USA(IEEE Fellow,President of IEEE Computational Intelligence Society)*

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)*

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

General Chairs

Youfeng Zou,*Henan Polytechnic University, China*

Zongpu Jia,*Henan Polytechnic University, China*

Zichen Li,*Beijing Institute of Electronic Science and Technology*

Program Committee Chairs

Jian Shu,*Nanchang HangKong University,China*

KARPUS NIKOLAY, *Peoples' Friendship University of Russia, Russia*

Lishen Yang,*Henan Polytechnic University, China*

Weihui Dai, *Fudan University,China*

Yiqin Lu, *South China University of Technology, China*

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

Organizing Chairs

Changsen Zhang, *Henan Polytechnic University, China*

Guangxue Yue, *Jiaxing University, China*

Jiexian Zeng,*Nanchang HangKong University,China*

Jien Deng,*Henan Polytechnic University, China*

Lei Shi,*Zhengzhou University, China*

Yan Gao, *Henan Polytechnic University, China*

Finance Chairs

Yan Gao, *Henan Polytechnic University, China*

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

Publication Chair

Youfeng Zou,*Henan Polytechnic University, China*
Fei Yu, *Peoples' Friendship University of Russia, Russia*
Zongpu Jia,*Henan Polytechnic University, China*
Zichen Li,*Beijing Institute of Electronic Science and Technology*

Local Organizing Chairs

Ziyi Fu,*Henan Polytechnic University, China*
Zhibin Zhang,*Henan Polytechnic University, China*

Workshop Chair

Yan Gao, *Henan Polytechnic University, China*

Secretary General

Fei Yu, *Peoples' Friendship University of Russia, Russia*
Jiyu An,*Henan Polytechnic University, China*

Third International Symposium Computer Science and Computational Technology

Committee Members

Prof. Changsen Zhang, Henan Polytechnic University, China
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, Institute of Automation, Chinese Academy of Sciences,China
Prof. Dingguo Wei, Guangdong University of Business Studies, China
Prof. Gao Yan, Henan Polytechnic University, China
Prof. Gary G. Yen, Oklahoma State University,USA
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. Jien Deng,Henan Polytechnic 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 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. Lishen Yang,Henan Polytechnic University, China
Prof. Martha Russell, Stanford University, USA
Prof. Ming LI, Nanchang HangKong University, China
Prof. Naiping Hu, Qingdao University of Science & Technology, China
Prof. Paul Werbos, National Science Foundation,USA

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 Technology, 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. Youfeng Zou,Henan Polytechnic University, 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. Zhibin Zhang,Henan Polytechnic University, China
Prof. Zhijian Wang, Guangdong University of Business Studies, China
Prof. Zichen Li,Beijing Institute of Electronic Science and Technology, China
Prof. Ziyi Fu,Henan Polytechnic University, China
Prof. Zongpu Jia,Henan Polytechnic University, China

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 001-004

Fast recognition based on color image segmentation in mobile robot

Liu Hai-bo [1], Wang Yu-mei [1], Dong Yu-jie[2]

[1] School of Electrical Engineering and Automation, Henan Polytechnic University, Jiao Zuo, China, 454000
Email: liuhaibo09@hpu.edu.cn
[2].WanFang College of Science and Technology, Henan Polytechnic University, Jiao Zuo, China, 454000
Email: lhb_1403@foxmail.com

***Abstract*—Real time segmentation is the first step in the color vision system on the robot system.A color image segmentation method using improved seed-fill algorithm in YUV color space is introduced in this paper. The new method dramatically reduces the work of calculation,and speeds up the image processing. The result of comparing it with the old method based on RGB color space was showed in the paper.The second step of the vision sub system is identification the color block that separated by the first step.A improved seed fill algorithm is used in the paper.The implementation on MiroSot Soccer Robot System shows that the new method is fast and accurate.**

***Index Terms*—autonomous mobile robot, image segmentation,target recognition,threshold**

I. INTRODUCTION

Robot soccer has developed into an adversarial game project in recent years.It comprehensively utilizes several fileds of theory and technology such as mechanics, electronics, control and pattern process. In the standard game of 5 to 5, both sides have a computer controller. First of all, the camera hanging above the venue shoots the game screen and sends the screen to the controller via image acquisition card. Secondly, the shooting pictuers are identified by an image recognition module to get the 11 target positions and orientation angles of the players of both sides and together the football.Then, according to these information, the decision-making procedures make decisions so that each robot car of its side can get its speed instructions. Finally, these instructions are sent to the cars on the pitch by radio communication module, and according to these instructions, the cars can play football.

In soccer robot colour-vision system, the procedures can identify the robots that which team it belongs to and which number it is by the colour labeled on the cars. Each robot has two colours, one of which is team colur, the other is ID colur. Therefore ,the first step of the identify work is classify each image pixel to different discrete colour class according to its colour.

The common methods which are used in colour classsification are linear color threshold value method, recent domain method and threshold vector method, etc.Among them, the linear colour threshold value method is to use linear graphic to segment the colour space, and the determination of the threshold value can be obtained by directly taking threshold value and by getting the target colour range using automatically training method. Also, in order to achieve a proper threshold value, we can use the methods of neural networks(NNs) and multivariate decision trees(MDTs) for self-learning. While the recent domain classification is used for image segmentation, it utilizes the method of membership functions, which determinates a colour belonging to which class according to the maximum membership degree. However, when the threshold vector method is used, it first segments the colour space into cuboid by using a set of certain threshold value, then it determines the pixel's colour by checking the position of the pixel's colour value in the space.

After colours' classification, it is should that every point of colour classes be disposed. Then we can identify the positions and the orientation angles of each player and the football.When the positions and angles is being identified, the common method is to scan all of the pixels which have been classified. This method is a method which pieces together the same colour of the adjacent pixels. However, the calculation work of this method is quite large since with this method, almost all the pixel points are to be disposed.

An identification method based on the regional projection algorithm is presented[4]. This method use the geometric method to calculate the center coordinates and the orientation angle(A angle between the robot car's forward direction and x axis positive direction). In practical application, I find that although the method has rapid identification speed, it needs the image to be very clear. Or the vertex coordinate errors will affect the precision, especially have great influences on the orientation angles. In the soccer robot system, for most cameras are common monitoring cameras, the clarity of the image is quite limited, therefore, this method is also difficult to be applied in practice.

Combining the high requirements of the soccer robot system, this paper uses a new improved colour determination method based on the threshold vector. This method can distinguish several colours through one operation. At the same time, the method uses an improved seed-fill algorithm. So it dramatically reduces the work of calculation. Besides, this method not only has no high requirements of the images, but also it has high precision.

Your goal is to simulate the usual appearance of papers in a Conference Proceedings of the Academy Publisher.

AP-PROC-CS-10CN007

We are requesting that you follow these guidelines as closely as possible.

II. RECOGNITION METHOD OF COLOR IMAGES

Color Image Transformation

The colour threshold values selected in this thesis are in 3-dimension colour space. There many kinds of colour spaces, and we commonly use Hue,Saturation and Intensity space, that is HIS space, YUV space and RGB (Red,Green,Blue) space.

The most common color space is using R,G,B three colours' mixture of different proportion to represent all kinds of colours. But the RGB colour model is easily affected by the sun-light. Because the sun-light of different positions have great differences in the venue, one colour's RGB of different positions also has great changes. Thus, we can't judge the colours by a set of single threshold vule. Therefore, the robustness of the identification procedures will be meaningless.

Different from RGB colour space, the HSI and YUV colour space use two-dimension to present the spectrum, and use the third dimension to present the colour's intensity. For example, in the HSI colour space, H and S present the colour information, while I present the intensity. Compared to RGB mode, these two kinds of colour spaces are superior to adapt to the changs of sun-light intensity. As a result, we often transform the RGB signals into HSI signals or YUV signals.

The YUV colour space is adopted in the MiroSot soccer robot vision system. The RGB images achieved from the camera can be transformed into the values of Y,U,V in the procedure by using the following formula：

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.169 & -0.331 & -0.500 \\ 0.500 & -0.419 & 0.081 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

B. Determination of Threshold and Color

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin in this template measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

The method of the colour's judgement in this thesis can be used to judge various colors in the colour space. Usually, every colour class can be described by 6 threshold values: there are two values in every dimension, and the two values present the maximum value and the minimum value respectively.

When we are determining the threshold value, first the upper and lower threshold of every dimension are determined in the colour space by taking samples. As is shown in figure 1, the 3 colour vectors(Y,U,V) can determine a cuboid in the colour space. When the cuboid can include the position of a pixel to be judged in the colour space, we regard the pixel as the colour to be found.

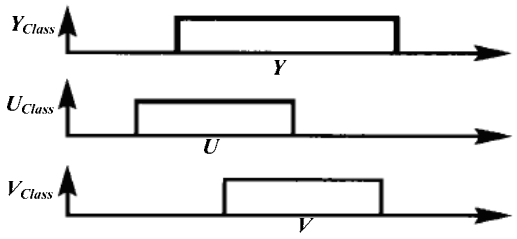


Figure1. The problem of two threshold vules in three-dimension color space

When we are judging the colours of the pixels, the colours to be identified can be put on the brightest parts and the darkest parts in the venue for sampling so that the influence of the sun-light can be minimum.

In MiroSot robot soccer system, because of its high real-time requirements, it must use the identification methods which have small amount of calculations and high speed. Thus, we select the methods that directly determine the threshold values.

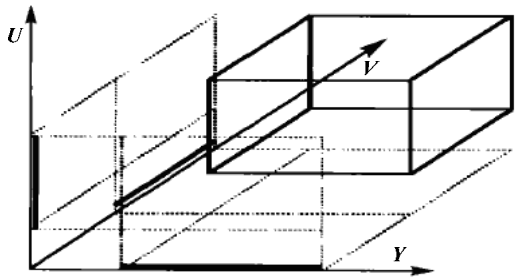


Figure2. The color space in the YUV space

Usually, after we have determined the 6 threshold of colour class, we can use the following method to judge whether a pixel is a member of one colour class.

if(($Y \geq Y_1$ lower threshold)
AND ($Y \leq Y_1$ upper threshold)
AND ($U \leq U_1$ lower threshold)
AND ($U \leq U_1$ upper threshold)
AND ($V \leq V_1$ lower threshold)
AND ($V \leq V_1$ upper threshold))
pixel colour=colour class 1;

This method requires 6 judgement sentences to process a pixel, while for a frame image of NTSC pattern, there are 640×480 pixels to be processed. In addition, it is just for colours. So the amount of processing is quite large. In the game, the positions of the robot and the ball are distinguished by different colours, therefore, it needs to distinguish 8 colours at least in the mean time. Such a large amount of calculation makes the processing efficiency quite low, no matter what kinds of computer hardware are used. As a result, it's difficult to satisfy the real-time requirements.

In the YUV colour space, since the value Y represents the brightness, and it has obvious changes, therefore, in this paper, we only consider the values U and V, that is,

the taking of colour judgement is based on the values U and V. When the colours are being judged, we should establish the threshold vectors of U and V first. As the values U and V are between 0 and 255, every vector has 256 elements. For example, by taking sample, the colour yellow's value U is between 7 and 9, its value V is also between 7 and 9, the resulting vectors respectively are:

$$U_{class}[] = \{0,0,0,0,0,0,0,1,1,1,0\cdots\};$$

$$V_{class}[] = \{0,0,0,0,0,0,0,1,1,1,0\cdots\};$$

So, when we judge a pixel in the image is whether or not the target colour that needs to be identified, we only need to take the colour's value of this point Bit AND calculations with above determinated vector, then we can get the results. For example, the values Y,U and V of one pixel are {1, 8, 9}, the follow procedures can be used for judgement: if the result of $U_{class}^{[8]}$ AND $V_{class}^{[9]}$ is 1, then the pixel is the colour that needs to be identified.

This method can also be used in various colors, and in order to judge what colour it belongs to, it also needs to take one time Bit AND calculation. When a variety of colours that need to be identified, for instance, we have to distinguish the blue colour except the yellow colour. By sampling, the U,V's threshold vectors of blue are:

$$U_{class}[] = \{1,1,1,0,0,0,0,0,0,0,0\cdots\};$$

$$V_{class}[] = \{0,0,0,1,1,1,0,0,0,0,0\cdots\};$$

We combine together the vectors of the two colours ,then we get a vector whose every element has two bit:

$$U_{class}[] = \{01,01,01,00,00,00,00,10,10,10,00\cdots\};$$

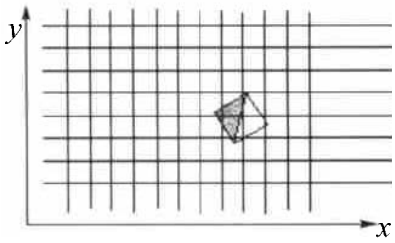
$$V_{class}[] = \{00,00,00,01,01,01,00,10,10,10,00\cdots\};$$

Similarly, for a data point whose values Y,U,V are {1,8,9}, when the result of $U_{class}^{[8]}$ AND $V_{class}^{[9]}$ is 10, then we can get a judgement that this point is yellow other than blue.

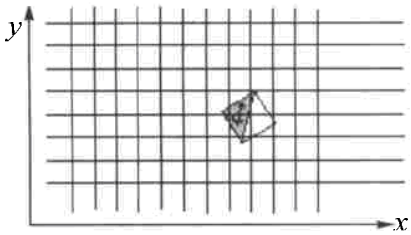
In real procedures, each element of the threshold vector has 8 bits, that is 1 Byte, in all 255 Byte. As a result, it can judge 8 colours simultaneously at most, which also satisfy the requirements of MiroSot 5 to 5 robot soccer game.

Ⅲ IMPROVED IMAGE SEGMENTATION METHOD

It used seed-fill algorithm in area filling procedure of paper, but improved it. The whole area filling of the improved seed-fill algorithm is synchronized to the judge of pixel color. In the beginning, the image is divided into several small pieces, obtain its center from each small pieces to judge color(the small pieces is 1/4 of the field ball generally, about 8×8 pixels, calculation is 1/64 of the original). When the center is the color to identify, it taked the center of the seed, spred around and judged the color of pixel around again until the whole color pieces is filled. It calculatied the center of gravity in filling time. In this process, it should identify and filter processing simultaneously, that,this piece is judged respectively to identify the target according to the coordinates of each pixel, if the distance between spread point and gravity point of filling color pieces has exceeded the size of the car, it is not the target and can be filtered out.The color recognition is carried out by class, firstly, identified yellow, blue, orange,they are team color and ball color respectively. When our team color is found out,it tried to find ID color around the team color and the center of gravity of ID color by using seed-fill algorithm.So it can improve the speed of recognition, and guarantee higher accuracy. The recognition process is shown in Figure 3.



(a) Color judgement of grid center



(b) Spread around when found color of the first object point

Figure3. The process of image segmentation and seed-fill algorithm

After obtaining the center of gravity of team color pieces and ID color pieces, its midpoint is the car's center of gravity. Connection direction is rotated 45° ,which is the direction angle(between the car direction and x-axis positive direction).

Ⅳ THE APPLICATION OF THE IMPROVED IMAGE SEGMENTATION IN SOCCER ROBOT SYSTEM

When we apply the improved image segmentation into robot soccer system, first, in the MiroSot robot soccer system, we use a pattern camera of NTSC whose frequency is 30 Hz to shoot pictures. Then, the image signal digitization is input to the processing computer which is equipped with pentium 4 whose primary frequency is 1.5GHz via Matrox MeteorII acquisition card. Besides, the computer has 256M RAM, its operation system is Windows98, and it use VC++ 6.0 to write procedures. The processed results based on the identification method of RGB and YUV colour space are shown in figure 5 and figure 6. The results shown in figure 5 and figure 6 are respectively segmentation result of the image. The figures which are shown are the segmentation target, and theirs backgrounds having been filtered, are black.

From figure 6, we can find that the position of the two cars are different, and the sunlight is also different. Using the method based on YUV colour space can better eliminate the light's interference, while the car in highlight will not be found when the method based on RGB is used. At the same time, the method based on RGB will take some miscellaneous points as effective ponits.

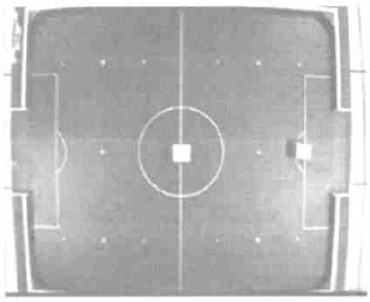


Figure4. The original robot soccer game image

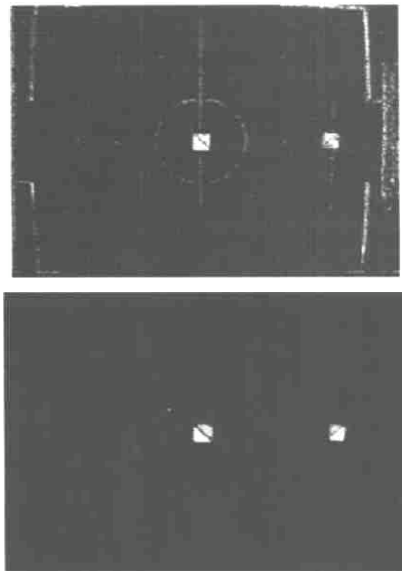


Figure6. The processing image using the identification method based in YUV colour space

In the recognition process, for the cars in the image, both of the methods have good recognition. For instance, whether you use the RGB method stilll the YUV method, that the position of the car whose real position is on the point x=0, y=0, θ=0,left you will both get is x=0, y=1, θ=1. While for a car whose real position is on the point x=60, y=0, θ=0 right, the position you will get is x=59, y=1, θ=0 when you use the YUV method. Practice has proved that the target lose very serious and the orientation angle's deviation will be above ±100 degree when the identification method based on RGB is used, so it can't identify targets effectively.

In the recognition rate, when you combine the method based on YUV colour space and the improved seed-fill algorithm together, it will take 15～20 ms to identify the 5 players and the ball in the venue, while it needs to take more than 50 ms when the traditional scanning method is used.

V CONCLUSION

The results prove that: when we apply the improved seed-fill colour image segmentation algorithm based on YUV colour space into real robot soccer system, it will not only have strong robustness over the changing sunlight, but also can it quickly complete image recognition under the conditions of ensuring the accuracy.

REFERENCES

[1] James Bruce, Tucker Balch, Manuela Veloso, "Fast and inexpensive color image segmentation for interactive robots," In: Proc IEEE/RSJ Intermation Conference on Intelligent Robots and System[C]. *Takamatsu. Japan*, pp. 2061–2066, 2000.

[2] Jain R, Kasturi R, Schunck B G. Machine vision[M]. New York: McGraw Hill,1995.

[3] Brown T A, Kopkwitz J. The weighted nearest neighbor rule for class dependent sample sizes[J]. IEEE Trans on Information Theory,1979,IT-25(5):617-619.

[4] Brodley C E, U tgoff P E Multivariate decision trees[J]. Machine Learning,1995,19(1):45-47.

[5] Grillo E, Matteucci M,Sorrenti G.D,Grtting the most from your color camera in a color-coded world[A]. D.Nardi et al. editors,RoboCup2004[C].Springer,2005:221-235.

[6] Maritinez-Gomez LA,Weitzenfeld A.Real time vision system for a small size league team[A],Proc.1st IEEE-RAS Latin American Robotics Symposium[C].Mexico City:ITAM,2004:343-349.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 005-008

Regional differentiation based on permutation entropy and its geographical explanation

C.Y. Hao[1], T.Q. Zhao[2]
[1]College of Surveying & Land Information Engineering, Henan Polytechnic University, Jiaozuo Henan Province, China
e-mail: haocy@hpu.edu.cn
[2]College of Resources and environment, Henan Polytechnic University, Jiaozuo Henan Province, China
e-mail: zhaotq@hpu.edu.cn

***Abstract*—The study area, located in the southwest of Yunnan Province, has a distinct regional difference in both topography and climate system. Western region belongs to the Southwestern Yunnan mountainous area with varied landforms while eastern region is a part of Eastern Yunnan Plateau whose terrain is rather flat relatively. And different regions are influenced by different climates, which have resulted from the multiplicity of monsoon systems and the significance of great topographic effects. By a complexity parameter--permutation entropy for time series from 1971 to 2000 based on comparison of neighboring values, the tolerance to the complexity of climate systematic is simple and effectual. And there is a good geographical explanation of both atmospheric circulation and localized topography to climate elements.**

***Index Terms*—permutation entropy, regional differentiation, atmosphere ciruculation, climate complexity, topographic effect**

I. INTRODUCTION

Regional differentiation, with big difference betweeen regions and rather inter-region similarity, belongs to the issue of cluster analysis[1]. In general, its traditional method is regional division based on understanding regional synthesis characteristics. And the next is to explore natural environmental characteristics and principal developmental laws between regions. However, it is not an easy thing to understand the regional integrated characteristics of different units. So the research on regional differentiation methodology has become one of hotspots in theoretical geography[2]. Then, the opportunity comes to here that some new ways or means are applied in regional differentiation.

Nowadays, phenomenon complexity has already become one of the most important issues in both social and natural science fields[3]. Entropy reflects the figure of micro-scale or thermodynamics probability and has direct proportion to the logarithm of thermodynamics probability, which was only used to describe the disorder degree of thermodynamics system. The more confused state, the greater the probability of thermodynamics is. Therefore, the entropy is the measurement of systematic complexity[4]. Among so many entropic calculation methods, permutation entropy, which is shorted with permutation entropy in the following passage, is a complexity parameter for long time series based on the comparison to neighbouring values. Simplicity, extremely fast calculation ability and practicality are the advantages of permutation entropy[5]. Based on permutation entropy, disease or health degree had already succeeded in being diagnosed from the complicated data of heart or brain conditions in medical field[6]. However, there are only a few applications of it on geological field by now[7]. Rind has opened up the road to the study on regional complex of varied terrains and landforms in the field of meteorology since he wrote an article entitled the complexity of topography and climate in Science[8]. And a Chinese scholar named Hou Wei has quantitated the process of abrupt climate change and provided a new way to the research of the regional climatic change by analyzing the day-to-day temperature time series of north China from 1960a to 2000a based on fifty-two meteorological stations[9]. But till now, there is no measurement of regional differentiation using permutation entropy[10]. Here, based on daily average temperature and daily total precipitation from 1971 to 2000 in the middle-south of Yunnan Province which is a typical complicated climate region in China, this study attempts to measure the complexity of climatic system, and probes into it in a geographical way.

II. STUDY AREA AND METHODS

A. Study Area

The study area locates in the middle-south of Yunnan Province, between 98º40′53″-106º11′33″E and 22º26'34″-24º27'35″N with a total area of 101 900km^2, including thirty county level administrative regions. In its western region, as a part of Southwestern Yunnan mountainous area, there are a series of longitudinal range-gorges including Laobie Mountain-Nandinghe River, Bangma Mountain-Lancangjiang River, Wuliang Mountain-Babianjiang River, and Ailao Mountain-Yuanjiang River. Among them, Ailao Mountain, with the highest peak exceeding 3100m, is a huge mountain from northwest to southeast in the middle part of the study area. The direction of Ailao Mountain is perpendicular to the air current of Southwest monsoon. Comparatively, the east of the study area has relatively gentle physiognomy, although it has been eroded by Yuanjiang River, Nanpanjiang River and their branches[11]. And the study area is influenced by two water vapour sources of subtropical seas (the Bay of Bengal and South Sea of China) at the same time; especially the western region lies in the southeast edge of the Tibetan Plateau, and is in the way of the southwest summer monsoon coming to

China[12]. In dry season (from November to April of next year), the study area is controlled by Plateau winter monsoon and the southern branch of westerly, and eastern part of region is influenced by East Asia winter monsoon, while in rainy season (from May to October), the study area is influenced by southwest summer monsoon and southeast summer monsoon. Namely, there is an obviously seasonal alternation between dry and rainy seasons[13]. In addition, the topography of the study area is complicated and various, and its main longitudinal ranges and gorges are nearly right angle with main air currents. The climate of the study area has remarkable particularity because of the factors mentioned above, and some synoptic climatic features are exclusive for our whole country, no matter in winter or summer. The three meteorological front subjects in the world, including low latitude problem, great topographic effect problem and tropical ocean problem, all appear remarkably in this region at the same time[14].

B. Definitions of Permutation Entropy

For practical purposes, $n=3\ldots7$ were recommend, and a series with seven values: $x=(4,7,9,10,6,11,3)$ was taken as an example[15]. Six pairs of neighbours were then organized according to their relative values, in which four pairs were $x_1<x_{t+1}$ and two pairs were $x_t>x_{t+1}$. The four pairs of values were represented by the permutation *01* ($x_1<x_{t+1}$) and two were *10*. The permutation entropy of order $n=2$ as a measure of the probabilities of the permutations *01* and *10* were defined. Thus there had the following formula:

$$H(2)=-(4/6)\log(4/6)-(2/6)\log(2/6)\approx0.918 \quad (1)$$

As usual, log is with base 2, H is given in bit. Then, three consecutive values were compared. (4,7,9) and (7,*9,10)* represent the permutation *012* since they are in increasing order, *(9,10,6)* and *(6,11,3)* correspond to the permutation *201* since $x_{t+2}<x_t<x_{t+1}$, and *(10,6,11)* has the permutation type *102* with $x_{t+1}<x_t<x_{t+2}$. The permutation entropy of order $n=3$ is

$$H(3)=-2(2/5)\log(2/5)-(1/5)\log(1/5)\approx1.522 \quad (2)$$

So, the definition of permutation entropy is as bellowed. Consider a time series $\{x_t/\ t=1,\ldots,T\}$, we study all *n!* permutations π of order n which are considered here as possible order types of n different numbers. For each π relative frequency was determined.

$$P(\pi)=(\#\{t/t\le T\text{-}n,\ (xt+1,\ \ldots,\ xt+n) \text{ has type}\pi\})/(T\text{-}n+1) \quad (3)$$

The frequency of π was estimated as good as possible for a finite series of values. To determine $p(\pi)$ exactly, an infinite time series $\{x_1,x_2,\ldots\}$ was assumed and the limit for $T\to\infty$ in the above formula was taken. This limit exists with probability *1* when the inner stochastic process was accord with a very weak stationary condition: for $k\le n$, the probability for $x_t<x_{t+k}$ should not depend on *t*.

The permutation entropy of order $n\ge2$ is defined as

$$H(n)=-\sum p(\pi)\log p(\pi) \quad (4)$$

Where the sum runs over all *n!* permutations π of order n. This is the information contained in comparing n consecutive values of the time series. It is clear that $0\le H(n)\le \log n!$ where the lower bound is attained for an increasing or decreasing sequence of values, and the upper bound for a completely random system where all *n!* possible permutations appear with the same probability. The time series presents some sort of dynamics when $H(n)<\log n!$. Actually, in our experiments with chaotic time series $H(n)$ did increase linearly with n. So it seems useful to define the permutation entropy per symbol of order n, dividing by $n-1$ since comparisons start with the second value:

$$Hn=H(n)/(n-1) \quad (5)$$

By definition of entropy, permutation entropy, as a parameter reflecting the nature of system, could measure the systematic dynamic complexity. So, when applied it in climate system, the larger entropy value, the higher the complexity or randomness of system is.

C. Calculations of Permutation Entropy

By Statistica Neural Networks software that Stastisoft Company produced in MATLAB 6.1 (MathWorks Inc. Copyright 1984-2001), permutation entropies of temperature and precipitation have been calculated. The value of temperature permutation entropy is from 0.9526 to 1.0245 while precipitation permutation entropy is from 0.7060 to 0.9210 (table Ⅰ).

III. RESULTS

A. Spatial Pattern of Permutation Entropies

From the spatial distsribution of temperature permutation entropy (the topside in figure one), the weather stations with larger entropies (larger than one) are distributed in the west side of Ailao Mountain while thoses with smaller entropies (smaller than one) are mainly in the eastern region. According to the definition of permutation entropy, the complexity of temperature system of the wester region is higher than that of the eastern region. Two weakening processes of precipitation permutation entropy are detected. The one is from west to east in the whole study area while the other is from southwest to north or east in the east side of Ailao Mountain (the underside in figure one). All these are tightly related with general atmospheric circulation and mid-latitude mountains in study area.

B. Spatial Analysis

At the whole scale, both temperature and precipitation permutation entropies are decreasing from west to east, which means the direction of atmospheric circulation system should be from west to east. So be it. Both the southern branch of westerlies trough and Tibetan Plateau winter monsoon have made temperature gradient point to

east in winter while the Southwest summer monsoon mainly comes from west in summer[16]. Therefore, the influence degree on temperature or precipitation in western region is deeper than in eastern region. And as a result of remarkable topographical effect on both rainfall and cold wave block, entropies spatial gradient have a signification correlation with the position of longitudinal ranges, especially temperature permutation entropy (the topside in figure one).

TABLE I. PERMUTATION ENTROPIES OF DIFFERENT WEATHER STATIONS IN STUDY AREA

Stations	LONG.	Lat.	T-PE	P-PE	Stations	LONG.	Lat.	T-PE	P-PE
Zhenkang	24.06	98.96	1.0245	0.8701	Yuanyang	23.16	102.75	0.9594	0.8759
Yongde	24.15	99.23	1.0022	0.8297	Jianshui	23.61	102.83	0.9937	0.7315
Cangyuan	23.15	99.26	1.0196	0.8402	Gejiu	23.38	103.15	0.9767	0.8151
Gengma	23.55	99.40	1.0143	0.8303	Jinping	22.78	103.23	0.981	0.8798
Shuangjiang	23.46	99.80	1.0166	0.8019	Mengzi	23.38	103.38	0.9818	0.7656
Lincang	23.95	100.21	1.0106	0.8129	Pingbian	22.98	103.60	0.9526	0.9210
Jinggu	23.50	100.70	1.0149	0.8344	Hekou	22.50	103.95	0.9578	0.8702
Zhenyuan	23.88	100.88	1.0122	0.8000	Qiubei	24.05	104.18	0.9891	0.8362
Puer	23.03	101.28	1.0080	0.8194	Wenshan	23.38	104.25	0.9706	0.7881
Mojiang	23.43	101.71	1.0074	0.8050	Yanshan	23.61	104.33	0.9675	0.8046
Xinping	24.06	101.96	0.9935	0.7450	Maguan	23.02	104.41	0.9701	0.8900
Yuanjiang	23.43	101.98	0.9847	0.7060	Xichou	23.45	104.68	0.9584	0.8682
Lvchun	23.00	102.40	0.9792	0.8457	Malipo	23.13	104.70	0.9585	0.8450
Honghe	23.36	102.43	0.9550	0.7268	Guangnan	24.06	105.06	9737	0.8137
Shiping	23.70	102.48	0.9883	0.7494	Funing	23.65	105.63	0.9745	0.8237

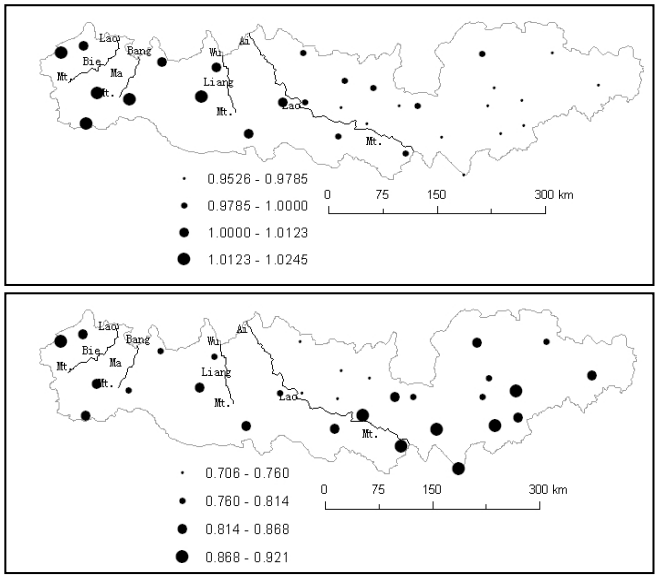


Figure 1. Spatial distributions of permutation entropies based on temperature(topside) and precipitation(underside)

At the regional differentiation, there is a significant contrast between western region and eastern region in whether temperature or precipitation permutation entropy. In western region, no matter of temperature or precipitation permutation entropy, its value is in a process increasing-decreasing with longitudinal ranges from west to east, again and again. The spatial distribution of permutation entropies are closed bound up with these longitudinal ranges, such as Laobie Mountain, Bangma Mountain, Wuliang Mountain and Ailao Mountain, i.e., there is instinct difference between high mountain and deep valley, between rainfall area and rain shadow area in surface terrain. So their relevance has showed that a huge mountain has a stronger effect on reallocating moisture and heat spatially. In eastern region, temperature permutation entropy is in a weakening process while precipitation permutation entropy is in a strengthening process from north to south. These conditions are resulted from both air current systems and mid-latitude mountains. Besides the southern branch of westerlies trough and Tibetan Plateau winter monsoon as already stated in our previous article, the eastern region is affected by East Asia winter monsoon characterized by strong cold coming from the north. So its northern area is characteristic of temperature system complexity, which results in high temperature permutation entropy in northern area, but low in southern area. In summer, it is under the control of East Asia monsoon coming from the South China Sea while it is influenced by the southwest summer monsoon at the same time. So the order of precipitation permutation entropy has a process of weakening from south to north.

Of the comparison between temperature and precipitation permutation entropies in spatial changes, the former is in a weakening process from west to east, especially nearby Ailao Mountain, but the latter has a process of weakening in eastern region from southwest to east or north. There are two factors to explain these. Firstly, the temperature spatial change is significantly related to the conversion of flow path in a year. The temperature permutation entropy has a process of weakening from west to east because there are three kind winds in a year, such as the southern branch of westerly flow, the Tibetan Plateau monsoon and the Southwest summer monsoon, all coming from the west. Secondly, the precipitation spatial change is mainly related to the direction of water source. There are two water vapour sources, the Bay of Bengaland South China Sea[17]. The former flows gradually across four longitudinal ranges in the sequence so that the precipitation permutation entropy does not change significantly because of their continuous rainfall interception. But in eastern region, especially along Ailao Mountain from southeast to northwest, the precipitation permutation entropy weakens gradually, which demonstrates fully the great topographical effect of Ailao Mountain on precipitation.

It was because of the complexity of several air current systems and the great topographic effects that both daily temperature and precipitation permutation entropies have demonstrated unique spatial patterns in the study area. In a word, permutation entropies based on the meteorological data for a long time series can reflect the systematic spatio-temporal complexity of regional climate elements, and give a good explanation for them.

IV. CONCLUSIONS AND DISCUSSIONS

Firstly, the spatial pattern of daily temperature permutation entropy can reflect the systematic complexity of regional general atmospheric circulation and the significance of great topographic effects. The weakening trend and its intensity of daily temperature permutation entropy have indicated the remarkable barrier functions of four longitudinal ranges in the study area. And in the east side of Ailao Mountain, the weakening process from north to south of permutation entropy have not only reflected the great topographic effects, but also showed the range and degree of East Asia winter monsoon influenced.

Secondly, the spatial pattern of daily precipitation permutation entropy can not only indicate the direction of vapour flowing, but also reflect the variation intensity of precipitation. The weakening process from west to east of daily precipitation permutation entropy has showed the topographic effects of longitudinal ranges and the direction of water vapour source. The South Sea vapour source has affected less and less from south to north in the eastern region, which is related with the smooth terrain of Yunnan Province Plateau.

Thirdly, permutation entropy based on a long time series meteorological data can reflect the systematic spatio-temporal complexity of regional climate elements. And most importantly, there is a good geographical explanation for them.

Lastly, by calculating permutation entropy based on meteorological data for time series, this research succeeds in carrying on the measurement of climate systematic complexity and its geographical explanation. But there are still a few of interesting questions to probe into. For example, for a region characterized by a relatively complicated climate system with remarkably topographic effects, its permutation entropies have remarkable spatial change characters. But what kind of situation would be in the case of single atmospheric circulation and flat terrain region? In addition, the time series of meteorological data in this research is from the year 1971 to 2000. Obviously, the length of time series would influence the entropy values, and therefore there should be a suitable time length in a certain study. Moreover, daily meteorological data seems to be overstaffed and random, what would be likely if taken ten-day or monthly data instead? All these questions are to be worth discussing.

ACKNOWLEDGMENT

This work was supported by the National Science and Technology Issues of Supporting Projects (No. 2006BAJ10B06-04) and Soft Science Project of He'nan Province (No.082400440750).

REFERENCES

[1] D. Zheng, Q. Y. Yang , M. C. Zhao, J. Z. Li, and F. H. Wang, *The Research on Regional Physical Geographical System*. Beijing, China: China Environmental Science Press, 1997.

[2] C.Y. Hao, S.H. Wu, and S. C. Li, "Study on the Method of Areal Differentiation Based on SOFM," Progress in Geography, vol. 27, pp.121-127, May 2008.

[3] E. Takimoto and A. Maruoka, "Top-down decision tree learning as information based boosting," Theoretical Computer Science, vol. 292, pp.447-464, Jannuary 2003.

[4] D. L. Camillo and R. Tristan, "The rectifiability of entropy measures in one space dimension," Journal Math Pures Application, vol.82, pp.1343-1367, October 2003.

[5] M. A. Jose, B. K. Matthew, and K. Ljupco, "The permutation entropy rate equals the metric entropy rate for ergodic information sources and ergodic dynamical systems," Physica D, vol.210, pp.77-95, October 2005.

[6] T.U. Schwartz, R. Walczak, and G. Blobel, "Circular Permutation as a Tool to Reduce Surface Entropy Triggers Crystallization of the Signal Recognition Particle Receptor β Subunit," Protein Science, vol.13, pp.2814-2818, July 2004.

[7] C. Bandt, G. Keller, and B. Pompe, "Entropy of interval maps via permutations," Nonlinearity, vol.15, pp.1595-1602, Septermber 2002.

[8] D. Rind, "Complexity and climate," Science, vol.284, pp.105-107, April 1999.

[9] W. Hou, G. L. Feng, W. J. Dong, and J. P. Li, "A technique for distinguishing dynamical species in the temperature time series of north China," Acta Physica Sinca, vol.55, pp.2663-2668, May 2006.

[10] J.A. Rial, "Abrupt climate change: chaos and order at orbital and millennial scales," Global and planetary change, vol.41, pp.95-109, April 2004.

[11] S.Y. Wang, *Geography of Yunnan*. Kunming, China: Yunan Nationality Press, 2002.

[12] L. X. Chen, Q. G. Zhu, H. B. Luo, J. J. Hai, M. Dong, and Z. Q. Feng, *East-Asia Monsoon*. Beijing, China: China Meteorological Press, 1991.

[13] X. M. Qiang, J. H. Jv, and H. H. Zhang, "A Diagnostic Analysis of the Summer Monsoon in Yunnan," Journal of Yunnan University (Natural Sciences), vol.20, pp.75-79, January 1998.

[14] J. Qin, J. H. Jv, and M. E. Xie, *Weather & Climate over Low Latitude Plateau*. Beijing, China: China Meteorological Press, 1997.

[15] C. Bandt, B. Pompe, "Permutation Entropy: a Natural Complexity Measure for Time Series," Physical Review Letters, vol.88, pp.174102.1-174102.4, 2002.

[16] D. Zheng, *The Formation, Environment and Development of the Tibetan Plateau*. Shijiazhang, China: Hebei Science & Technology Press, 2003.

[17] K. Y. Zhang, "The Characteristics of Mountain Climate in the North of Ailao MTS," in *Research of Forest Ecosystem on Ailao Mountains, Yunnan*, Laboratory of Ecology Kunming Branch, Academic Sinica, Eds. Kunming, China: Yunnan Science and Technology Press, 1987, pp. 20-29.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 009-013

A Reliable Time Synchronization Protocol for Wireless Sensor Networks

Fuqiang Wang[1,2], Peng Zeng[1], Haibin Yu[1] and Xiaoquan Zhao[1]
[1] Key Laboratory of Industrial Informatics Shenyang Institute of Automation, Shenyang, China
Email: wangfuqiang@sia.cn
[2] Chinese Academy of Sciences & Graduate School of the Chinese Academy of Sciences Beijing, China
Email: { zp & yhb & zhaoxiaoquan2000}@ sia.cn

***Abstract*—Reliability is crucial for time synchronization in Wireless Sensor Networks (WSNs). Existing time synchronization algorithms provide on average good synchronization in laboratory environment, however, outdoor environment associated with radio interference will influence the performance of time synchronization. In this paper, we proposed a Reliable Time Synchronization Protocol (RTSP) which is designed to adapt topology changes due to link failures or node mobility. RTSP works in a level fashion: each level means the nodes in this level have the same hop. MAC layer time-stamping and linear regression is adopted to compensate clock drift. Compensation mechanism and random time source choice mechanism is introduced to make synchronization robust against link and node failures. The protocol is implemented on the SIA2420 platform using TinyOS and the result show the reliability of our protocol.**

***Index Terms*—WSNs; time synchronization; clock drift; random time source; RTSP**

I. INTRODUCTION

WSNs consist of large populations of wirelessly connected nodes, capable of computation, communication, and sensing. As one of the key technologies in WSNs, time synchronization plays an important role in node localization, low power listening, data fusion, TDMA, synchronized hopping system etc.

Although each sensor node is equipped with a hardware clock, these hardware clocks can usually not be used directly, as they suffer from severe drift. No matter how well these hardware clocks will be calibrated at deployment, the clocks will ultimately exhibit a large skew. In order to get an accurate common time, nodes need to exchange messages from time to time, constantly adjusting their clock values.

Recently, two main network time synchronization methods GPS and NTP are widely used. The GPS method is by receiving time from a Global Position System (GPS) [1], which can provide a high precision; however this information is not available in some situations especially in the battlefield. In addition, GPS receivers are expensive. The other method is Network Time Protocol (NTP), which is operated for internet, and by which we can get high precision for the network time synchronization, following with intensive computing [2]. In WSNs, with battery-powered nodes, the limits for energy supplying, bound of the size and the cost, even the Harsh Environment for the node distributed, GPS and NTP are not suitable for WSNs.

Several time synchronization protocols have been developed to deal with the special requirements of WSNs application. Some of the notable ones are Reference Broadcast Synchronization (RBS) algorithm [3], Timing-sync Protocol for Sensor Networks (TPSN) [4] and Flooding Time Synchronization Protocol (FTSP) [5], Simple Time Synchronization [6], Tsync [7] and Lightweight Time Synchronization (LTS) [8]. There are protocols implementations for these protocols that can achieve synchronization of a few microseconds. However, all these experiments are implemented in laboratory environment with less radio interference. Most of the protocols employ node to node time synchronization and once the parent node failed, all the child of this node will out of synchronization and a process of refresh the topology or re-synchronization may occur. As in figure 1, node 0 is the root of the whole network and it act as the reference time source of other nodes. If node 1 is invalidation all the nodes synchronized through node 1 such as node 3, 4, 5, 6 will out of synchronization soon.

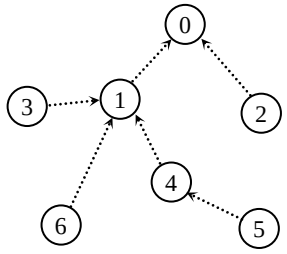


Figure 1. one example of time synchronization

In this paper we propose an error prediction compensation mechanism to provide better average synchronization precision and random time source choice mechanism to decrease the probability of synchronization failure. The remainder of this paper is organized as follows. In Section 2 Time Mode of Time-synchronization shows the model of packet delay and oscillator frequency. Time synchronization algorithm and error prediction compensation mechanism will be described in Section 3. In Section 4, the random choice of

This work was supported by Chinese National 863 High Technology Plan under grant number 2007AA041201, National Natural Science Foundation of China 60804067 and National Excellent Young Leader Foundation under grant number 60725312

time source is discussed. In Section 5, the experiment results are given and the conclusion is presented in Section 6.

I. The Influence on Time-Synchronization

Clocks in nodes, in general, based on crystal oscillators which provide a local time for each network node. The time in a node clock is just a counter that gets incremented with crystal oscillators and is referred to as software clock. The software clock must be increased by the interrupt handler every time an interrupt occurs. Most hardware oscillators are not so precise because the frequency which makes time increase is never exactly right. Even a frequency deviation of only 0.001% would bring a clock error of about one second per day [9]. Considering the physical clock synchronization in a distributed system to UTC (Universal Time Controller), the node clock shows time *C(t)*, which may or may not be the same as *t*, at any point of real time t. For a perfect clock, the derivative *dC(t)/dt* should be equal to 1. This term is referred to as clock rate. The clock rate can actually vary over time due to environmental conditions, such as humidity and temperature, but we assume that it stays bounded and close to 1, so that:

$$1-\rho \le \frac{dC(t)}{dt} \le 1+\rho\,. \qquad (1)$$

The clock rate *dC(t)/dt* denote as *f(t)*, so the clock value in a node *i* at time *t* can be defined as [10]

$$T_i(t) = \int_{t_0}^{t} f_i(\tau)d\tau + \Psi_i(t_0)\,. \qquad (2)$$

Where $\psi_i(t_0)$ is the hardware clock offset of node *i* at time t_0.

From equation (1) the *f(t)* in equation (2) can be denote as $1 - \rho \le f(t) \le 1 + \rho$. Where $0 \le \rho < 1$, which means the hardware clock never stops and always makes progress with at least a rate of $1 - \rho$. This is a reasonable assumption since common sensor nodes are equipped with external crystal oscillators which are used as clock source for the clock of the nodes. Dealing with equation (2) a Taylor series expansion of *T(t)* yields [11]

$$T_i(t) = \beta_i + \lambda_i t + \gamma_i t^2 \cdots . \qquad (3)$$

Here the subscript β is the offset, and α is the skew, and γ can be used to model and detect time variation, i.e., departure from the linear model.

Some protocol such as TPSN use constant model (α_i=0 and γ_i=0) that concern noting more than time offset between two clocks. Several protocols such as FTSP employ linear model (γ_i=0) and estimation of α_i and β_i is readily accomplished via linear regression (least squares). This is optional if the error is Gaussian and can provide the best linear estimator for any error pdf commonly. When the skew and offset are varying with time, a quadratic model can be employed. If αi and βi change in the observation interval, a linear fit will get biased estimates. In this case, a quadratic model can be used to detect clock drift [5].

Without concern the drift between two clocks, constant model should exchange messages from time to time, constantly adjusting their clock values. A quadratic model will lead to higher computational complexity and this is not suitable for resource constrained wireless sensor networks. Current state-of–the–art linear model is widely used for time synchronization in WSNs, and the design such as FTSP can optimize the clock skew to get good performance in WSNs.

II. Synchronization Algorithm

In this section, we describe our clock synchronization algorithm. The basic idea of the algorithm is to use linear regression to achieve time drift between nodes and an error prediction approach to obtain one-step look-ahead prediction is taken charge for compensation mechanism.

Given a time window of n observations, (T_{Ai}, T_{Bi}), we can predict the time at node B give a new time at node A, T_A using ordinary least squares (OLS) to get a linear regression estimation with equation (3) as:

$$\hat{T}_B = \hat{\beta} + \hat{\lambda} T_A\,. \qquad (4)$$

Follow standard regression theory [12], a (1-α) confidence interval can be constructed for this prediction as:

$$\hat{T}_B \pm \left[t_{(1-\alpha)/2,n-2} * SE(\hat{T}_B) \right]. \qquad (5)$$

The first term in the product in equation (5) stands for an upper quantile of the t distribution with n-2 degrees of freedom, and the last is the standard error of the predicted value [13]. The estimate of the prediction error (E_p) we represent as δ. Using 95% confidence interval is the typical choice made in literature due the Gaussian assumption on the error distribution.

For clock used in networks, we will pay the whole attention to the present and future clock time. Arithmetic of linear regression takes charge of the time segment of collecting data for calculation and the parameters acquired gives a good approach for the past time but additional error may occur at present time or the future so the prediction of time error mentioned before is necessarily and a compensation algorithm implemented to compensate the prediction error.

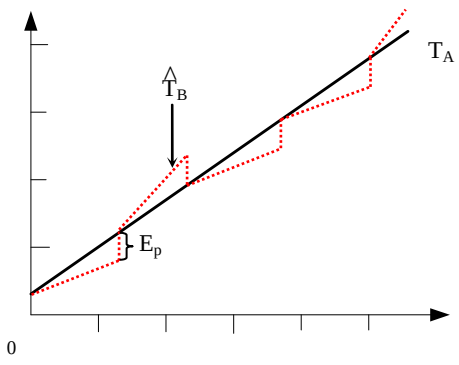


Figure 2. The compensation algorithm

The compensation algorithm will work once E_p is derived while new time information arrived. As show in figure 2, the black line acts as the time of node A, and the red dashed line denotes as the logical time of node B relative to T_A. When E_p is compensated after the calculation, the static error of the logical time in node B can be eliminated and the same process with the next period and so on.

III. Random Time Source

The protocols mentioned earlier always make use of node to node time synchronization. The time precision and the survival period of synchronization depends on the time source the synchronized node selected and if any problem occurs for the time source, all the children of this node will be out of synchronization. If wireless devices deploy in places accompany with other wireless equipments or in industrial plant full of electronic interference, communication fails will be familiar.

In this protocol, node in the network employs several potential nodes as time source and randomly selects one for time synchronization every time period. Some rules deploy as follow:

- The better of the link quality implies much more probability to be chosen as time source.
- If the node success received time information from its recent time source, the parameters according to link quality will be updated, otherwise decrease the parameters. Once the parameters of the potential time source lower the limit of link quality, the corresponding node will be deleted from the potential time source queue.
- If the node fails to synchronize to all the potential time sources, a resynchronization process will be held.

One example of the choice of time source show in figure 3 and the choice of time source depends on the hops from the root. The node 1 and 2 can hear from the network root 0 and they can only synchronize to node 0. The node within the radio radius of node 1 and node 2 can make these two nodes as potential time sources like node 4 and node 7 in figure 3. In the same way, node 5 can acquire 3 potential time source as 4, 6 and 7 in figure 3.

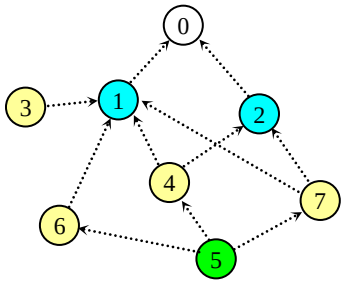


Figure 3. Example of random time source algorithm

Random selecting can reduce the probability of re-synchronization due to link failure and get robust of synchronization. As depict in figure 3 if one of the parents of node 5 fails, it can remain synchronized with the other two nodes.

IV. Experiment Results

This section describes the implementation of our Reliable Time Synchronization Protocol on the SIA2420 sensor nodes using the TinyOS operating system.

Target Platform

The hardware platform used for the implementation of the protocol is the SIA2420 sensor node from Shenyang Institute of Automation in China. It features a TI MSP430 microcontroller with 10kB RAM. The CC2420 radio module has been designed for low-power applications and offers data rates up to 250 kBaud using Direct Sequence Spread Spectrum (DSSS).

The MSP430 microcontroller has 10 build-in 16 bit timers in which 3 timers denoted as timer A and the other 7 as timer B. The SIA2420 board is equipped with two different quartz oscillators (32 kHz and 8 MHz) which can be used as clock sources for the timers. Timer A is configured to operate at 1/8 of oscillator frequency (8 MHz) leading to a clock frequency of 1 MHz. Since Timer A is sourced by an external oscillator it is also operational when the microcontroller is in low-power mode. We employ Timer 2 in Timer A to provide our system with a free-running 32-bit hardware clock which offers a precision of a microsecond.

TinyOS Implementation

The implementation of RTSP on SIA2420 platform is done in TinyOS 2.1. The protocol implementation provides time synchronization as service for an application running on the mote. The architecture of the time synchronization component and its relation to other system components is shown in figure 4

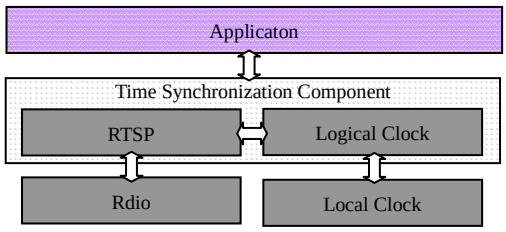


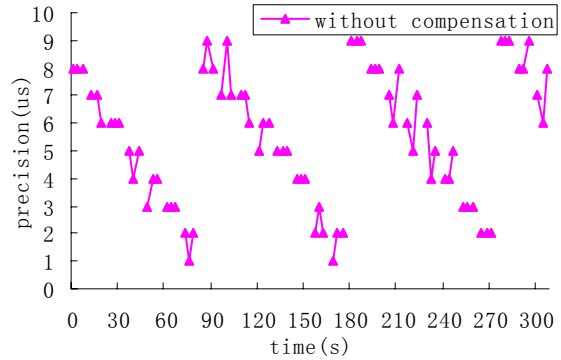
Figure 4. Architecture of time synchronization service and its integration within the hardware and software platform

The RTSP module periodically broadcasts a synchronization beacon containing the sending timestamp. Each node overhearing messages sort by the choice of time source managed by RTSP module and run compensation mechanism is disposed by Logical Clock module.

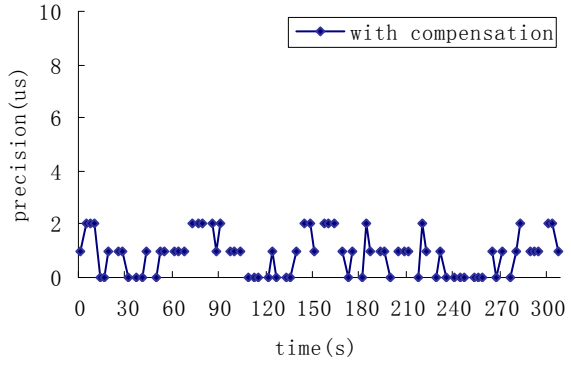
Experiment Results

We tested the protocol focusing on the precision after compensation and the reliability with our random time source select mechanism. The precision test is carried out to make compare with the widely used FTSP. The instant error of two protocols is show in figure 5. Part (a) shows the precision using FTSP and part (b) shows the RTSP with compensation. Analyzing the two curves our protocol can acquire more stable precision under our compensation mechanism.

A scene show as figure 6 is adopted to test the reliability of our random time source mechanism. Node 0 acts as the root and the reference time of the network, and the one hop with nodes (1, 2 and 3) act as the first level of network maintain the unique time source of node 0. Node 5 deploys within the radio range of the three one hop nodes but beyond the scope of node 0 and the same as node 4 and node 6. Node 5 can randomly synchronized to it three potential time sources of one hop and the other two nodes (4 and 6) run unique time source selecting mechanism. The result shows in figure 7 and the node that runs a random selection with 3 time sources achieves better performance. Although nodes with single source can not always choose time source with the most stable as figure 7 in which the node with single source selects its time source of the curve with diamond shape and following with our protocol the node with random choice can obtain a compromise performance. The point at the peak means the node out of synchronization, which we adopt for the convenient of exhibiting the relation of our experiment since the error has great value. The node with single time source may fail of synchronization for the failure of the link with its source node as the peak point in figure 7 and the node performs random choice can always under good synchronization for not relying one single time source.



(a) time precision of FTSP



(b) time precision of RTSP

Figure 5. The compare of FTSP and RTSP with compensation

V. CONCLUSION AND FUTURE IMPROVEMENT

We have described the Reliable Time Synchronization Protocol for WSN. The protocol was implemented on SIA2420 platforms running TinyOS. The average precision is no more than 5μs and robustness to link failure. This performance is markedly better than those of other existing time synchronization approaches on the same platform.

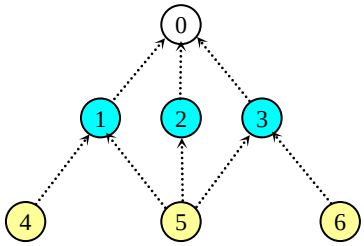


Figure 6. Test scene of the experiment

The RTSP was tested and its performance was verified in a real world application. This is important because the service had to operate not in isolation, but as part of a complex application where resource constraints as well as intended and unintended interactions between components can and usually do cause undesirable effects. Moreover, the system operated in the field with factory device and other interference. This is a testimony to the robustness of the protocol and its implementation.

Several further researches can be done in the future. The first is the influence of temperature to the accuracy of our protocol should be concerned and eliminated. Also synchronization information can be carried by some normal message in the network to reduce the energy cost by synchronous messages.

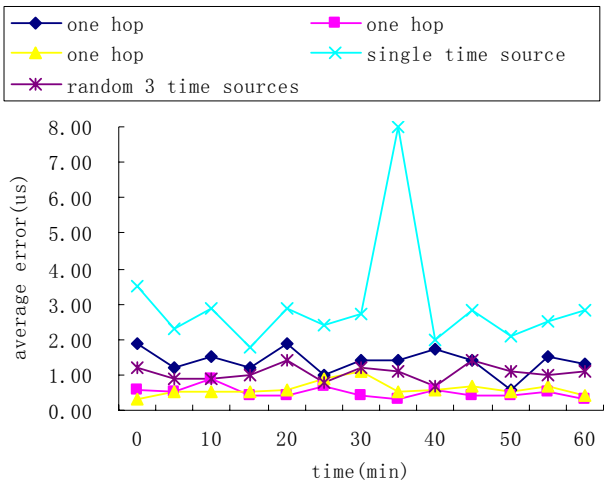


Figure 7. The compare of single time source and random time sources

ACKNOWLEDGMENT

This work is supported by Chinese National 863 High Technology Plan under grant number 2007AA041201, National Natural Science Foundation of China 60804067 and National Excellent Young Leader Foundation under grant number 60725312.

REFERENCES

[1] W. Zhu and Ieee, *TDMA Frame Synchronization of Mobile Stations Using a Radio Clock Signal for Short Range Communications*, 1994.

[2] K. Guanlin, W. Fubao, and D. Weijun, "Survey on Time Synchronization for Wireless Sensor Networks," *Computer Measurement & Control*, vol. 13, pp. 1021-1023,1030, 2005.

[3] J. Elson, L. Girod, D. Estrin, and U. Usenix, "Fine-grained network time synchronization using reference broadcasts," Boston, Ma, 2002, pp. 147-163.

[4] P. Ganeriwal, P. Kumar, and M. B. Srivastava, "Timing-sync protocol for sensor networks " *Conference On Embedded Networked Sensor Systems*, pp. 138 - 149 2003.

[5] M. Maroti, B. Kusy, G. Simon, and A. Ledeczi, "The flooding time synchronization protocol," Baltimore, MD, United states, 2004, pp. 39-49.

[6] M. L. Sichitiu, C. Veerarittiphan, and I. Ieee, "Simple accurate time synchronization for wireless sensor networks," New Orleans, La, 2003, pp. 1266-1273.

[7] H. Dai and R. Han, "TSync: a lightweight bidirectional time synchronization service for wireless sensor networks " *ACM SIGMOBILE Mobile Computing and Communications Review* vol. 8, pp. 125 - 139 2004.

[8] J. v. Greunen and J. Rabaey, "Lightweight Time Synchronization for Sensor Networks," *Proceedings of the 2nd ACM international conference on Wireless sensor networks and applications* pp. 11-19, 2003.

[9] I. K. Rhee, J. Lee, J. Kim, E. Serpedin, and Y. C. Wu, "Clock Synchronization in Wireless Sensor Networks: An Overview," *Sensors*, vol. 9, pp. 56-85, Jan 2009.

[10] P. Sommer, R. Wattenhofer, and Ieee, *Gradient Clock Synchronization in Wireless Sensor Networks*, 2009.

[11] B. M. Sadler, A. Swami, and Ieee, "Synchronization in sensor networks: an overview," Washington, DC, 2006, pp. 1983-1988.

[12] C. R. Rao, *Linear Statistical Inference and Its Applications*. New York: Wiley, 1973.

[13] S. Ganeriwal, I. Tsigkogiannis, H. Shim, V. Tsiatsis, M. B. Srivastava, and D. Ganesan, "Estimating Clock Uncertainty for Efficient Duty-Cycling in Sensor Networks," *Ieee-Acm Transactions on Networking*, vol. 17, pp. 843-856, Jun 2009.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15, August 2010, pp. 014-016

Realization of Information Security in Electronic Commerce

Li Fu-Guo[1] , Dong Yu-Jie[2]

[1]WanFang College of Science and Technology of HeNan Polytechnic University, Jiaozuo,China
E-mail: lfg@hpu.edu.cn
[2]WanFang College of Science and Technology of Henan Polytechnic University, Jiaozuo,China
E-mail: hpudyj@hpu.edu.cn

***Abstract*—With the wide application of E-commerce, E-commerce security issues become more prominent and urgent. This article discussed information security issues in electronic commerce activities, some solutions are proposed to realize E-commerce security in operation and E-commerce environment.**

Index Terms—**Electronic commerce,Information security, Realization, Network**

I. INTRODUCTION

E-commerce is the use of advanced electronic technology to general business activities. E-commerce includes two aspects: First, business activities; second electronic means. Modern electronic commerce is a new business activities based on Internet technology, is the main mode of business operation of 21st century market economy. With the globalization and opening of the internet, without restriction of time and space bring all sorts of e-commerce transactions insecurity. Network information security issues has become an important factor affecting the development of electronic commerce. Therefore, research in an open network environment e-commerce security becomes a very urgent and important field.

II. MEASURES AGAINST E-COMMERCE SECURITY ISSUES TO BE TAKEN

E-commerce security issues relate to various aspects of e-commerce and participate in all aspects of e-commerce transactions, it is a systems engineering and social issues to solve the e-commerce security issues, need participation of whole society. But at the operational level, the following measure should be adopted.

First of all, we should build e-commerce security technology framework systems. In the e-commerce transactions, e-commerce security is mainly network security and transactions security. The network security is the network operating system against network attacks, viruses, so that keep a continuous and stable network operation, commonly used firewall technology protection measures. Transaction security is the data protection of the parties are dealing not be destroyed, as both non-disclosure and transaction identity confirmation, you can use encryption, digital certificates and authentication, SSL (Secure Socket Layer)security protocol, SET(Secure Electronic Transaction) and other technologies to protect.

A. Computer virus prevention technology

Computer viruses are actually a kind of function program running in the computer system, it can destroy and infect against computer system. Virus transmission through the system after a successful attack or breach of license, the attackers usually implant in the system procedures such as Trojan horses or logic bombs, facilitate conditions for the subsequent attack to computer system or network[1]. The current anti-virus software is facing the challenge of the Internet. Currently, hundreds of new viruses were produced in the world every day, and more than 90% of viruses are spread via the Internet. In order to effectively protect the enterprise's information resources, required anti-virus software can support all internet protocols and e-mail systems may be used by e-commerce users, so that it can adapt in time and keep up with the rapidly changing pace of the times. Most anti-virus software mostly focus on stand-alone anti-virus, although some manufacturers introduced a network version of the antivirus products, it only be used in the desktop and file server for protection, the scope of protection is still relatively narrow, so anti-virus vendors should try to enhance protection of the gateway or e-mail server as quickly as possible .Only effectively cut off the entrance of the virus, it will be possible to avoid economic losses of e-business users caused by outbreak of virus.

B. Firewall technology

Firewall as a separator, limiter and analyzer, it mainly used for implementing the access control policy between the two networks, it effectively monitored the network and all activities of the network, it provided necessary access control for the internal network without causing the network bottlenecks, control the data access system of the network through security policy to protect critical resources within the network.

C. Intrusion detection system

With the increasing risk factor of network security, IDS (Intrusion Detection System) as a beneficial complement to firewall, it can help the network system quickly find the coming possible attack, IDS expanded the security management capacities of system administrator (including Security audits, Monitoring, Attack recognition and responsion) and improved the integrity of the information security infrastructure.

Intrusion detection system is a dedicated system which give a real-time monitoring to network activities and behind a firewall, it can work with firewalls and routers and used for checking all communications, records and prohibited network activities of a LAN(Local Area Network) segment, it can be re-configured to prohibit the malicious data traffic come from the outside of the firewall[2]. IDS can quickly analyzed the information on the network or did user audit analysis in the host, manage and monitor network access through the centralized console, so that realized linkage between IDS and network switching equipment. The information of various data streams reported to the safety equipment, IDS can detect network access based on reported information and data streaming content, carry out targeted actions quickly when network security events were discovered, and send these actions response to security incident to firewall or switch, the switch or firewall closed and disconnected accurate port, IDS tried to cut off the connection initiatively and respond immediately when network attacks were discovered.

D. Information encryption method

The purpose of information encryption is to protect the data, files, password and control information within the network and also to protect data transmitted online. Network encryption methods commonly used link encryption, endpoint encryption and node encryption[3]. Link encryption is designed to protect the link information security between network nodes, endpoint encryption's objective is to protect data transmitted from source user to the destination user, the purpose of node encryption is to protect the transmission link between the source node and the destination node. Users can select encryption methods appropriately according to network conditions.

E. Digital certificates and authentication

Digital certificates and certification is a series of data in network communication which marks identity of the communication parties and also a rigorous identity authentication system established through the use of symmetric and asymmetric cryptography. Digital certificates and authentication have following functions :the data not to be stolen by other people except the sender and the receiver and not to be altered during transmission, the sender can confirm the identity of the receiver through digital certificate , the sender can not denied for the information which to be sent.

F. SSL security protocol

SSL is a secure communications protocol. SSL provides a secure connection between two computers, the entire session is encrypted, thereby ensuring the security of transmission. SSL has three characteristics: using symmetric cryptography to encrypt data; using authentication algorithm to proceed the integrity test; using asymmetric cryptography for authentication of the end entity identification.

G. SET technical standards

SET (Secure Electronic Transaction) is a technical standards which pay for the security funds through open network, SET provides real security rule to the applications of electronic transaction based on credit card: ensure secure transmission of the internet and transmission data is not stolen by hackers; orders and personal account isolated, when the order contains the cardholder account sent to businesses, the business can only see the orders but not the cardholder's account; cardholders and merchants authenticated mutually, so that to determine the identity of both communication sides, usually a third party responsible for providing credit guarantee to both sides of the online communications; requires the software follow the same protocol and message format so that software developed by different manufacturers have compatibility and interoperability function and may be run on different hardware and operating system platforms.

III. ENVIRONMENT MEASURES TO KEEP E-COMMERCE INFORMATION SECURITY

A. Construct a sound e-commerce system

Actively participate in international cooperation and integrate international e-commerce framework, construct e-commerce system suitable for China's national conditions. As a sovereign state, in order to safeguard national interests and economic security, we must pay attention to proprietary technology development which related to e-commerce technology, not all rely on imports. Therefore, we must increase investment, focus on the research and development of e-commerce security technology.

B. Strengthen the laws and regulations

To against the new types of crime related to information technology and information systems, government departments should organized forces quickly and combined with the objective of e-commerce needs so that to strengthen the existing laws and regulations related to electronic commerce, it is a usual practice of human history to fight against crime with the use of law. Chinese government can strengthen the laws such as: "The People's Republic of China Criminal Law", "the National People's Congress Standing Committee decision on Internet security", "Contract Law", "Copyright Law" and other related laws. In these laws, it can properly increase the penalties provisions for cyber crime and increased the terms of copyright protection of network works. Relevant authorities of the government should develop departmental rules and regulations firstly so that to against related issues need to be solved quickly with the development of e-commerce such as electronic payments, tax Administration, security certification, network and information security, intellectual property rights protection, consumer protection and so on, when necessary, administrative rules and regulations can be issued by the State Department, and then rose to the legal procedures.

C. Speed network infrastructure construction

Information infrastructure is the material basis and the carrier for development of electronic commerce. Speed up the network infrastructure construction, promote the process of enterprise information, it is the direct driving force to enhance the research of science technology and application in the related applied fields with which we can obtained the "innovation" and "sustainable development" in the information security field and information field. The development of information infrastructure needs support of variety of disciplines and talents, the joint efforts of government and industry, in particular the Government's strong investment and macro-control.

Ⅳ. CONCLUSION

Chinese Government should strengthen the research of e-commerce information security, so that to establish a flexible legal framework to regulate e-commerce, so that to make e-commerce open, reasonable and legalization. This will ensure not only the interests of all e-commerce sides but also the smooth progress of e-commerce. Enhance the security of e-commerce, in addition to using advanced science and technology arm, but also its own e-commerce companies to take proactive security measures. Enterprises must carry out its internal security awareness education for all staff so that they can fully understand the importance of enterprise information security, and take appropriate preventive measures to against insecurity factors, this is the only way to ensure the safe operation of e-commerce information.

REFERENCES

[1] JIN Bo, SONG Ping, “A Probe into the Information Security of Electronic Documents in Electronic Government,” Journal of Shanghai University (Social Science Edition), vol. 16, pp. 128–129, March 2009.

[2] YUAN Jia-bin, GU Kai-kai, YAO Li, “Security Grid Technology Based on Information Security Control Theory,” Journal of Nanjing University of Scinece and Technology (Natural Science Edition), vol. 31, pp. 423–425, April 2007.

[3] WANG Da-kang, DU Hai-shan, “Encrypt & Crack Technology in Information Security,” Journal of Beijing University of Technology, vol. 32, pp. 498, June 2006.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 017-020

Improved Routing Algorithm Research for ZigBee Network

Zhao Hong-tu [1], Ma Yue-qi [2]
[1] College of Computer Science & Technology, Henan Polytechnic University, JiaoZuo, China
Email: HT-ZHAO@163.com
[2] College of Computer Science & Technology, Henan Polytechnic University, JiaoZuo, China
Email: myq3636@163.com

***Abstract*—Aiming at the problems of ZigBee AODVjr routing algorithm that the RREQ packets flooding and the routing may not be optimal and some node may use up all the energy because of heavy transmissions in the Cluster-Tree, an improved routing algorithm for ZigBee networks is proposed. AODVjr algorithm is combined to control the range and the direction of the RREQ packets in this improved algorithm. At the same time, the neighbour table was introduced to make sure that the routing was also considered to avoid selecting some node with low residual energy in the Cluster-Tree algorithm. The simulation results indicate that the energy consumption is reduced efficiently, the problem of unbalance load is resolved and the lifetime of the whole network is maximized in this improved algorithm.**

***Index Terms*—ZigBee network, AODVjr algorithm, Cluster-Tree algorithm, residual energy**

I. INTRODUCTION

ZigBee is a newly developing short-range, low-rate, low cost, low-power wireless network technology. The technology was designed targeting at low-rate wireless sensor and control network, which can be widely used in industry, families, medicine and other low-power, low cost wireless communication applications which ask for less demanding on data rate and quality of service. With the development and improvement of the ZigBee technology, its extensive use will necessarily bring great convenience to people's daily lives [1].

ZigBee networks adapt Cluster-Tree algorithm and AODVjr algorithm. As a simplified version of AODV, AODVjr algorithm can support end-to-end transmission. In the route discovery process a large number of flooded RREQ packet will result in significant additional energy consumption, so that the overall consumption of the network can be excessive. To minimize RREQ packet costs as possible is one of the ways to reduce the overall consumption of the network. This paper, through controlling the scope and direction of the RREQ packet transmission with AODVjr algorithm as well as introducing a neighbour table into the Cluster-Tree algorithm to make routing choices to avoid the remaining low-energy nodes as far as possible, thus reducing network overhead, saving the network's overall energy consumption to extend the network life.

II. AODVJR ALGORITHM AND CLUSTER-TREE ALGORITHM DESCRIPTION

A. *AODVjr algorithm*

AODVjr is a simplified version of AODV. AODVjr can execute the main function of AODV, but take into account the lower cost, energy-saving, ease of use and other factors to simplify some of the features of AODV. First of all, in order to reduce controlling overhead and simplify the process of route discovery, AODVjr did not use the serial number of the destination node, and in order to ensure loop-free routing, AODVjr stipulates that only the destination node can reply packet RREP, and intermediate nodes cannot reply RREP even if the route to access destination nodes exist through the nodes. At the same time, not as in AODV, there is not a pioneer table in AODVjr, which simplifies the routing table's structure. Additionally, AODVjr node does not send a HELLO packet information, and update neighbour nodes list based solely on the received pocket information or the information provided by the Mac layer, thus saving some controlling overhead [2].

B. *Cluster-Tree Algorithm*

In the Cluster-Tree algorithm, the nodes according to the network address of the destination node group to calculate the next hop of the pocket. For ZigBee routing nodes whose address of A, depth of d, if met the following inequality, then the the destination node whose

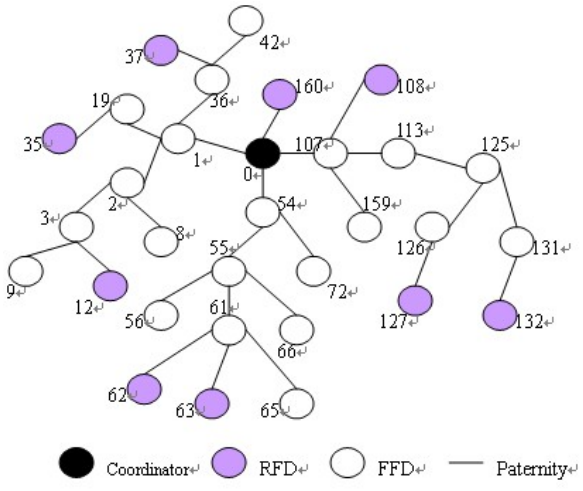


Figure1. Network Address Assignment

address of D is one of its offspring:

$$A < D < A + Cskip(d-1) \quad (1)$$

If it is determined the packet destination node is the descendant of the accept node, the node will send packets to a sub-node. This time, if satisfied:

$$D > A + Rm \times Cskip(d) \quad (2)$$

It means that the destination node is its terminal child node, then the next node address N as follows:

$$N = D \quad (3)$$

Otherwise, if the destination node is not an offspring of receiving node, , it will send the packet to its parent node.

C. *Shortcomings of AODVjr algorithm and Cluster-Tree Algorithm for the*

Although AODVjr comparing Cluster-Tree algorithm can find the optimal path, but in the route discovery process the AODVjr algorithm will still generate redundant RREQ packet, while the redundant RREQ packet also involve in route discovery process, but does not play much in finding an optimal path eventually. Shown in Figure 1, the node 107 will send data to node 61,and if from the node 107 to node 61, there is no routing table entry, then the node will initiate the routing discovery process that the node 107 send RREQ packet to all its neighbour nodes .Because node 61 is not a descendant node of the node 107, the node 107 send RREQ packet to its descendant nodes, so it plays a minor role in the process in finding the optimal path from node 107 to node 61 .

In addition, the closer the nodes near the root the faster the energy consumes. If frequently all of the descendants of node 1 send data through node 1 to the descendants of the node 107 or node 54, it may lead the battery power of node 1, node 54 or node 107 runs out too quickly, so that not only led node 1, node 54 or node 107 cannot properly communicate with other nodes, but also led the entire network partitioning, resulting the descendants of the node 1 communicate with the descendants of node 54 or node 107. Therefore, if the remaining energy of node 1 is in a low-energy situation, the descendants of the node can choose the other nodes forward data to avoid an early exhaustion of the energy value of the node1. To deal with these problems, considering the appropriate restrictions on RREQ packet flooding in the process of route discovery and nodes' residual energy, we propose the following improved algorithm.

III. IMPROVED ALGORITHM DESIGN

A. *Neighbour Table Definition*

If the two nodes can communicate directly within one hop, we say that the two nodes are neighbours. Due to the RFD device storage capacity is weak, so it only stores its neighbour list for the FFD equipment, and RFD equipment, particularly need to store the neighbour list of it. FFD nodes in the network record the adjacency between the nodes and other nodes through the neighbour list Nlist, as shown in Table I.

TABLE I. NEIGHBOR LIST ENTRIES

ADDR	DT	NP

In the Nlist, there are three fields: ADDR: neighbour nodes address DT: neighbour nodes device identity bits, 1 indicates that the neighbour node is FFD device, with a routing function; 0 indicates that the neighbour node is RFD device, does not have routing functions, and only send and receive data. NP: neighbour node residual energy identity bits. 1 indicates saturated node, 0 indicates non-saturated node.

B. *Cluster-Tree Algorithm*

At first we divided the nodes into two distinct areas based on its remaining battery power:

1) saturated nodes: If the value of the current residual energy of the node is greater than Ec, it locates in a saturated node. FFD points in the region can participate in data forwarding.

2) Non-saturated nodes: If the current residual energy of the node is less than Ec, it locates in the non-saturated nodes. The transmission of data should be away from the node as far as possible.

Suppose the node's initial energy is Ew, and Ec is set as follows:

$$Ec = \frac{a}{f(x)}\sqrt{Ew} \quad (4)$$

Where is a specific factor, whose role is to slow down the speed of Ec value decreasing (in simulation test, we value a = 2). $f(x)$ is a function changing as x is changed, which is defined as follows:

$$f(x) = \begin{cases} 1, & x = 0 \\ \dfrac{Nt + x}{Nt + x}, & 1 \le x \le Nt \end{cases} \quad (5)$$

Where Nt refers to the total number of the nodes of the network, as a constant. X as the variable, defined below in detail. From the formula 4 can be seen that $f(x)$ is increasing function on the x, therefore, Ec can be deduced as a decreasing function on the x, namely, Ec increases as x decreases. When Ec reduces to a certain extent, some nodes in the non-saturated zone may become the node of saturated zone and continue to be used. When x approaches the Nt, value of Ec tends to zero. At this time the node still lower than the Ec could be regard as dead node, and the decreasing rate of Ec will be increasingly smaller[3].

We set two internal counters C1, C2 at the central coordinator. If the current spare capacity of the node is lower than current Ec value, then the node will send early warning information to the central coordinator. When the central coordinator receives a warning message for each, and then will plus one on the counter C1. Through the counter C1, the central coordinator allows you to count the proportion which the non-saturated nodes across the network may account for as P, we set a threshold T (0 <T <1, in the simulation experiment, we take T = 0.2). When P <T, the counter C2 will plus 1, and X in the formula 3

and 4 are the value of the counter C2, so that the value of x plus I to change, and so as to update the value of Ec. When updated Ec value, the counter C1 will be cleared, and recount the number of non-saturated nodes.

When sending data packets, node can according to the different regions of the neighbour node to determine the routing forwarding mechanism. The nodes in the saturated zone whose residual energy is sufficient, when choose the routing only need to take into account number of hops; the nodes in the non-saturated zone whose residual energy is short , when select the routing should avoid such a node as much as possible [4].

C. *Improved Routing Algorithm*

ZigBee network routing algorithm can according to the expression (1) determine whether the destination node is the descendants of the node of an intermediate forwarding node. So, if the destination node is the forwarding node's descendants, at this time if still allow the forwarding node's parent node to forward RREQ, the possibility to find the optimal path through the parent node is very small. If the destination node is not a descendant node of the forwarding node, and we still allow the descendants of the forwarding node to forward RREQ packet, it also does not mean much to find the optimal path. Therefore, through the improved algorithm with the basic idea of the tree routing determine the general direction of a RREQ packet, thus avoiding RREQ packet flooding along the opposite direction to the destination node to save the overall network energy consumption.

Based on AODVjr algorithm the improved algorithm adds the RREQ packet flag: flag = 0 indicated that the current node's parent node should not forward the RREQ packet; flag = 1 indicated that the current node's descendant nodes should not forward the RREQ packet. Routing method as follows:

(1) If the RFD node wants send data to other nodes in the network, the data will be forwarded by the RFD node directly to its parent node, and then forwarded by the parent node.

(2) When the FFD nodes with routing function send data to other nodes in the network, if the source node's routing table has not the routing table entry to the destination node, will start the route discovery process.

(3) In the route discovery stage, when the node as an intermediate forwarding node, sends the RREQ packet from the source node, the node will detect its residual energy value at first.

(4) If the node's residual energy value <Ec, then it will send a warning message to the central coordinator. At the same time, the node will detect the value of the flag in the RREQ packet: If flag = 0, shows that the parent node of the node is not suitable for forwarding the RREQ packet, and then should choose the descendants of the node directly whose NP value is 1 by looking at the neighbour's menu to forward PREQ, and if the NP values of all the descendants of the node are 0, then will abandon the RREQ.

(5) If flag = 1, shows that the descendants of the node is not suitable to forward the RREQ packet ,then should directly view the parent node in the neighbour table to determine if the NP value is 1, if for 1,then forward PREQ ; for 0,then discard RREQ.

(6) When the destination node receives the RREQ packet, irrespective of the number of residual energy, always send back a RREP packet.

(7) When the source node receives the RREQ packet sent by the destination node, and then according to the route discovery path transmit data.

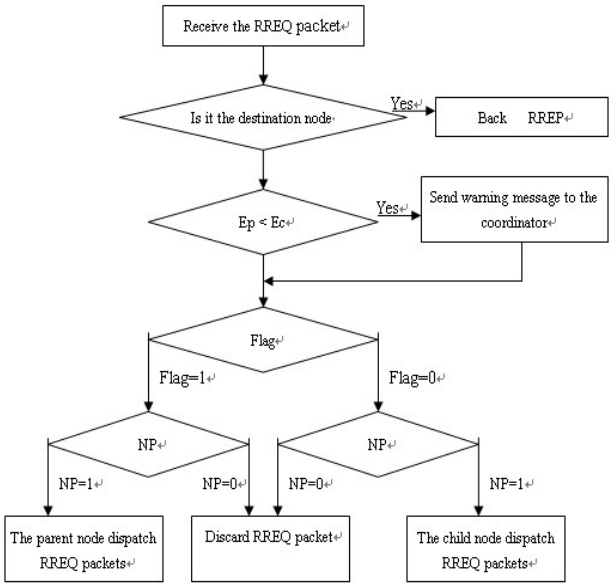


Figure2. The treatment of RREQ by intermediate nodes flow chart

IV. ALGORITHM AND RESULTS ANALYSIS OF STIMULATION EXPERIMENTAL

A. *Algorithm Analysis*

This improved algorithm through a combination of AODVjr algorithm and Cluster-Tree algorithm set pre-control on the direction of the PREQ sending, so as to reduce the flooding of the RREQ packet, and can balance residual energy of PREQ packet each node of the network, extending the life of the nodes of the network. Meanwhile, the improved algorithm is simple, and costs of maintaining neighbour node are also smaller, additionally the algorithm's time complexity is O (n).

B. *Simulation Result Analysis*

Improved algorithm was compared with the traditional AODVjr algorithm by simulation experiments, focusing on comparing the total network energy consumption and the number of failure nodes in transmission. Simulation results have proved the effectiveness of the improved algorithm.

Simulation tool was Omnet + +3.2 p1. Network covers an area of 400 × 550m, with the number of network nodes is set to 80, and the data packs all are length of 256B, the channel bandwidth of 2M, the initial energy of the node all are 5000J. We set Cm = 4, Rm = 3, Lm = 4. The simulation results shown in Figure 3 ~ 4.

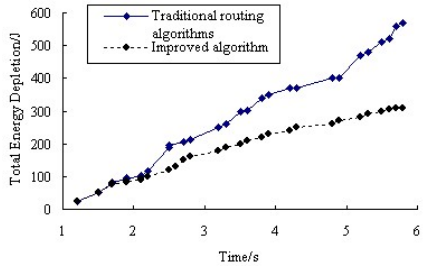


Figure3. Overall network energy consumption curve.

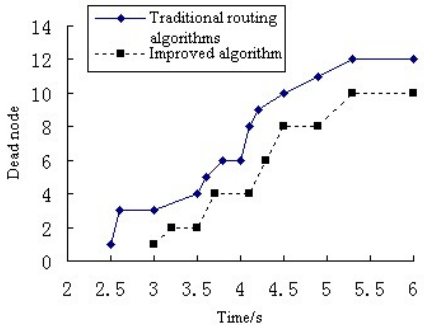


Figure4. The death node number of the network curve

In Figure 3, the curve 1 represents the network's overall energy consumption with the traditional algorithm run-time, and curve 2 is on behalf of the network's overall energy consumption with this improved algorithm. As the improved algorithm has control the RREQ packet flooding, and balanced the residual energy of network nodes so as to have saved energy.

In Figure 4, curve 1 represents the death node number of the network with the traditional algorithm run-time, and curve 2 indicated that the number of dead nodes with this improved algorithm run-time. At the initial stage, the energy of each node is sufficient, there will not produce the dead node. As networks' running time increase, some nodes frequently as a forwarding node consume significant energy, and the traditional method does not take into account the value of the node's residual energy, therefore, the time of emergence of the dead node is earlier than the improved algorithm. For this improved algorithm avoids node of the low residual energy, and select the nodes of more energy for data forwarding, so as to avoid the premature death of individual nodes, balance the network load, maximum of the network's survival.

V. CONCLUSIONS

This paper presents an improved algorithm basing on the traditional AODVjr algorithm combined with Cluster-Tree algorithm; control the PREQ packet in the routing discovery process, and introducing a neighbour table, considering the node's residual energy in the data transfer process, which makes the network energy consumption to achieve parity. The simulation results show that the algorithm have saved the network's overall energy consumption and extend the network life.

REFERENCES

[1] ZigBee Document 053474r06[S].Version 1.0.ZigBee Alliance,2004

[2] Chakeres LD, Klein-Berndt. AODVjr, AODV simplified [J].Mobile Computing and Communication Review,2002.

[3] Yanli. B, Improved routing algorithm for ZigBee Network[J], computer engineering and applications, 2009.

[4] Yanli. B, ZigBee Network tree routing algorithm based on energy balance[J], Computer Applications, 2008

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 021-024

Research and Analysis of Adaptive Failure Detection Algorithm

Lei Shi [1], Shifei Yang [1], and Qian Zhang [1, 2]
[1]School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China
Email: shilei@zzu.edu.cn
[2]Henan Provincial Key Lab on Information Network, Zhengzhou 450052, China
Email: {sfy2008, zhangqian}@yahoo.com.cn

***Abstract*—Due to the variations of the network in actual distributed system, failure detectors without adaptive mechanism cannot meet the requirements of QOS of applications. Adaptive failure detectors should dynamically adjust the detecting quality according to the real-time state of the network. Assuming that delay of the message and loss of the messages is a random probability, the failure detection model based on the predicted message delay is proposed in this paper. A P_{AC}-AFD adaptive failure detection algorithm is realized based on the above model based on the prediction from historical message delay and it contains checking idea. Experimental results show that the algorithm can relieve the effect of delay of the message and loss of the message on the failure detection while ensuring the accuracy and completeness of detection.**

***Index Terms*—failure detection, QOS, distributed system, adaptive, checking.**

I. INTRODUCTION

As an important building block for fault-tolerant systems, failure detector plays a central role in such dependable systems. Therefore, ensuring QOS of failure detector is very important for ensuring fault tolerance of distributed systems. Chandra and Toueg [1] firstly proposed metrics of unreliable failure detectors which can resolve some radical problems of unreliable system.

The state of network is multivariant in actual distributed system. At the same time, there are many kinds of applications in distributed system and different one of them has different requirements of failure detection [2]. And then, Adaptive failure detectors are presented to meet different requirements of QOS [3].

II. RELATED WORK

Fetzer [4] firstly proposed a simple adaptive failure detection mechanism which gained a maximal delay time to be the upper limit of overtime by collecting the reached heartbeat message delay. After proposing the measurement system of QOS, Chen [5] presented some adaptive algorithms based on probability network model to realize quantitative control of adjusting the parameters of failure detectors by OOS. This algorithm presents a good prediction for the next arrival time. Bertier [6] and Hayashibara [7] improved Chen's adaptive failure detection algorithm realizing less detection time.

In this paper, a new adaptive failure detection method is proposed.

III. BASAL THEORY OF FAILURE DETECTOR

A. Failure Detector Model

The model of failure detector can be defined as follows [1]: assume a system with N processes: $\Pi = \{P_1, P_2, \ldots P_n\}$. There is an independent global clock in the system and a time set T obtained from clock time signal which is natural number. Suppose $p \in \prod$, $t \in T$, then failure detector can be defined as: $FD_p(t)$: $\Pi \times T \rightarrow 2^{\Pi}$. For $q \in \prod$, if $q \in FD_p(t)$, the failure detector of p deems that q is failed at t. The output of FD is a set of failed objects: Failed= $\cup_{t \in T} FD_p(t)$.. Failure model of nodes fits Fail-stop [8] model.

B. Failure Detector Level

Failure detector has two basic metrics: completeness and accuracy and it can be classified eight classifications [9] according to completeness and accuracy, as Table1.

TABLE I
Eight Classifications of Failure Detector

	Strong completeness	Weak completeness
Strong accuracy	P	Q
Weak accuracy	S	W
Eventual strong accuracy	◇P	◇Q
Eventual weak accuracy	◇S	◇W

A perfect failure detector should agree with the definition of ◇P (a set of eventually perfect failure detector) that should fits:

Strong Completeness: Every failed process will eventually be judged eternally failed by all the correct applications in any operation.

Eventual Strong Accuracy: Every correct process will not be falsely judged as failed process after some moment t in any operation.

IV. FAILURE DETECTION ALGORITHM BASED MESSAGE DELAY PREDICTION

In this paper, P_A called query accuracy is used to represent the QOS demand of application, here, $P_A \in (0, 1)$. It can be set flexibly according to different applications to meet different requirements of QOS.

When detection accuracy is more important, P_A is set with a large value. When detection speed is more important, P_A is set with a small value.

A. Basic Failure Detector Algorithm

Definition 1: Let sup{x:p(t≤x)=1} and low{x:p(t≥x)=1} which are called sup and low for short respectively be the upper and lower bounds for the random variable t.

Definition 2: If the random variable t has its bound, then we call $A(t)=\frac{\sup(t)+\mathrm{low}(t)}{2}$ the arithmetic mean of t. And if low≥ 0, we call $G(t)=\sqrt{\sup(t)\times\mathrm{low}(t)}$ the geometric mean of t.

Theorem 1: Let t be a random and bounded variable, the mathematical expectation and variance of t are E(t) and V(t), and low>0, then

$$V(t)\le 2E(t)*\left[A(t)-G(t)\right] \quad (1)$$

Theorem 2: Assume that E(D) is the mathematical expectation of D which is the interval between two heartbeat messages delay. Then for arbitrary t>0,

$$P(D>t)\le\frac{2E(D)*\left[A(D)-G(D)\right]}{\left[t-E(D)\right]^2} \quad (t>E(D)) \quad (2)$$

Proof: Because the interval between two heartbeat messages is a bounded random variable, so it can be obtained from theorem 1 that:

$$V(D)\le 2E(D)*\left[A(D)-G(D)\right]$$

Here we assume that V(D) is the variance of D which is the interval between two heartbeat messages delay. The interval between two heartbeat messages delay is a random probability event, so it can be obtained from Chebyshev inequality that:

$$P(D>t)\le\frac{V(D)}{[t-E(D)]^2} \quad (t>E(D)) \quad (3)$$

So, it can be obtained from (1) and (3) that:

$$P(D>t)\le\frac{2E(D)*\left[A(D)-G(D)\right]}{\left[t-E(D)\right]^2} \quad (t>E(D))$$

Statistic the intervals of recent N heartbeat messages from detected nodes calculating E(D)、A(D) and G(D), to propose the possible delay time of the next heartbeat message. Assume that the predicted time of this time is T_1 with a weighted value P_1, the previous predicted time is T_2 whit a weighted value P_2……the predicted time of the previous ith time is T_{i+1} with a weighted value P_{i+1}……the predicted time of the previous w-1th time is T_w with a weighted value P_w. Here w is the window size of historical record. In addition , $\sum_{i=1}^{w}P_i=1$ and $P_i=\frac{K}{i}$ can be obtained from the first Zip law. K is a parameter which can be known from normalization calculation.

$\sum_{i=1}^{w}P_i=\sum_{i=1}^{w}\frac{K}{i}=K\sum_{i=1}^{W}\frac{1}{i}\approx K*(\ln w+r)=1$, r is Euler constant. $K\approx\frac{1}{\ln w+r}$.

The overtime value of timer for the next heartbeat message reaching can be set according to the reaching interval of historical heartbeat messages. If the detector does not receive the heartbeat messages from the detected node in predicted delay interval, it will be checked. The checking process is:

Detection process p sends inquiring messages to the detected process q. The format of inquiring message is ask(q, count), here q represents the process inquired and count represents the sequence number of heartbeat messages sent to inquired processes. If process p receives the response message ack(q, count, yes) from process q in the predicted time, it will be considered normal. If process p does not receive the response message in the predicted time, process q will be considered failed. This can improve the accurate of the failure detector.

The P_{AC}-AFD algorithm is described as follows:

Input: heartbeat messages;

Output: status of the process being detected;

1. for process p and q, initialize UDP socket;
2. initialize others correlative arguments;

Process q:

3. if current time is i* t
4. send heartbeat message to process p; /* t is the period of sending heartbeat message */

Process p:

5. initialize that process q is live;
6. loop
   {
7. timer (t_n) start; /* t_n is the predicted interval of heartbeat delay */
8. wait for receiving message from q;
9. if $t_c\le t_n+t_p$ and $k<s_{min}$ /* received overtime message t_c is the reaching time currently of the heartbeat message, t_p is the reaching time of the previous heartbeat message, s_{min} is the smallest sequence number of the heartbeat message which does not receive its response message up to now */
   {
10. judge that process q is live;
11. timer(t_{n+1}) restart;
    }
12. else if tc≤tn+tp and k≥smin /* received message which is being waiting */
    {
13. j←k%w; /* k is the sequence number of the response message of q */
14. $t_d\leftarrow t_c-t_p$; /* t_d is the detection time of heartbeat message */
15. add(t_d) to sw[j];/* Save the detection time into the sliding window*/
16. $s_{min}\leftarrow k+1$; $t_p\leftarrow t_p+$ t;
17. calculate out E(D)、A(D)、G(D);

18. $T_d = \sqrt{\frac{2E(D)*[A(D)-G(D)]}{1-P_A}} + E(D);$

/* the possible arrived delay time of the next message*/

19. $t_n = T_d P_1 + \sum_{i=2}^{w} T_i P_i;$

(4)

/* the predicted arrival time of the next message*/}

20. else
{
21. send inquiring message ask(q, count);
22. if receive ack(q, count, yes) in t_r /* t_r is the predicted response time of inquired processes*/
{
23. judge that process q is live;
24. timer(tn+1) restart;
}
25. Else
{
26. judge that process q is dead;
27. tp←tp+ t;
28. add tn to sw[smin%w];
29. smin←smin+1;
}}}

B. Failure Detection Level of the Algorithm

The algorithm proposed in this paper meets strong completeness and eventual strong accuracy. We assume that there is an upper limit of message delay.

Property 1(Strong Completeness): **E**very correct process will receive the last heartbeat message from process q at sometime T_i (i＝1, 2,…,n-1). For correct process P_1, assume that it receives the last heartbeat message from process q at time T_1. We can know from the algorithm that process P_1 will calculate an interval t_n of message delay for the next heartbeat message sent by process q after receiving a heartbeat message from it. If process P_1 does not receive the heartbeat message from process q in that interval, then the checking will be carried out. During checking, if P_1 cannot receive the message ack(q,count,yes) in predicted interval, then P_1 will consider q failed. We know that all the data needed to calculate the next heartbeat message delay in this algorithm, including E(D)、A(D) and G(D), has a determinate value. So, the interval of message delay is bounded.

We know from the algorithm that the total detection time is bounded. We use T_{fi} to represent the value of this upper limit and each process has that moment. Let T_{fm}= max $\{T_{fi}\}$ as its maximum value, so all processes consider that process q has failed after T_{fm}.

Property 2(Eventual Strong Accuracy): If the message delay time calculated by the algorithm is smaller than the actual message delay time, then the process will be checked to judge whether it is failed or not. During checking, if detecting process does not receive the response message from detected process in the predicted time, the detected process will be considered failed. Process p will dynamically increase detection time to adapt the changes of networks of the process q.

The detection time will increase with time increasing. Until after a certain time t_0, process p will receive heartbeat message sent by process q in detection time.

V. Experiment and Analysis

The configuration of the environment is as follows: Two computers with the same configuration are connected via the internet. The detected machine sends heartbeat messages to the detector periodically, and the detector will return a response message to the detected machine after receiving the heartbeat message from the detected machine. The interval between two adjacent heartbeat messages is one second. The detection metrics are error rate and average detection time in different window size and different value of P_A.

A. Impact of P_A Values on the Detection

Firstly, we set the window size of historical a fixed value 1000, and P_A is set a value 0.6, 0.7, 0.8 and 0.9 respectively. Fig1 shows a group of heartbeat message sequence selected randomly in 24 hours which has a consecutive sequence number. We can see from the chart: The larger the P_A is, the longer the detection time is.

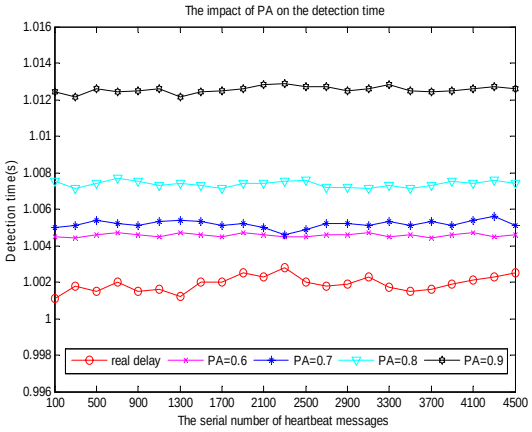


Figure 1. Impact of P_A on the detection time

It can be seen from the table 2: P_A has larger influence on error rate, that is, the larger the P_A is, the longer the average detection time is and the smaller the error rate is. The result meets predicted effect.

TABLE II
Impact of P_A on the average detection time and error rate

P_A	0.6	0.7	0.8	0.9
Average detection time(s)	1.004 257	1.005 719	1.007 347	1.012 531
Error rate	0.011 697	0.005 238	0.001 072	0.000 142

B. *Impact of Historical Window size on the Detection*

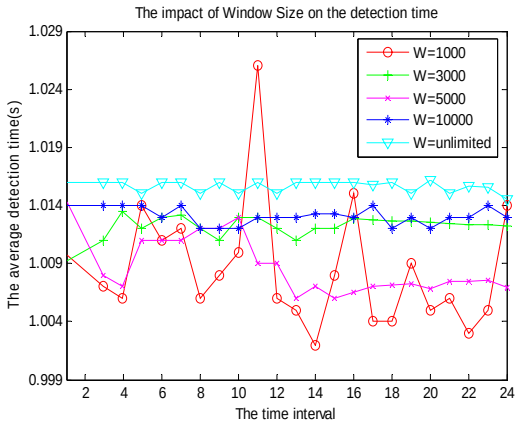


Figure 2. Impact of window size on the detection time

From Fig2 we can see that the average detection time becomes larger with the size of window becoming larger and the average detection time becomes smaller with the size of window becoming smaller. Here, the average detection time is the longest when the size of window is unlimited.

TABLE III
Impact of window size on the average detection time and error rate

Window size	1000	3000	5000	10000	unlimited
Average detection time(s)	1.012 532	1.012 792	1.013 082	1.013 962	1.015 672
Error rage	0.000 143	0.000 117	0.000 085	0.000 051	0.000 026

Fig 2 and Table 3 show that: The size of window has a large impact on the fluctuation of average detection time. The smaller the size of window is, the larger the fluctuation of average detection time is and the higher the error rate is.

C. *Analysis of Performance*

We compare our algorithm the with Chen's failure detector in the same network environment. In our algorithm, the size of window is set a value 1000 and the period of sending heartbeat message is one second. In Chen's algorithm, safe margin α is set a value 0.005. We set P_A a value 0.85. Table 4 shows: When the size of window is 1000, α is 0.005 and P_A is 0.85, P_{AC}-AFD algorithm has smaller error rate than Chen's algorithm ensuring almost same average detection time. The performance of failure detector is improved.

TABLE IV
Comparison with Chen's algorithm

	P_{AC}-AFD	Chen
Parameter	P_A =0.85	α=0.005
Window size	1000	1000
Average detection time(s)	1.011792	1.011835
Error rate	0.001023	0.002442

D. *Conclusions*

A predicted method on the basis of the prediction from historical message delay is studied in this paper. The checking idea is used in this method. The algorithm predicts the next message delay time according to the historical record of message delay. The analysis of experiment and performance proved the failure detection level of the new algorithm. Experimental results show that the algorithm has strong adaptability and it can relieve the effect of message delay and message loss on the failure detection while the detection accuracy and detection completeness is satisfied.

REFERENCES

[1] T. D Chandra and S. Toueg, "Unreliable Failure Detectors for Reliable Distributed Systems," Journal of the ACM. vol. 43(2), pp. 225-267, 1996.

[2] J. Dong, Research on key techniques of failure detection in distributed systems [Ph.D. Thesis]. Harbin institute of Technology. 2007.

[3] N. Xiong, A.V Vasilakos and L Yang, "Comparative analysis of quality of service and memory usage for adaptive failure detectors in healthcare systems," IEEE Journal on Selected Areas in Communications. vol. 27(4), pp. 495-509, 2009.

[4] C. Fetzer, M Raynal and F Tronel, "An adaptive failure detection protocol," IEEE the 8th Pacific Rim International Symposium on Dependable Computer. pp. 146-153, 2001.

[5] W. Chen, S. Toueg and M.K Aguilera, "On the quality of service of failure detectors," IEEE Trans. on Computers. vol. 51(5), pp. 561-580, 2002.

[6] M. Bertier, O. Marin and P Sens, "Implementation and performance evaluation of an adaptable failure detector," Martin DC, ed. Proc. of the 15th Int'l Conf. on Dependable Systems and Networks. Bethesda, IEEE CS Press. pp. 354-363, 2002.

[7] N. Hayashibara, X. Defago and T Katayama, "Implementation and performance analysis of the φ -failure detector," JAIST Research Report. IS-RR-2003-013, 2003.

[8] Yair, Amir, Danny, Dolev, Shlomo, Kramer, Dalia and Malki, "A communication sub-system for high availability," in the proceedings of the 22nd Annual International Symposium on Fault-Tolerant Computing. Boston, pp. 76−84, 2002.

[9] I. Gupta, T. Chandra and G Goldszmidt, "On scalable and efficient distributed failure detectors," in the proceedings of 20th Annual ACM Symposium on Principles of Distributed Computing. pp. 170-179, 2001.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 025-027

Research on Covering's Reduction

Zhi Dongjie[1], Zhi Huilai[1], Liu Zongtian[2]
[1] School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, China
Email: zhidongjie,zhihuilai@126.com
[2] School of Computer Engineering and Science, Shanghai University, Shanghai, China
Email: ztliu@shu.edu.cn

***Abstract*—When using reducible formula to reduction a covering, the outcome exist redundant. In order to overcome redundant, precision formula and reduction algorithm are promoted. Information quality and compression ratio are defined to evaluate covering precision reduction. Although covering precision reduction may contain different sub-sets, their information quality is same. Covering precision reduction is better than the existing covering reduction theory in the perspective of redundancy eliminating.**

***Index Terms*—covering, precision reducible formula, precision reduction, information quality**

I. INTRODUCTION

In the study of set theory, besides doing comparisons and calculations on different sets directly, usually divide large sets into smaller ones at first, which including covering and divide. Cover and divide are two associate and fundamental ways to split large sets into smaller ones, and also the basis of rough set theory.

Rough set theory is based on equivalence relations and is put forward by Pawlak at 1982. At the beginning, equivalence relations are found on sets' divide. But in order to deal with incomplete and complex information, the initial model is extended. As a result, rough set theory based on covering is put forward [1-3].

Covering reduction is the theoretical basis of rough set theory based on covering. In the view of importance of covering reduction, it is necessary for its in-depth research. In reference [4], promote reducible formula as core concept of covering reduction theory, and prove that every covering has one reduction. In this paper, we will give the definition of precision reducible formula, and put forward covering reduction method based on precision reducible formula, and discuss basic properties of covering reduction.

II. PRECISION REDUCIBLE FORMULA

Covering reduction theory are only sporadic in a number of documents appeared, and all of these are not systematic. Definitions 1 to definition 4 originate from reference [5-7], and are the core concepts of covering reduction theory.

Definition 1: Given universe of discourse U, C is a family of series subsets of U, and C doesn't include empty sets, $\cup C=U$, then we call C is a covering of U, <U,C> is a covering approximate space.

Definition 2: Given covering approximate space <U,C>, $x \in U$, $Md(x)=\{K \in C | x \in K \ x \in S \ S \ K \ K=S\}$, then we call Md(x) is the minimal description of x.

Definition 3: Given that C is a covering of U, $K \in C$, if K can be got by the union of the sets in C-{K}, then we call C is a reducible formula, otherwise we call C is a non-reducible formula.

Definition 4: Given that C is a covering of U, if every set in C is non-reducible formula, then we call C is simplified, otherwise we call C is non-simplified and reducible. After reduction, get a covering on U, we call it is a reduction of C which is denoted as red(C).

Example 1: U={a,b,c,d,e,f}, K_1={a,b}, K_2={a,c}, K_3={a,c,d,f}, K_4={b,d}, K_5={b,e,f}, K_6={a,b,c,d}, C={K_1,K_2,K_3,K_4,K_5,K_6}.

Because $K_6=K_2 \cup K_4$, then the reduction of C is red(C)={K_1,K_2,K_3,K_4,K_5}. □

According to the above discussion we can see that, if a subset family doesn't has a subset that can be got by union of other subsets, and then this subset family is simplified. After carefully study, the result has redundant information, and can be reduction further.

In Example 1, K_3 in the outcome red(C) is redundant, as we have Md(a)={K_1,K_2}, Md(c)={K_2}, Md(d)={K_4}, Md(f)={K_5}. In other words, the other subsets have all the information that K_3 contains. In application, apparently we need discard K_3. For the sake of application, it is necessary to find a new covering reduction method.

Definition 5: Given that C is a covering of U, $K \in C$, if $x \in K$, x can be minimal described by C-{K}, then we call K can be described by C-{K}, K is a precision reducible formula. Otherwise, we call K can not be described by C-{K}, K is a precision non-reducible formula. After reduction, we can get a covering of U, we call it is precision reduction of C, and denote as RED(C).

Proposition 1: Reducible formula must be a precision reducible formula.

If a subset K is a reducible formula, then k can be got by the union of several sets in C-{K}.Apparently, every element in K can be minimal described by these sets in C-{K}, and K must be a precision reducible formula.

Proposition 2: Non-reducible formula may be precision reducible formula.

Given K is a non-reducible formula, if every element in K can be minimal described by C-{K}, then K is a precision reducible formula.

Example 2: U={a,b,c,d,e,f}, K_1={a,b}, K_2={a,c}, K_3={a,c,d,f}, K_4={b,d}, K_5={b,e,f}, K_6={a,b,c,d}, C={K_1,K_2,K_3,K_4,K_5,K_6}.

$RED(C)_1$={K_2,K_4,K_5}, $RED(C)_2$={K_5,K_6}, $RED(C)_3$={K_3,K_5 }.

Both $RED(C)_1$ and $RED(C)_2$ as well as $RED(C)_3$ are the precision reduction result of C.

Ⅲ. COVERING'S PRECISION REDUCITON

Example 2 shows that the result of covering's precision reduction has uncertainty to some extent. So it's necessary to put some constraint to covering's precision reduction, and make the result more reasonable and certain.

A. Covering's precision reduction algorithm

From the intention of covering's reduction, the purpose of reduction is to eliminate redundant information, and make sure reduction result and the original cover has the same information quality. So we can use information quality to measure the effect of the covering's precision reduction and design reasonable algorithm.

Definition 6: universe of discourse U, and a covering $C=\{c_1,c_2,\ldots,c_r\}$, we define the information quality of covering C is $K(C)=\sum_{i=1}^{r}|c_i|*|U-c_i|$.

Intuitively we can see that information quality of a covering is determined by two factors: the number of subset and the Uniformity of the subset. More evenly distributed, more information quality a covering has; the contrary, less information quality the covering has. This is because the subset of family implies the classification information, and if a covering has plenty of evenly distribute subsets, it implies the covering has more classification information. Apparently, when $C=\{U\}$, $K(C)=0$. A detailed discussion about information quality, can refer to the reference [8-9].

After calculation, we get $K(RED(C)_1)=25$, $K(RED(C)_2)=17$, $K(RED(C)_3)=17$, result $RED(C)_1$ has the largest information quality.

According to the analysis on information quality, we can get the main idea of covering precision reduction algorithm: descending find precision reducible formula and eliminate it, until the set family doesn't have precision reducible formula, then the rest sets in the set family is the results of the covering's precision reduction.

Eliminate larger sets first is for the reason that: firstly, it aims at getting the largest information quality; secondarily, large set can be got by the union of smaller sets.

(1) Covering's precision reduction algorithm

Input: universe of discourse U, and its covering $C=\{c_1,c_2,\ldots,c_n\}$

Output: precision reduction of C (RED(C))

Step1 initialization: sort $c_1,c_2,\ldots,c_n$ in a descending order, denote as $c_1',c_2',\ldots,c_n'$, RED(C)=C, i=1;

Step2 if c_i' can be described by $RED(C)-c_i'$, turn to step3; otherwise turn to step4;

Step3 eliminate c_i', $RED(C)=RED(C)-c_i'$;

Step4 i=i+1, if i<=n, turn step2;

Step5 output RED(C), exit from the algorithm.

(2) Step2 refinement

c_i' contains n_i elements, that is $c_i'=\{c_{i1},c_{i2},\ldots,c_{ini}\}$, if every element $c_{ij}(j=1,2,\ldots,ni)$ belongs to $RED(C)-c_i'$, then c_i' can be described by $RED(C)-c_i'$.

Definition 7: Given a covering C, redundant information eliminated quality of precision reduction is $KRED(C)=K(C)-K(RED(C))$; redundant information eliminated quality of reduction is $Kred(C)=K(C)-K(red(C))$.

Definition 8: Given a covering C, compression rate of precision reduction is $QRED(C)=KRED(C)/K(C)$; compression rate of reduction is $Qred(C)=Kred(C)/K(C)$.

Example 3: $U=\{a,b,c,d,e,f\}$, $K_1=\{a,b\}$, $K_2=\{a,c\}$, $K_3=\{a,c,d,f\}$, $K_4=\{b,d\}$, $K_5=\{b,e,f\}$, $K_6=\{a,b,c,d\}$, $C=\{K_1,K_2,K_3,K_4,K_5,K_6\}$.

After calculation, we get $RED(C)=\{K_2,K_4,K_5\}$, $red(C)=\{K_1,K_2,K_3,K_4,K_5\}$; $KRED(C)=24$, $Kred(C)=8$; $QRED(C)=0.51$, $Qred(C)=0.16$.For papers published in translated journals, first give the English citation, then the original foreign-language citation [6].

B. Properties of covering's precision reduction

Proposition 3: Given a covering $C=\{c_1,c_2,\cdots,c_n\}$, if any two sets in C satisfies $|c_i|\neq|c_j|$,then there is only one result of the precision reduction of covering C.

Proposition 4: Given a covering, the result of the precision reduction may be different, but the information quality must be the same.

Proof: in precision reduction, descending arrange the sets in the set family, $c_1,c_2,\cdots,c_i,\cdots,c_j,\cdots,c_n$, $|c_1|\geqslant\cdots\geqslant|c_i|\geqslant\cdots\geqslant|c_n|$. Not lose generality, suppose we have reduction $c_1,c_2,\cdots,c_{i-1}$, the rest sets are $c_i,c_{i+1},\cdots,c_j,\cdots,c_n$. If $|c_i|=|c_{i+1}|=\cdots=|c_j|$, and there are at least two precision formulas in them, as the size of the sets are same, the then have the same information quality. Randomly select one set and eliminate, and therefore reduce of the information quality are the same. So, select different precision formula doesn't change information quality, and the proposition is proved. □

Example 4: $U=\{a,b,c,d,e\}$, $K_1=\{a,b\}$, $K_2=\{a,c\}$, $K_3=\{b,c\}$, $K_4=\{c,d,e\}$, $C=\{K_1,K_2,K_3,K_4\}$.

$RED(C)_1=\{K_1,K_2,K_4\}$, $RED(C)_2=\{K_1,K_3,K_4\}$, $RED(C)_3=\{K_2,K_3,K_4\}$, their information quality is the same, and is 18.

Proposition 5: Given a covering C, redundant information eliminated quality of precision reduction RED(C) is more than redundant information eliminated quality of reduction red(C), at least equal to the later.

According to Proposition 2, this proposition is easily proved. This proposition illustrates that precision reduction is better than reduction.

Ⅳ. CONCLUSION

Covering reduction based on reducible formula has redundant information, and isn't fit for application. In this paper, put forward the term precision reducible formula, and design precision reduction algorithm based on precision reducible, and use information quality to measure and compare the effect of precision reduction

and reduction. Result shows precision reduction is better than reduction.

Covering and division are two related methods in set theory, a lot of research has been done on division, but covering is overlooked all the time. So, it's necessary to carry on research roughly on covering as well as the on relations that are established on covering.

ACKNOWLEDGMENT

The work presented in this paper is supported by National Science Foundation of China (60575035, 605975033) and Shanghai University Innovation Foundation (SHUCX091010, A.16-0108-08-002).

REFERENCES

[1] Zhu Feng, Wang Feiyue, Some results on covering generalized rough stes[J] Journal of Pattern Matching and Artificial Intelligence, 2002,3(1):6-13(in Chinese).

[2] Pawlak Z. Rough Sets: Theoretical Aspects of Reasoning about Data. Dordrecht, Kluwer Academic Publishers, 1991.

[3] Wang Jue, Miao Duoqian, Zhou Yujian. Rough set theoty and its application: a survey[J] Journal of Pattern Matching and Artificial Intelligence, 1996,9(4):337-344 (in Chinese).

[4] Pei Daowu, Fu Li. On two kinds of generalized rough sets[J], Journal of Pattern Matching and Artificial Intelligence, 2004,9(3):281-285(in Chinese).

[5] Zakowski W. Axiamization in the Space(U,π).Demonstratio Mathematica,1983,16:761-769.

[6] Pomykala J A. Approximation Operations in Approximation Space. Bulletin of the Polish Academy of Science, 1987,35{9-10:653-662.

[7] Bonikowsld Z,Bryniariski E,Skardowska U W. Extensions and Intentions in the Rough Set Theory. Information Sciences, 1998,107:149-167.

[8] Chen Tangmin, Research of the heuristic reduced algorithm based on the separating capacity [J].Chinese Journal of Computers, 2006, 29(3):480-497 (in Chinese).

[9] Zhang Haiyun, Liang Jiye, Liang Chunhua, An attribute reduction algorithm based on the knowledge quantity [J]. Journal of Chinese Computer Systems, 2007,28(11):1968-1971(in Chinese).

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 028-032

Interference Analysis between the CDMA 1X and DO Co-site and Co-antenna

Zi-yi Fu [1], Yun Song [2]
[1] College of Computer Science & Technology, Henan Polytechnic University, Jiaozuo, China
Email: fuzy@hpu.edu.cn
[2] College of Computer Science & Technology, Henan Polytechnic University, Jiaozuo, China
Email: yuner20078@163.com

***Abstract*—This paper gives the interference theoretical analysis, under the co-site and co-antenna of the CDMA2000 1x and CDMA2000 1xEV-DO system. The results show that the reverse capacity will be reduced by 4.8%, and the system with the bilateral adjacent channel interference will be reduced by 9.2% under the co-site and co-antenna system. If the impact of the adjacent channel interference for coverage is very small, this change can be ignored.**

***Index Terms*—coverage; capacity; interference**

I. INTRODUCTION

With the growing of service requirement from the customer, many CDMA operators build CDMA 1x and CDMA 1xEV-DO network at the same time. The limited radio resources can not provide too much frequency to set up a wide protection band, so there is adjacent channel interference in the system.

Adjacent channel interference (ACI) is provided by out-of -band interference and spurious interference. One is the out-of-band interference. Another is spurious interference. It will lead to the decrease of the receiver sensitivity, and make the adjacent system performance decline, when signal level from this system exceeds a certain value of sensitivity adjacent to the receiver.

In order to avoid the interference between these systems, wireless devices have a stringent set of specifications.

This article will focus on analyzing and discussing the impact on the interference in the adjacent band between the two systems and the loss the system produces. Usually, there are two kinds of circumstances in adjacent channel interference. One is unilateral adjacent channel interference, which is the interference adjacent to each other between the two frequency points. The other is the bilateral adjacent channel interference.

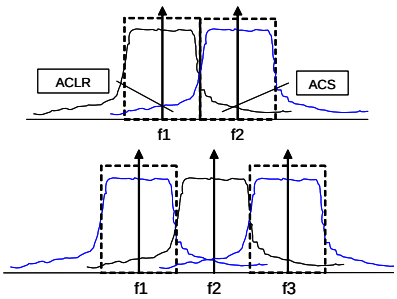


Figure 1. Diagram of the unilateral and bilateral interference

Adjacent channel interference is usually associated to indicators, including Adjacent Channel Leakage Ratio (ACLR) and adjacent channel selectivity (ACS). This is shown in Fig. 1. Adjacent Channel Leakage Ratio (ACLR) is used to measure the radiation characteristics of the transmitter-of-band, whose mainly reason is that the out-off-band power leakage results in the interference because of the non-ideal characteristics of the interference object of in the transmitter. Adjacent Channel Selectivity (ACS) measures suppression capabilities which receive filter produced to the adjacent channel, whose mainly reason is that it is influenced by a partial-band adjacent channel power brought by the non-ideal filter characteristics of the influenced object receive [1].

The mainly adjacent interference analysis of the article will focus on co-site and co-antenna scenario. This paper will only consider the same frequency band adjacent channel spectrum and no special instructions, analysis by default to 800MHz frequency band. The interference impact analysis takes into account the transmitter adjacent channel leakage ratio (ACLR) and the receiver adjacent channel selectivity (ACS). A unified by symbols λ (i.e. $\lambda = \lambda tx\_ach + \lambda rx\_acs$), at the following called adjacent channel interference factor. The results are not the same, because the equipment relevant indicators are slightly different in different manufacturers and band frequency. So the following related indicators are only a reference. The Coverage area due to different propagation model is different, the analysis using the classical propagation model Okumura-Hata as a reference.

II. IMPACT OF ADJACENT CHANNEL INTERFERENCE WITH CO-SITE

A. Insertion loss with co-site and co-antenna

The co-site network is divided into co-antenna system and un-co-antenna system. Both of them in the adjacent channel interference analysis is basically similar, the difference is that for co-antenna system the impact of a certain insertion loss needs to be considered because of adding a set-top sharing module but for un-co-antenna system, there are no such problem but it need to consider, the insulation between the antennas of mutual independent work Increasing the set-top shared module

will bring in a certain degree of insertion loss, thereby increasing the loss in reverse link. Usually, under the transmission signal the merger of the insertion loss is 0.4dB, whose influence is small and can be ignored, and the merger of the insertion loss for the receiving signal is around 3.5dB~4.0dB. It will increase around 3.5~4.0dB when using the passive program in order to achieve the diversity receiving and reduce the interference noise. It could be considered an LNA in front of the receiver, in order to compensate for loss caused by noise, such as the choice gain 12dB, noise figure of the LNA is 0.8dB, noise figure can be improved by 2.3~2.8dB, when it includes the set-top sharing of modules and LNA is the system sensitivity only reduced 1.5dB. To some extent, this change can be accepted. For 1x or DO, the insertion loss will affect coverage of regional, the efficiency of the edge of coverage and the quality of service, where insertion loss is mainly about the introduction of performance produced coverage [2].

The following Table.1 is coverage theoretical analysis of the impact of Insertion Loss (co-site and co-antenna) Okumura-Hata propagation model formula:

$$L\ (dB) = 69.55 + 26.16 \times lgf_c - 13.82 \times lgh_b - \alpha\ (h_m) + 44.9 - 6.55 \times lgh_b \times lgd - K \quad (1)$$

Where the parameter f_c is the working frequency (MHz), $h_b(m)$ is the base station antenna effective height, $h_m(m)$ is the terminal antenna effective height and $d(km)$ is the horizontal distance between the base station antenna and the terminal antenna.

$$K = \begin{cases} 4.78 \times (\log_{10} f_c)^2 - 18.33 \times \log_{10} f_c + 40.94 & Open\ area \\ 2 \times [\log_{10}(f_c / 28)]^2 + 5.4 & Suburban \\ 0 & Urban \end{cases} \quad (2)$$

When in the Middle and Small city, in this case we obtain the following expression:

$$\alpha\ (h_m) = (1.1 \times log_{10} f_c - 0.7) \times h_m - (1.56 \times log_{10} f_c - 0.8) \quad (3)$$

When in the big city, we have

$$\alpha\ (h_m) = 3.2 \times [log_{10}(11.75 \times f_c)]2 - 4.96 \quad (4)$$

The coverage impacts uplink more than downlink due to insertion loss. It can be seen from the analysis of the Table.I.

For the urban and dense urban areas, the coverage radius of the cell is about 200m ~ 700m. Take 700m for example, the loss of reverse link path is 121.3dB, when the loss 3.5 ~ 4dB introduced by the antenna insertion makes the radius reduce to 559m, adding the LNA will improve 1.5dB and cell radius will reduce to 636m.

For the suburban and rural areas, the cell coverage radius is more than 1000m. The co-antenna the radius will reduce to 798m, when adding the LNA the cell radius will reduce to 908m.

To sum up, the insertion loss of the antenna the coverage ratio basically makes the same effects for different coverage area, and the absolute value of the impact of different coverage radius is not same. For the link budget, the main consideration is the forward link power-constrained because the reverse link mobile terminals are non-firing full-power state. The co-site with the antenna LNA cases, 1.5 dB of the impact of insertion loss can be ignored. For wide coverage, the main consideration is reversely limited in link budget, but when adding LNA, the impact of insertion loss to the antenna, 1.5dB will cause the reduce of the coverage radius, which can't be ignored.

TABLE I.
COMPARE THE IMPACT OF INSERTION LOSS WITH DIFFERENT ENVIRONMENT

Configuration	Coverage area	Dense urban and urban	
	Coverage radius d(m)	200 ~ 700	
	EdgeCoverage Probability	90%	
	Link	Forward	Reverse
NO LNA	The changes of loss（dB）	↑0.4	↑3.5
	The changes of Coverage radius （m）	5~18 (↓2.6%)	40~141 (↓20.1%)
With LNA	The changes of loss（dB）	↑0.4	↑1.5
	The changes of Coverage radius d(m)	5~18 (↓2.6%)	18~64 (↓9.1%)
Configuration	Coverage area	Suburban and Rural	
	Coverage radius d(m)	1000 ~ 5000	
	Edge Coverage Probability	75%	
	Link	Forward	Reverse
NO LNA	The changes of loss（dB）	↑0.4	↑0.4
	The changes of Coverage radius （m）	25~127 (↓2.5%)	25~127 (↓2.5%)
With LNA	The changes of loss（dB）	↑0.4	↑0.4
	The changes of Coverage radius （m）	25~127 (↓2.5%)	25~127 (↓2.5%)

B. Analysis the Impact of coverage and capacity with co-site and co-antenna

When the 1x and DO work in the same frequency band, because of various out-of-band radiation and spurious interference, there is a certain of interference in the devices, so they will influent the system coverage or capacity [3] .

1) Analysis on impact of coverage

The interferences are of two main types. One is forward, it is meant the 1x base stations interfere with DO terminals, while DO base stations interfere with 1x terminal. The other is reverse; it is meant the 1x terminal interference DO base stations, while DO terminals interfere with 1x base stations.

In the CDMA system, all users transmitting in the same band, each user's signal interference with other users, so a total interference power received by forward and reverse link and thermal noise power can be expressed as follows.

In the Pilot channel, the forward can be expressed:

$$I_{t\text{-}f} = N_{o\text{-}m} + P_{host} + \beta \cdot P_{host}. \quad (5)$$

In the traffic channel, the forward can be expressed:

$$I_{t\text{-}f} = N_{o\text{-}m} + \xi P_{host} + P_{host}\beta. \quad (6)$$

For the Reverse, this equation is expressed as follow:

$$I_{t\text{-}r} = N_{o\text{-}b} + I_{sc} + I_{oc}. \quad (7)$$

Where $I_{t\text{-}f}$ is the total forward link interference power, $N_{0\text{-}m}$ is thermal noise power of the terminal, P_{host} is a total power the user receives in this district, ξ is the Orthogonal factor in the forward Traffic Channel (a value is 0~ 1), β is the interference factor in other cells. $I_{t\text{-}r}$ is the total reverse link interference power, $N_{o\text{-}b}$ is the thermal noise power of the base station, I_{sc} is interference power of the based cell, I_{oc} is the interference power in other cells.

Assuming that link budget, the situation of co-site with difference antenna, the user distribution and user load is the same. The total interference power for forward and reverse are as follows.

In the Pilot channel, the forward can be expressed:

$$I'_{t\_f} = N_{o\_m} + P_{host} + \beta P_{host} + \lambda_b P_{host} + \lambda_b \cdot \beta \cdot P_{host}. \quad (8)$$

In the traffic channel, the forward can be expressed:

$$I'_{t\_f} = N_{o\_m} + \xi P_{host} + \beta P_{host} + \lambda_b P_{host} + \lambda_b \beta . P_{host.} \quad (9)$$

For the Reverse, this equation is expressed as follow:

$$I'_{t\_r} = N_{o\_b} + I_{sc} + I_{oc} + \lambda_m I_{sc} + \lambda_m I_{oc.} \quad (10)$$

Where λ_b is adjacent channel interference factor of the base station, λ_m is the adjacent channel interference factor in terminal. The uplift impact of the adjacent channel interference on of receivers is expressed as follows.

In the Pilot channel, the forward can be expressed:

$$R_{im\_f} = \frac{I'_{t\_f}}{I_{t\_f}} = \frac{N_{o\_m} + P_{host} + \beta P_{host} + \lambda_b P_{host} + \lambda_b \beta . P_{host}}{N_{o\text{-}m} + P_{host} + \beta \cdot P_{host}} \quad (11)$$

In the traffic channel, the forward can be expressed:

$$R_{im\_f} = \frac{I'_{t\_f}}{I_{t\_f}} = \frac{N_{o\_m} + \xi P_{host} + \beta P_{host} + \lambda_b P_{host} + \lambda_b \beta . P_{host}}{N_{o\text{-}m} + \xi P_{host} + P_{host}\beta} \quad (12)$$

For the Reverse, this equation is expressed as follow:

$$R_{im\_r} = \frac{I'_{t\_r}}{I_{t\_r}} = \frac{N_{o\_b} + I_{sc} + I_{oc} + \lambda_m \cdot I_{sc} + \lambda_m \cdot I_{oc}}{N_{o\_b} + I_{sc} + I_{oc}} \quad (13)$$

In the circumstance that the thermal noise power for the total received interference power can be ignored when it is smaller, and the user at the cell edge, at this time the interference from the other cell will be dominant.

For the Forward, this equation is expressed as follow:

$$R_{im\text{-}f} \approx 1 + \lambda_b. \quad (14)$$

For the Reverse, this equation is expressed as follow:

$$R_{im\text{-}f} \approx 1 + \lambda_m. \quad (15)$$

Reference Okumura-Hata propagation model, the influence on the uplift of floor noise for unilateral adjacent channel interference is illustrated in Table.II. [3]. Base on the theoretical analysis, in the case of co-site, for unilateral adjacent channel interference, and the noise floor of the base station receive in the reverse link r increase about 0.21dB, and the noise floor of terminal receiver increase about 0.11dB in the forward link. Here adjacent channel situation will be considered. The interference factor of the bilateral adjacent channel interference is about twice time more than the unilateral Base on the above equation, it can be seen that the impact of noise floor is still relatively small (forward is about 0.21dB, in this case the reverse is about 0.41dB). Because of the difference of performance between the base station and terminal, the adjacent channel interference on the reverse link is more than the uplink.

2) Analysis the impact of Capacity

The ultimate capacity of CDMA 1x voice model can be known, according to the interference of the reverse traffic model [4].

The formula of the ultimate capacity of the reverse for the Omni coverage of the cell is expressed as follows.

$$N_{max} \approx \frac{G_p}{a \cdot d} \cdot \frac{1}{1+\beta}. \quad (16)$$

(when sector direction coverage, $N_{max\_s} \approx N_{max} \cdot G_s$)

Where: G_P is the spreading gain ($G_P = W/R$ is 1.2288MHz, R for traffic rates, voice traffic of 9.6kbps), a voice-activating factor, default value is 0.4, d is the receiver demodulation threshold E_b/N_o, G_s is the fan factor. For the three sectors, G_s = 2.55.

The ACI of the co-site as follows.

$$N'_{max} \approx N_{max} \cdot \frac{1}{1+\lambda_m}. \quad (17)$$

When Sector direction coverage, the adjacent channel interference can be expressed as follow:

$$N'_{max\_s} \approx G_s \cdot N'_{max}. \quad (18)$$

TABLE II.
COMPARE THE IMPACT OF COVERAGE WITH DIFFERENT ENVIRONMENT

fc(MHz)=850, h_b (m)=30m, h_m(m)=1.5m , λ_b =-16dB, λ_m =-13dB		
Coverage area	Dense Urban and urban	
Coverage radius d(m)	200 ~ 700	
Edge Coverage Probability	90%	
Link	Forward	Reverse
The change of receiver noise floor	↑0.11dB	↑0.21dB
The changes of Coverage radius d(m)	1~5 (↓0.7%)	3~9 (↓1.5%)
Coverage area	Suburban and Rural	
Coverage radius d(m)	1000 ~ 5000	
Edge Coverage Probability	75%	
Link	Forward	Reverse
The change of receiver noise floor	↑0.11dB	↑0.21dB
The changes of Coverage radius d(m)	7~36 (↓0.7%)	14~68 (↓1.4%)

By equation (16), (17), the impact of capacity with the co-site stations is expressed. In this case, the impact of system limit capacity reduces 4.8% compare with the original. The impact of interference on the capacity to be know through the relationship between interference and system capacity.

The interference margin can be expressed in difference situation. Its computation formula is as follows:

Non-adjacent channel interference:

$$R_{im\_r} = \frac{1}{1-X} \cdot \tag{19}$$

Unilateral adjacent channel interference:

$$R_{im\_r} = \frac{1}{1-X-\lambda_m \cdot X} \cdot \tag{20}$$

Bilateral adjacent channel interference:

$$R_{im\_r} = \frac{1}{1-X-2\lambda_m X} \cdot \tag{21}$$

Where X is the system load; λ_m is the interference factor for the terminal adjacent channel. Base on the formulas above, we can show the following curve. This result is shown in Fig.2. (To simplify the expression, with N-ACI expressed the non-adjacent channel interference, with U-ACI expressed the unilateral adjacent channel interference, with B-ACI expressed bilateral adjacent channel interference, V-ACI expressed the different between N-ACI and B-ACI) . CDMA system with 50% load (interference margin is 3dB) is designed as a reference. In the same user and the same coverage area, because of the adjacent channel interference, the network load is only 47.6%, 4.8% of the system capacity is reduced. From the previous figure, with the users increasing, the relative values of influence which adjacent channel interference impact on system capacity is the same, but the absolute capacity of the impact is raising.

With the different of load, the adjacent channel interference system margin computation formula can be expressed as following.

$$R_{im\_r} = \frac{1-X_2+\lambda_m X_2}{(1-X_1)(1-X_2)-\lambda^2_m X_1 X_2} \cdot \tag{22}$$

Take X_2 at different loading, the variance of capacity vary with distinct loading. The relation of loading and interference can be seen from Fig.3, the larger network load of source of the disturbance, the larger impact which is generated by the disturbed system. The design of 50% load system is referenced, when the interference system load reaches 70%, the impact of the adjacent channel interference on the capacity is 12.2%.

According to the theoretical analysis results, in the same load condition and same coverage area, non-adjacent channel interference of the system loading can be reached 50%, in the Bilateral adjacent channel interference situation it only can be reached 45.4% and system capacity is relatively decreased by 9.2%. These variables are related by the same loading given in Fig. 4. And From Table III shows the compare of the capacity. From the previous figure, with the users increasing, the relative values of influence which adjacent channel interference impact on system capacity is the same, but the absolute capacity of the impact is raising.

III. CONCLUSIONS

They need an addition share set-top module for the co-site with same antenna system, there are more impact for uplink capacity. Its values range from 3.5dB to 4.0dB. If add an addition LNA, the insertion loss can reduce about to 1.5 dB, which is can be negligable in dense urban with height traffic density. The design and installation only need to follow the principles that avoid interference and ensure certain of antenna isolation. The impact of the adjacent channel interference can be negligible when co-site with different antenna system.

The main impact of unilateral adjacent channel interference is for capacity with the same coverage and user. The reverse capacity will be reduced 4.8% in theoretical analysis and capacity will be reduced 9.2% with the bilateral adjacent channel interference.

TABLE III.
THE COMPARE OF THE CAPACITY

X_2	0%	40%	60%	70%
X_1	50%	48.3%	46.1%	43.9%
($\Delta X1/X1$)%	0%	3.4%	7.8%	12.2%

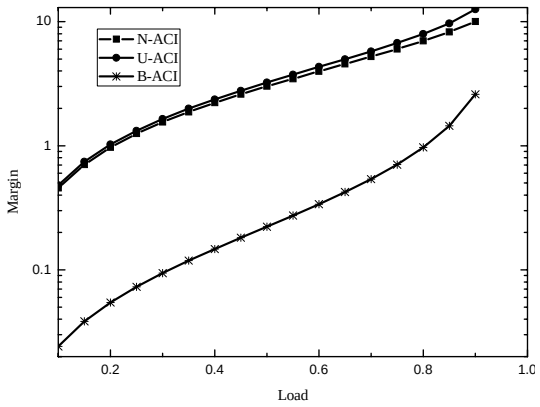


Figure 2. Capacity changes with same loading

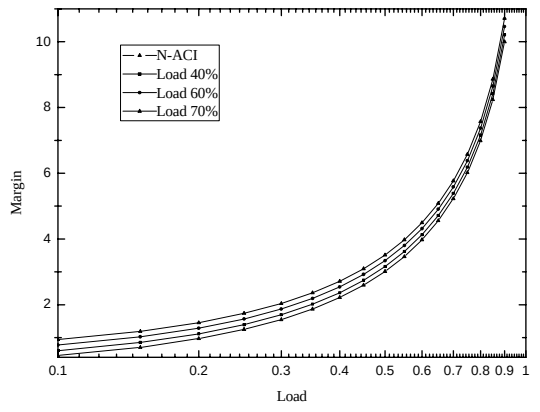


Figure 3. Capacity changes with difference loading

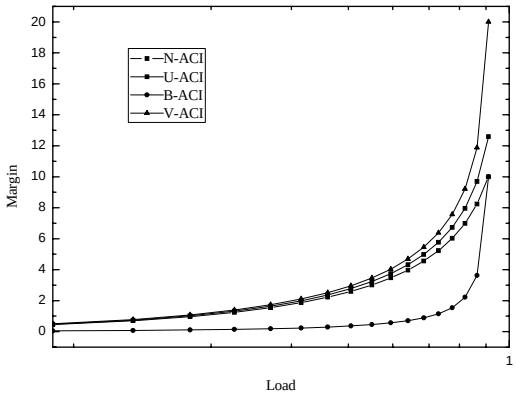


Figure 4. Capacity changes with same loading

ACKNOWLEDGMENT

The project of this thesis is supported by the funds from Henan Education Department (serial number: 2008B470002) and Major key project in Henan Province(serial number: 082102210079).

REFERENCES

[1] ZhangChuanfu, “cdma2000 1x/EV-DO communication network planning and design,”Posts & Telecom Press, vol. 11, 2009.

[2] 3GPP2. C.S0033-A v2.0. Recommended Minimum Performance Standards for cdma2000 High Rate Packed Data Access Terminal. USA, 2007.

[3] MaYue, XieWeihao, JiangMinggang, “CDMA2000 1X EV-DO wireless network planning features and recommendations,” vol. 09, pp. 66-69, 2006.

[4] WuChangguo, JiGuoqing, JiangGuangxing “CDMA and WCDMA system, adjacent channel interference problems ofoutdoor,”vol.27,pp.49-54,Dec2007.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15, August 2010, pp. 033-036

Distributed Trust Rating Scheme Based on Feedback Confidence over P2P Networks

Jianli Hu[1], Bin Zhou[2], Xiaohua Li[1], Yonghua Li[3]
[1]Information Department, Guangzhou General Hospital of Guangzhou Military Command, Guangzhou 510010, China
[2]School of Computer, National University of Defense Technology, Changsha 410073, China
[3]Department of Hematology, Guangzhou General Hospital of Guangzhou Military Command, Guangzhou 510010, China
lxman82@gmail.com

***Abstract*—An important challenge regarding peer's trust valuation in peer-to-peer (P2P) networks is how to cope with such problems as dishonest feedbacks of malicious peers and collusions, which cannot be effectively tackled by the existing solutions. So a feedback confidence (FC)-based distributed trust management scheme for P2P networks, named FCTM for short, is proposed to quantify and evaluate the trustworthiness of peers. In FCTM, the factor of FC is introduced to be used as the confidence metric for depicting the extent to which the feedback peer trusts its feedbacks, which is related to three aspects, including the density of interaction experiences with the trustee, the altering scope of interaction experiences and the interacting time. Besides, we consider the feedback credibility (RC), which is constructed with the similarity function to describe the veracity of its feedbacks, and time fading characteristics of trust. Theoretical analyses and experimental results demonstrate that FCTM has advantages in combating some malicious activities over the existing models, and show more robustness and effectiveness.**

***Index Terms*—P2P, trust management, reputation, feedback confidence**

I. INTRODUCTION

Most of the existing reputation-based trust models [1-4] regard the trust value of one peer as the index to choose the service needed, in other words, they compute the trusted rank of the peer with its transaction histories with others. The peer with the highest trust value will be preferred when there are many options. To a certain degree, this approach has some effects on the simple malicious behavior patterns, but has little effects in coping with the complex attack and disturbance activities on reputation systems [2].

Therefore, in this paper, we proposed a FC-based distributed trust model for P2P networks. FCTM introduces the index of FC as the confidence metric of its feedbacks to others, and uses the factor of RC as the veracity metric of its feedbacks, which put different weights to feedbacks from different peers. Furthermore, in order to reasonably distinguish the degree, to which the transactions in different periods affect the computation of trust values, this model also introduces the concept of time fading feature. Compared with the existing trust model such as EigenTrust [2], simulation experimental results reveal that FCTM exhibits more robustness and effectiveness in combating some malicious behaviors such as the dishonest feedbacks and collusions from malicious peers.

The rest of the paper is structured as follows. Section 2 formally introduces our trust management model FCTM. Section 3 simulates and discusses FCTM. Finally, we conclude the paper.

II. FEEDBACK CONFIDENCE BASED TRUST MODEL

Definition 1 (Peer's Trust Value) The trust value of one peer is composed of two parts: the direct trust value (DTV) and the indirect trust value (ITV). DTV and ITV are used to describe the trust ratings that the trustor gives to the trustee according to the direct transaction experiences with the trustee, and the rating information from feedback peers (recommenders), respectively. Let i, j, k denote trustor, trustee and the recommender respectively, T_{ij} denote the trust value peer *i* puts to peer *j*, and we can get it from the following computing formula:

$$T_{ij} = \alpha * D_{ij} + (1-\alpha) * R_{ij} \qquad \alpha \in [0,1] \tag{1}$$

in which D_{ij} denotes DTV peer *i* puts to peer *j*, and R_{ij} denotes ITV peer *i* puts to peer *j*; α is the trust regulatory factor.

Definition 2 (DTV) After transacting with each other, one peer (the service consumer) will submit its ratings of satisfactory degree to the other peer (the service provider), which can be defined as the following map function $f(i,j)$:

$$f(i,j) = \begin{cases} 1, & totaly\ \ satisfactory \\ 0, & tataly\ \ unsatisfactory \\ e(\in (0,1)), & else \end{cases} \tag{2}$$

In the time fraction *t*, supposing *m* denotes the number that peer *i* has interacted with peer *j*, so the DTV peer *i* puts to peer *j* can be defined:

$$D_{ij}^{t} = \begin{cases} \dfrac{\sum_{k=1}^{m} f(i,j)}{m}, & m \neq 0 \\ 0, & m = 0 \end{cases} \tag{3}$$

Thus, the final DTV model is defined as follows:

$$D_{ij} = \frac{\sum_{k=1}^{n} g_k * D_{ij}^{t_k}}{\sum_{k=1}^{n} g_k} \tag{4}$$

in which $g(k) = \rho_{fade}^{n-k}$ is the fading factor within the time fraction t_k, and $0 < f_k < f_{k+1} < 1$, $1 \le k < n$.

Definition 3 (ITV) ITV denotes the trust evaluation to the trustee, in which the trustor aggregates the ratings (DTVs) from different recommenders and ratings the trustor puts to all the recommenders. ITV is relevant to several factors, including DTV the recommender puts to the trustee, and the RC and FC of the recommender. Moreover, it has the feature of time fading, meaning the newly recommending actions deserve more trustworthiness. So we define ITV that peer *i* puts to peer *j* as follows:

$$R_{ij} = \frac{\sum_{k \in K} D_{kj} * Cr_{ik} * RCon_{kj} * g_k}{\sum_{k \in K} Cr_{ik} * RCon_{kj} * g_k} \quad (5)$$

in which, peer *i, j, k* denote trustor, trustee and the recommender, respectively, *K* denotes the recommender set, $RCon_{kj}$ denotes the FC when peer *k* submits the DTV of peer *j* to peer *i*, and Cr_{ik} is the RC peer *i* puts to peer *k*. We give the trustworthier recommender with the higher FC a larger weight in formula (5), and the DTV from this recommender is more credible.

Definition 4 (Feedback confidence) FC is used to depict the confidence metric of the recommender when providing feedback information to others. In computing ITV, we introduce this index as a weight of the DTV to the trustee submitted by the recommender, which is related to the QoS of feedback information. FC is relevant to the following factors:

(1) The transaction frequency with the trustee. Normally, the higher the frequency, the stronger the FC is. (2) The difference between DTV and ITV. The recommender can also obtain ITV information of the trustee with the identity of a trustor. The difference can affect FC of the recommender, and the smaller the difference, the stronger the FC is. (3) The time when the transaction happens. Recent transactions have greater effect on the recommenders' FCs. (4) The consistency of the trustee's behaviors. The more consistent the trustee's behaviors, the stronger the FC is.

We introduce the transaction density factor $TNum_{ij}$ to describe the transaction frequency between peer *i* and peer *j*, which can be defined:

$$TNum_{ij} = \frac{m}{n} * \beta^{\frac{1}{m}} \quad \beta \in (0,1) \cap m \neq 0 \quad (6)$$

in which *m* denotes the transaction number between peer *i* and peer *j*, and *n* denotes the total transaction number of peer *i* with other peers.

The variable used to describe the difference between DTV and ITV is named the transaction difference factor denoted by $TDif_{ij}$, which depicts the consistency of peers' actions, and correlates with the factor of time. Supposing *i* and *j* denote the trustor and the trustee, respectively; $K = \{k_1, k_2, \cdots, k_n\}$ denotes the recommender set, and the corresponding transaction time fraction set is $TSpan = \{t_1, t_2, \cdots, t_n\}$. So we define this variable as:

$$TDif_{ij} = \frac{\sum_{k \in K, t \in TSpan} g_k * \left| D_{ij} - D_{kj} \right|^{1/2}}{\sum_{t \in TSpan} g_k} \quad (7)$$

Finally, based on the above two factors, we can define the FC $RCon_{ij}$ peer *i* puts to peer *j* as:

$$RCon_{ij} = TNum_{ij} * * TDif_{ij} \quad (8)$$

Therefore, we can see from the above analyses, the larger the transaction number is; or the smaller the difference between DTV and ITV of a recommender, and or the more consistent the trustee's actions, the larger the value of $RCon_{ij}$ is, and in other words, the stronger FC is.

Definition 5 (Feedback Credibility) RC is used to describe the veracity of the feedback information. Here, we use the correctional cosine similarity measure to build the RC computing model, which is as follows:

$$Cr_{ik} = \frac{Sim(i,k)}{|CS|} \quad (9)$$

in which, $Sim(i,k) = \frac{\sum_{c \in CS} (D_{ic} - \overline{D_i}) * (D_{kc} - \overline{D_k})}{\sqrt{\sum_{c \in B} (D_{ic} - \overline{D_i})^2} * \sqrt{\sum_{c \in KS} (D_{kc} - \overline{D_k})^2}}$ denotes the correctional cosine similarity function, which utilizes the way of average rating to avoid the errors deriving from different rating angles. IS and KS denote the peer sets, which has ever interacted with peer *i* and peer *k*, respectively; $CS = IS \cap KS$ denotes the common peer set, which has ever interacted with peer *i* and peer *j*; $\overline{D_i}$ and $\overline{D_k}$ denote the average value of DTV peer *i* and peer *k* put to the peers in CS, respectively.

III. EXPERIMENTAL EVALUATION

We apply the file sharing application as the simulation case. The detailed simulation setup is shown in Table I. In simulation, assuming that all the files can be located successfully, that each file is at least possessed by one normal peer, and that the newly joined peer has a probability of 10% to be chosen as the service provider. Here, we simulate 100 query cycles, and each peer can execute transactions for 100 times.

TABLE I. SIMULATION SETTINGS

N	# of the total number of peers in community	1000
N_f	# of the total number of files	10000
P_{res}	% of the probability in response to query requests	1
α	% of the trust regulatory factor	0.5
ρ_{fade}	% of the time fading rate	0.9
β	% of the transaction density regulatory constant	0.5
TTL	# of the forwarding depth of query requests	4
D	# of the adaptive time window	100

To compare, we also simulate EigenTrust trust model. The evaluation standard is the successful transaction rate (STR), which is described as the percentage of the number of successful transactions to the total transaction number. The index of STR intuitionistically reflects the applying effect of the trust model.

A. Behavior Pattern Definition

In P2P networks, the malicious peer may provide malicious services or dishonest feedbacks to others. Here, We choose two malicious behavior patterns to evaluate FCTM as follow:

(1) The simply malicious peer, being the basic type of malicious peer, only provides malicious uploading service, which is named SMP for short. (2) The dishonest feedback peer only present dishonest feedbacks to other

peers, including two types, namely, the exaggerated feedback and the denigrated feedback (both named DFP). (3) The collusive malicious peers are organized into a group, which give exaggerated ratings and highly trusted services to members within the group, and denigrated ratings and malicious services to members outside the group. We named this type of peer as CMP.

B. SMP Simulation and Discussion

This experiment is mainly used to evaluate the effectiveness of FCTM and EigenTrust, when there only exist varying scale SMP peers in the simulated community. As a comparison, we also simulate under the circumstances of not deploying any reputation system (named this setting as *without RS*). The setting without RS means no trust model is deployed into the experimental system, and each peer randomly chooses downloading sources to download. We can see from Fig. 1: Without any malicious peers, STRs of the three experimental results are close to 100%. With the increasing of the number of SMP peers, the curve labeled by Without RS drops most quickly. When the percentage of SMP peers reach 50%, the STR for *without RS* drops to only about 20%. However, the same values keep over 60% for the other two models. The results show that with the varying SMP peers, the STR for EigenTrust drops greatly for the fact that it has no enough punishing mechanisms against this malicious behavior. While FCTM utilizes the approach of aggregating the DTV and IDV, and the RB and RC mechanisms, to get the more veracious ratings, and recognize and withhold SMP peers, it keep its STR in a higher level.

C. DFP Simulation and Discussion

The malicious behaviors of DFP peers are harder to be recognized. As the role of service provider, it can provide the highly trusted service, and keep its trust value in some level. However, as the role of feedback provider, it provides dishonest feedbacks. As for the varying scale DFP peers (Here, we only discuss the DFP peers with exaggerated feedbacks, and for the denigrated type, we can get the similar results), our experimental results are illustrated in Fig. 2. Because EigenTrust only simply substitutes the quality of services (QoS) for the quality of feedbacks, without the concept of RC, it is not very obvious for the dishonest feedbacks to affect the STR of EigenTrust. As far as FCTM is concerned, as shown in Fig. 2, with the increasing of the percentage of DFP peers, its STR drops not too much. When DFP peers account for up to 50% of all peers, FCTM's STR still reaches 86%. The reason is that FCTM can make use of the transaction difference factor in FC to distinguish the behaviors of dishonest recommenders effectively, and restrain the negative effect of DFP peers. Furthermore, FCTM can exploit the RC mechanism to verify further the veracity of the feedback information. The dishonest feedbacks will enlarge the similarity difference, and reduce RC. So the negative effect of DFP peers can be suppressed effectively. Moreover, confronted with the complex strategic dishonest recommenders such as the recommenders who start to submit dishonest feedbacks after obtaining a higher ITV by submitting honest feedbacks, or the opposite case, FCTM has shown its sensitiveness and effectiveness. We simulate for the two cases when the two types of peers account for 80%, respectively, and the results in Fig. 3 prove the above conclusions.

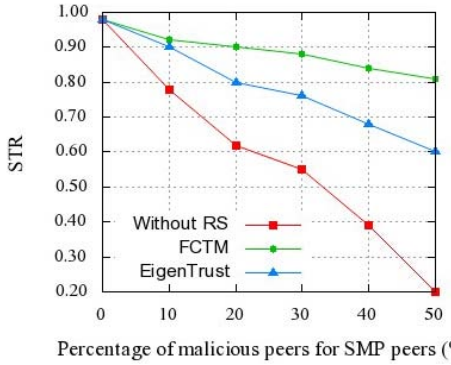


Figure 1. STR vs. different percentage of SMP peers

D. CMP Simulation and Discussion

CMP peers are familiar with each other in their group, and will produce more threats to the trust model itself. From Fig. 4 we can see that CMP peers can easily get a large trust value. The STR of EigenTrust decreases evidently without effective punishing mechanisms to CMP peers. Oppositely, FCTM is integrated with the mechanism of RB, which can effectively cope with the malicious behavior pattern of CMP peers. The detailed discussions are as follow:

(1) FCTM can effectively suppress the dishonest recommendation behaviors of CMP peers by utilizing the index of RB. If the CMP peer gives exaggerated ratings to the member inside the group, then the value of its transaction difference factor may become larger, which will decrease the value of RB, and so do the ITV and the trust value of the trustee. So, the RB has a punishing effect on the malicious behaviors of CMP peers. Regarding the denigrated type of CMP peers, we can also get the similar results. (2) Similarly, FCTM can take advantage of the correctional cosine similarity function in RC, effectively distinguishing the collusive behaviors by comparing the ratings to the common transaction peer set from the members inside the group and outside the group, respectively.

IV. CONCLUSIONS

In this paper, we proposed a FC-based distributed trust model for P2P networks FCTM introduces the index of FC as the confidence metric of its feedbacks to others. This index is related to three aspects, including

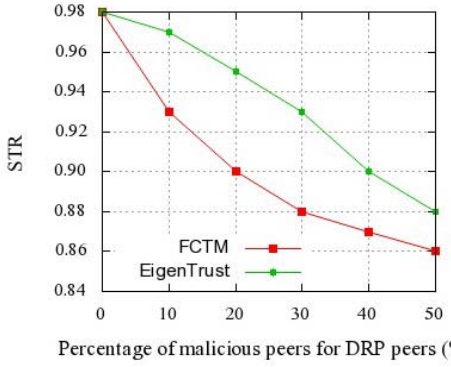


Figure 2. STR vs. different percentage of DFP

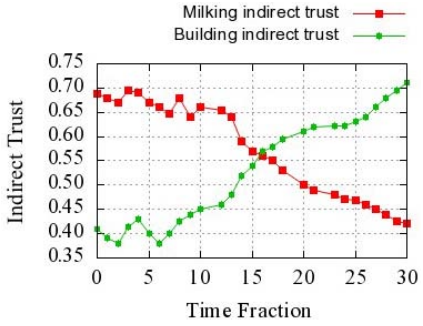


Figure 3. ITV vs. strategic dishonest recommenders with time

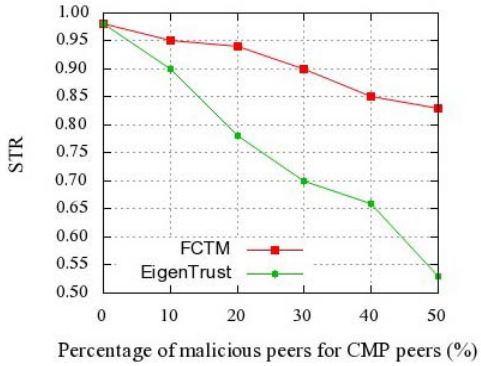


Figure 4. STR vs. different percentage of CMP peers

the density of interaction experiences with the trustee, the altering scope of interaction experiences and the interacting time. Besides, we consider the feedback credibility, which is constructed with the similarity function to describe the veracity of its feedbacks. Furthermore, in order to reasonably distinguish the degree, to which the transactions in different periods affect the computation of trust values, this model also introduces the concept of time fading feature. Compared with the trust model EigenTrust, simulation results demonstrate that FCTM has marked advantages against some malicious behaviors in P2P networks, and show more effectiveness and robustness, and can be easily integrated into some existing distributed engineering applications.

ACKNOWLEDGMENT

The authors acknowledge the useful comments from the anonymous reviewers. This research was supported by the doctorial start-up project of the Natural Science Foundation of Guangdong Province under Grand No. 9451001002003920, the National Grand Fundamental Research Program (973 Program) of China under Grand No. 2005CB321800, the National High Technology Research and Development Program (863 Program) of China under Grant No. 2007AA010301, and the National Science Foundation for Distinguished Young Scholars of China under Grant No. 60625203.

REFERENCES

[1] Dou Wen, Wang Huaimin, Jia Yan, et al. A feedback-based peer-to-peer trust model [J]. Journal of Software, 2004 ,15(4): 571-583 (in Chinese)

[2] Kamwar S. D, Schlosser M. T, Hector Garcia-Molina. The EigenTrust algorithm for reputation management in P2P networks. In: Proceedings of the 12th International Conference on World Wide Web, Budapest, Hungary, 2003 , 640-651

[3] Damiani E, De Capitani di Vimercati S, Paraboschi S, Samarati P. Managing and sharing servents' reputations in P2P systems. IEEE Transactions on Data and Knowledge Engineering, 2003, 15(4): 840-854

[4] Xiong L, Liu L. PeerTrust: Supporting reputation-based trust in peer-to-peer communities. IEEE Transactions on Data and Knowledge Engineering, Special Issue on Peer-to-Peer Based Data Management , 2004 , 16(7) : 843-857

[5] C Dellarocas. Immunizing online reputation reporting systems against unfair ratings and discriminatory behavior[C]. ACM Conf on Electronic Commerce, Minneapolis, Minnesota, USA, 2000

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 037-040

The Synchronization Algorithm of IEEE802.11a System

Mao Yan[1], Xu Qi[2], Zhang Chang-sen[2]
[1] Wanfang College of Science and Technology, Henan Polytechnic University,Jiao Zuo,China
Email: myzcs@hpu.edu.cn
[2] College of Computer Science & Technology, Henan Polytechnic University,Jiao Zuo,China
Email: xuqitongxin@163.com, zhangchangsen@hpu.edu.cn

***Abstract*—The standard of IEEE802.11a was introduced first in this paper.And then presents the design and implementation of a simulation model for physical layer in the WLAN system with the simulation tool MATLAB and has realized the simulation, which include timing synchronization and frequency synchronization. The simulation result shows that the algorithm presented possesses low complexity in implementation as well as high precision.**

***Index Terms* -IEEE802.11a; algorithm; synchronization**

I. INTRODUCTION

IEEE802.11a working in the 5GHz band was a higher speed standard version of the IEEE802.11 protocol suite,which maximum rate up to 54Mbps in Physical layer.It had he advantage of using OFDM technology,which basic principle was transmiting high speed serial data stream into many parallel low speed subdata stream.The subcarriers were orthogonal between us by setting frequency interval,which had more efficient use of frequency resources and was better performance in anti-multipath fading than single-carrier system.However, synchronization technology for OFDM systems was a problem,which was good or bad would affect performance of the system directly.The reason was that OFDM system made high-speed data to distribut a number of parallel and low data rate sub-carrier transmission,which were orthogonal between them and resistance to interference due to multipath transmission.But systematical timing synchronization error would break the sub-carriers' orthogonality,which led to a sharp deterioration in overall system performance.

II. THE ESTABLISHMENT OF SYNCHRONIZATION MODEL AND IEEE802.11A STANDARD

A. IEEE802.11a standard

The standard of IEEE802.11a defined Media Access Control and Physical include IEEE802.11a,IEEE802.11b and IEEE802.11g.International standards urgent to be developed to adapt to the growing interoperability of WLAN equipment.IEEE802.11 WLAN communication standards were promulgated by IEEE in June 1997,that provided all kinds of interface protocol of a variety of mobile clients machine and wireless link information.

IEEE802.11a frequency range was from 5.15 to 5.25 ,from 5.25 to 5.35 and from 5.125 to 5.825GHz.

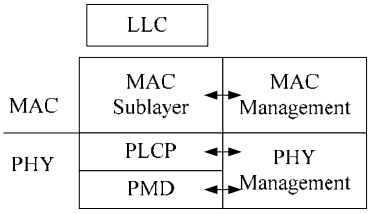


Figure 1. The logical Structure of IEEE802.11a

WLAN provided an effective loading for 6,9,12,18,24,36,48 and 54Mbps data communication services.There were 52 sub-carriers in OFDM system,and the transmitted signals were modulation by BPSk,QPSK,16QAM or 64QAM. IEEE802.11a physical layer consisted of the following three components:

*a) Physical Layer Management:*It connected to medium access control for the physical layer to provide management capabilities.

*b) Physical Layer Convergence Protocol:*MAC connected to PLCP with physical layer service access point by primitives.When the MAC layer gave directions, PLCP began to prepare to transfer media protocol data unit.

*c) Physical medium dependent:*PMD supported through a wireless medium to achieve the physical entities to send and receive between the two stations. In order to achieve this functionality,PMD need directly to access the wireless medium and provide modulation and demodulation.for the frame transmiting.

B. IEEE802.11a System Synchronization Simulation Model

IEEE 802.11a simulation program was composed of many m-files in simulation software matlab6.5 environment.This system mainly included the following sections: transmitter, channel, receiver ,calculation and comparison,that were construction of the physical layer of IEEE802.11a WLAN system model.

1) The transmitter

2) The channel

There were several parts in channel such as generating

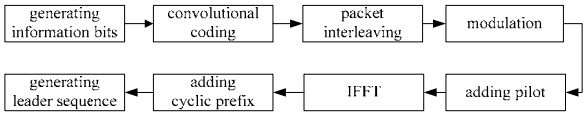


Figure 2. Diagram of the transmitter

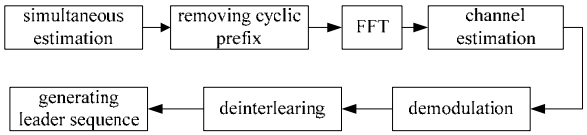


Figure 3. Diagram of the receiver

gaussian and index decline channel, adding the signal to noise ratio that was SNR for the noise,oscillator generating phase noise, joining the channel delay and so on.

3) The receiver

The receiver was made from making use of delaying related packet detection algorithm of the preamble structure, adoptting the two algorithms that the one was proposed by IEEE802.11a system rules and another one was new proposed algorithm to separate frequency synchronization estimation, appling the two algorithms that the one was proposed by IEEE802.11a system rules and another one was new proposed algorithm to separate symbol synchronization estimation, removing cyclic prefix, fast fourier transform algorithm calculation, channel estimation, demodulation, deinterlearing, viterbi decoding and so on.

4) Calculation and comparison

The primary function of this module was generating different results on the basis of different simulation conditions .And the model depend on front result to calculated and compared to BER-SNR with receiving and transmitting signals in different SNR.

III. THE SYNCHRONIZATION ALGORITHM OF IEEE802.11A SYSTEM

A. IEEE802.11a System frequency synchronization algorithm

The most effective way is to use Maximum Likelihood Algorithm to estimate and correct frequency deviation when the receiver got the true useful information under the leading information of data frame.

The reference[3] showed algorithm functions that made use of cyclic prefix for maximum likelihood estimation, as the function of equation (1) and (2).

$$\Lambda_{cp}(\theta)=\left|\sum_{k=\theta}^{\theta+L-1} r^*(k)r(k+N)\right| \qquad (1)$$

$$\varepsilon=\arg\left[\sum_{k=\theta}^{\theta+L-1} r^*(k)r(k+N)\right] \qquad (2)$$

But this algorithm would affect the effectiveness of the cyclic prefix algorithm as the impaction of intersymbol interference would destroy the cycle of cyclic prefix in the multi-path channel . And the scope was limited if cyclic prefix was using for frequency offset estimation.

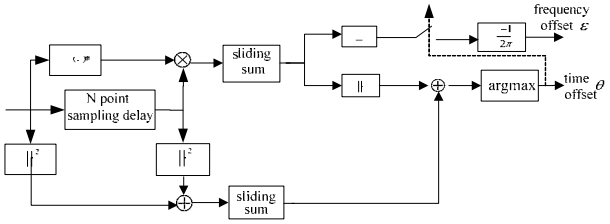


Figure 4. The synchronization Diagram base on CP

In this paper,the author used a maximum likelihood estimator in time domain analysis which had been mentioned in many articles.The only diffrernce is slightly different in form. Training information had to contain at least two repeated symbols in this way,and long and short training symbols met this requirement in the leading information WLAN standards defined.

It was set to be x(n) that sending a signal,and the transmitter carrier frequency was ftx. The received baseband signal r (n) that was carried down-converted by receiver and the effect of noise had been neglected as followed

$$r(n)=x(n)e^{j2\pi f_{tx}nT_s}e^{-j2\pi f_{rx}nT_s}=x(n)e^{j2\pi n\varepsilon/N} \qquad (3)$$

In this function *"N"* was the number of subcarriers,and coefficient of frequency deviation was ε , $\varepsilon=\frac{\Delta F}{\Delta f}=\frac{\Delta F}{f_s/N}$ The symbol *"D"* was defined as a duplicate symbol,so offset frequency estimator was analyzed as followed.

$$z=\sum_{n=0}^{D-1} r(n)r^*(n+D)=\sum_{n=0}^{D-1} x(n)x^*(n+D)e^{j2\pi n/N}e^{-j2\pi(n+D)n/N}$$

$$=e^{-j2\pi\varepsilon D/N}\sum_{n=0}^{D-1}\left|x(n)\right|^2 \qquad (4)$$

From analysis above we can make a conclusion,the offset frequency estimator was

$$\hat{\varepsilon}=-\frac{N}{2\pi D}angle(z) \qquad (5)$$

The function *angle(z)* was defined the interval $[-\pi,\pi]$,so the range of frequency offset was estimated as followed:

$$\hat{\varepsilon}\le\frac{N}{2D} \qquad (6)$$

For the short training symbols,it's length was 16.So the maximum frequency deviation was estimated to be 625KHz,at this time offset was 212KHz. Similarly, for the long training symbols can also be achieved within the estimated.

B. IEEE802.11a System timing synchronization algorithm and simulation

OFDM system timing estimation algorithm can be divided into two steps to complete.First of all, packet inspection for coarse synchronization,it was called packet synchronization estimation algorithm.Furthermore, symbol synchronization for fine synchronization,it was called symbol synchronization estimation algorithm.Simultaneous estimation of two parts can be taken WLAN system leader sequence to complete in IEEE802.11a system.Packet detection was to determine a data packet to reach the exact location which was the basis of frequency synchronization and symbol synchronization after this.Rest of the synchronization process depended on the pros and cons of group testing completion.

Packet detection algorithm performance can be summed up in two generous rate: detection probability

P_D and error alarm probability $P_{FA.}$ The former was detecting indeed a generous packet rate,which standed for the quality of detection desiring to reach. The latter was probability that misjudged group appear but actually it did not appear in probability. Therefore, P_{FA} should be as small as possible. Normally, with the P_{FA} increasing, P_D had also increased; with the P_{FA} reducing, P_D had also reduced.

The most simple algorithm of discovering rising edge of data packet was measuring received signal energy.Set *r (n)* for the receiver to receive the first *"n"* samples of signal *r (t), N(n)* for the receiver to receive the first *"n"* samples' noise, *s(n)* for the sender to send the first *"n"* samples of signal *s (t)* . When it did not receive data packet, the first *"n"* samples of receiving signal *r(n)* was equal to *N (n).*When the data packet arrived, the received signal *r (n)* added to the signal components, which was *r(n)=s(n)+N(n).*

Decision variable was equal to m_n .

$$m_n = \sum_{k=0}^{L-1} r(n+k)r^*(n+k) = \sum_{k=0}^{L-1} \left| r(n+k) \right|^2 \qquad (7)$$

In the function, m_n was used as the energy of sliding window which length was *L* when it received the signal.The energy of *s (n)* was usually higher than that of *N (n).*So when receiving the energy changed can be detected packet.

The structure of short training symbols as follows.

ShortTrainingSymbols$_{-26,26}$=sqrt(13/)×(0,0,l+j,0,0,0, 1-j,0, 0,l+j,0,0,0-1-j,0 0 0,-1-j,0,0,0,l+j,0,0,0,0,0,0,0,-1-j,0,0, 0,-1-j,0,0,0,l+j,0,0,0,l+j,0,0,0,l+j,0,0,0,l+j,0,0)

The structure of long training symbols as follows also.

LongTrainingSymboIs$_{-26,26}$=(1,1,-1,-1,1,1,-1,1,-1,1,1,1,1,1,1,-1,-1,1,-1,l,-1,1,1,1,1,0,1,-1,-1,1,1,-1,1,-1,1, -1, 1,-1,-1,-1,1,1,-1,-1,-1,1,-1,1,1,1,1)

In above diagram the window $C = \sum_{k=0}^{L-1}$ *, window* $P = \sum_{k=0}^{L-1}$ *, L is equal to the length of sliding window.*

Figure 5 showed the signal delay and the related algorithms process.There were two sliding window *C* and *P*.The former called delay-related was cross correlation coefficient of receiving signal and receiving signal delay. Time-delay Z^D was equal to the cycle of leading the start.In IEEE802.11a WLAN systen *D* was equal to 16, which was also the cycle length of short training symbols. The latter calculated receiving signal energy during the window of cross-correlationcoefficient.

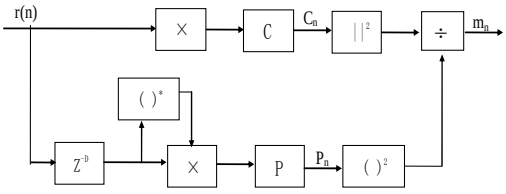


Figure 5. Delay and correlation algorithm signal flow diagram

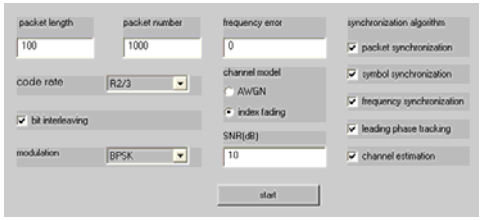


Figure 6. Timing synchronization simulation parameters in index fading channel frequency

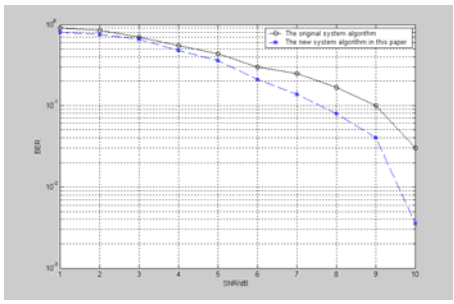


Figure 7. Synchronization simulation results in index fading channel frequency

And the value of this window was judged normalization of statistics.C_n was calculated by the formula (8) ,also P_n was calculated by the formula (9).

$$C_n = \sum_{k=0}^{L-1} r(n+k)r^*(n+k+D) \qquad (8)$$

$$P_n = \sum_{k=0}^{L-1} r(n+k+D)r^*(n+k+D) = \sum_{k=0}^{L-1} \left| r(n+k+D) \right|^2 \qquad (9)$$

Decision variable mn was calculated by the formula (10) .

$$m_n = \frac{|C_n|^2}{(P_n)^2} \qquad (10)$$

In the system simulation parameter was set to channel model for the index fading channel, the number of packet length was 100,the number of group was 1000,convolutional coding rate was R2/3, frequency error was 0kHz, modulation was BPSK and SNR was 10dB.

It can be seen from figure 7 the system bit error rate reflected the performance of the whole system. As the SNR increased, BER gradually reduced. The system's bit error rate had been a marked improvement,compared with the existing synchronization algorithms.According to data from structural characteristics of IEEE802.11a coarse synchronization was carried out by packet detection,fine synchronization was worked out by symbol synchronizing,and simulation estimation was realized by delay correlation algorithm. Experimental results demonstrated that the algorithm above in this paper can effectively overcome the frequency offset and effect of gaussian noise. Thereby,we can get accurate timing information so as to cchieve good communication results.

IV. CONCLUSION

In wirelesscommunication system the research of synchronous technology had been current trends research hotspot.The Problem of synchronization in IEEE802.11a system was researched;timing synchronization and frequency synchronization algorithm IEEE802.11a system proposed was brought up and made a detailed description of achieving algorithm modules in this paper. Simulation results revealed that the algorithm can get good synchronization performance in low noise ratio and index of fading channel,which provided a good idea and simulation platform for taking a step forward study of OFDM system and synchronization in IEEE802.11a WLAN in the future.Also it provided a theoretical analysis basis for the chip design of the IEEE802.11a system.

ACKNOWLEDGMENT

This work was supported by the Planned Science and Technology Project of Henan Province of China (Grant No.052422047) and the third postgraduate degree thesis innovation fund of Henan Polytechnic University.

REFERENCES

[1] Wensheng,Sun;Yuanyuan,Zhang.A frame synchronization and symbol timing synchronization algorithm in burst OFDM communication based on IEEE802.11a. Proceedings-2009 International Forum on Information Technology and Applications, IFITA 2009, v1, p190-193, 2009.

[2] HanYan-chun,Yang shi-zhong.A New Symbol Synchronization Algorithm Based on IEEE 802.11a[J]. TELECOMMUNICATION ENGINEERING,2007,47(2):29-31.

[3] Wensheng, Sun; Yuanyuan, Zhang .A frame synchronization and symbol timing synchronization algorithm in burst OFDM communication based on IEEE802.11a.Proceedings - 2009 International Forum on Information Technology and Applications, IFITA 2009, v 1, p 190-193, 2009.

[4] Toshiya, Shinkai; Nishimura, Haruki; Inamori, Mamiko; Sanada, Yukitoshi.Experimental investigation of fractional sampling in IEEE802.11a WLAN system.2008 11th IEEE Singapore International Conference on Communication Systems, ICCS 2008, p 1368-1373, 2008.

[5] LuoRen-ze.A ew generation of wireless mobile communication system key technology[M].Bei Jing: Beijing University of Posts and Telecommunications Press,2007.

[6] T.M.Schmidl,D.C.CoxLow-Overhead,Low-Complexity Synchronization for OFDM,"IEEE International Conference on communications,Vol.3,1996,p1301-13.

[7] IEEE Std 802.11a-1999 Information technology and telecommunication and information exchange between systems-Part11: wireless LAN MAC and PHY specifications: High speed physical layer in the 5GHz band.

[8] Chen Shi,Zhang Jing-mei.An Improved Symbol Synchronization Algorithm Based on IEEE 802.11a[J].JOURNAL OF WUHAN UNIVERSITY OF TECHNOLOGY(INFORMATION & MANAGEMENT ENGINEERING) .2009,31(3):354-356.

[9] J-J van de Beek,M.Sandell,P.O.Borjesson, " ML Estimation of Time and Frequency Offset in OFDM systems," IEEE Transaction on Signal Processing Vol.45 No.7,July 1997.

[10] Mehta, Mahima; Gupta, Kamlesh.On error performance analysis of sub-channel modulation schemes of IEEE802.11A.IET Seminar Digest, v 2007, n 2, p 903-907, 2007.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 041-043

Research on Optimization Strategy of Relational Schema based on Normalization Theory

Dong Yu-Jie[1] ,Li Fu-Guo[2]
[1]WanFang College of Science and Technology of Henan Polytechnic University,Jiaozuo,China
E-mail: hpudyj@hpu.edu.cn
[2]WanFang College of Science and Technology of Henan Polytechnic University,Jiaozuo,China
E-mail: lfg@hpu.edu.cn

***Abstract*— Optimization of relational schema is critical and difficult for designing a relation database system, the ultimate goal of optimization is to construct a good relational schema for database system. Article giving a systematic and detailed analysis on normalization theory, introduced a good optimization method of relational schema based on normalization theory, proved the feasibility of the method using a real instance.**

Index Term—s**Relational database, Normalization, Schema optimization, Functional dependency**

Ⅰ. INTRODUCTION

Establish a good relational schema is critical for the whole relational database system, because all of the programs and data of the relational database system are based on the relational schema. Relational schema is established in the primary stage for designing database system, it is the core and infrastructure of the whole database management system, all of the operations of the relational database system are executed according to the relational schema requirements. In other words, if the database schema we build out of order, this will affect the whole database system disastrously. The goal of optimization of the database schema is to ensure that the database schema is correct, good and health, a criterion of good database schema is that there does not exist data redundancy, update anomaly, insertion anomaly and deletion anomaly in the relational schema, the main theoretical basis for determining a relational database schema is healthy or not is normalization theory of relational database.

Ⅱ. NORMALIZATION THEORY

Relational schema normalization theory is a set of standards and protocols closely related to functional dependency, this theoretical system contains a total of 6 paradigm (That is Normalization Function, referred to as NF), they are 1NF, 2NF, 3NF, BCNF, 4NF and 5NF, different paradigms corresponding to different schema optimization level, the higher-level paradigm related to the higher-level schema optimization, the following content give the details of the six paradigms.

(1) 1NF: the minimum requirements to relational schema known as 1NF, it demands that all properties items in the relational schema are no longer sub-atomic item.

(2) 2NF: if R ∈ 1NF, and all of the non-primary attributes of relation schema is entirely dependent on the key, then R ∈ 2NF.

(3) 3NF: if R ∈ 2NF, and all of the function dependencies which non-primary attributes depend on the key is not transitive function dependency, then R ∈ 3NF.

(4) BCNF: if each determining factor contains key, then R ∈ BCNF.

(5) 4NF:for relational schema R ,if each non-trivial multivalued dependencies X →→ Y ($Y \not\subset X$), X contains a key, then R ∈ 4NF.

(6) 5NF: if R ∈ 4NF, and does not exist connection dependency in the schema, then R ∈ 5NF.

Ⅲ. OPTIMIZATION METHODS OF RELATIONAL SCHEMA

The basic idea of relational schema optimization is gradually eliminate the inappropriate function dependency in relational schema, to make the various parts of the relational schema achieve a certain degree of "separation." The basic steps are:

(1) Projection decomposition of 1NF relational schema, to eliminate the functional dependency of the original schema which non-primary attributes partly depend on the key, so that 1NF schema will be decomposed into a number of 2NF relational schemas.

(2) Projection decomposition of 2NF relational schema, to eliminate the transitive functional dependency which non-primary attributes depend on the key, so that 2NF schema will be decomposed into a number of 3NF relational schemas.

(3) Projection decomposition of 3NF relational schema, to eliminate the transitive functional dependency and the partial functional dependency which primary attribute depend on the candidate key that not including this primary attribute, so that 3NF schema will be decomposed into a number of BCNF relational schemas.

(4) Projection decomposition of BCNF relational schema, to eliminate the non-trivial and multi-valued functional dependency decomposition in the original relations, so that 3NF schema will be decomposed into a number of BCNF relational schemas..

These are the basic methods of relational schema optimization based on normalization theory, 5NF is the ultimate paradigm, also the best schema that optimization can achieve to the state. But in fact, for the

general relational database systems, it can meet the basic needs as long as the relational schema achieved to 3NF or BCNF, that is eliminate the partial functional dependencies and the transitive functional dependencies in the relational schema. Here's how to use the normalization theory to optimize the relational schema.

A..Methods of elimination of partial functional dependency

If there are relational schema of R (U), U = (X, Y, Z_1, Z_2), X is the key of the relational schema , X_1, X_2 are two subset X, and X ∩ Y ∩ Z_1 ∩ Z_2 = ϕ. There are functional dependencies: X → Y, X_1 → Z_1, X_2 → Z_2. The relational schema R exists attribute set Z_1 and Z_2 that partly depends on the key X: $X \xrightarrow{p} Z_1$, $X \xrightarrow{p} Z_2$.

The decomposition steps are the following:

The first step: See functional dependencies and decompose the schema attributes into two parts, one is entirely depend on the key, the other is partly depend on the key. So the result of the decomposition is: the first part of the attributes (Y) and the second part of the attributes (Z_1, Z_2).

The second step: For the first part of the attributes, directly add into the key to form a new relational schema: R_1 (X, Y).

For the second part of the attributes, according to the different parts of the key on which these attributes fully depend, and then decomposed. Here, attributes Z_1, Z_2 are dependent on the X_1, X_2, and will get the following relations: R_2 (X_1, Z_1), R_3 (X_2, Z_2).

Thus, we get three relational schemas R1, R2 and R3 through schema decomposition, eliminating the partial functional dependencies in the original relational schema R.

B. Methods elimination of the transitive functional dependency

If there are relational schema of R (U), U = (X, Y_1, Z_1, Y_2, Z_2), assume that X is the key of the schema R.Y_1, Y_2 and Z_1,Z_2 are the attribute groups of the schema R, and X ∩ Y1 ∩ Y2 ∩ Z1 ∩ Z2 = ϕ. There are functional dependencies X → Y_1, Y_1 → Z_1, X → Y_2, Y_2 → Z_2 in the relational schema R, so attribute groups Z_1, Z_2 are transitively depends on the key X.

The decomposition steps are the following:

The first step: decompose the schema attributes into two parts, one part is direct functional dependency on key X, the other part is transitive functional dependency on key X. So the first part of the schema attribute is: (Y_1, Y_2), the other is: (Z_1, Z_2).

The second step: For the first part of the attributes, directly add into the key to form the first new relational schema: R_1 (X, Y_1, Y_2).

For the second part of the attributes, according to the different attributes of the relational schema on which these attributes depend, then decomposition and combination. Here, the attributes Z_1, Z_2 are respectively depend on Y_1, Y_2, so get the following relations: R_2 (Y_1, Z_1), R_3 (Y_1, Z_2).

Thus, we get three relational schemas R1, R2 and R3 through schema decomposition, eliminating the transitive functional dependencies in the original relational schema R ,and decomposition is completed.

Ⅳ. EXAMPLE APPLICATION

Assume that relational schema as follows:

SLC (Sno, Sname, Sdept, Mname, Sloc, Cno, Cname, Grade). Sno is the students number, Sname students name, Sdept is department of the students, Mname is the Head of Department, Sloc is the student residence, Cno is the number of course students selected, Cname is the course name, Grade is the result. A student can only learn in a department, a department has only one Head of Department, each student can select a number of courses, a course may also be selected by a number of students, each student has a score for each course, assuming that the students of one department live in one place. The key of the schema SLC is (Sno, Cno).

First analyze and write the functional dependencies of the schema as following:

(Sno,Cno)→Sname, (Sno,Cno)→Sdept, (Sno,Cno)→ Mname, (Sno,Cno) → Sloc,(Sno,Cno) → Cname, (Sno,Cno) → Grade,Sno → Sname,Sno → Sdept,Sno → Sloc,Sno→Mname, Sdept→Sloc, Sdept→Mname, Cno →Cname.

After analysis, we know that this relational schema exists following four exception issues:

(1)Insert exception: can't insert student information that no elective courses;

(2) Delete Exception: when a students elective records were cleared, at same time the basic student information are deleted;

(3)Huge data redundancy: if a student select 10 courses, the value of Sdept and Sloc of the student must be storage 10 times;

(4) Modify complex: if a student turn from Mathematics (MA) to Information Department (IS), originally only need modify the value of Sdept in the student tuple, but here also must modify corresponding residence (Sloc).

Reason of the four exception questions is that there exist non-primary attributes(Sname, Sdept, Mname, Sloc and Cname) partly depend on the key (Sno, Cno).In the following, we will decompose the schema SLC according to the normalization theory and the above schema optimization methods to make the schema SLC achieve to BCNF.

First, each attribute of the schema SLC are the basic data items that can not be separated, so SLC ∈ 1NF;

Second, eliminate the partial functional dependencies, using the method described in 3.1 of this thesis.

(1) See functional dependencies and decompose the schema attributes into two parts, one is entirely depend on the key, the other is partly depend on the key.

The first part of the attributes is: (Grade)

The second part of the attributes is: (Sname, Sdept, Mname, Cname, sloc)

(2) For the first part of the attributes, directly add into the key and form a new relational schema: SC(Sno,Cno,Grade).

For the second part of the attributes, according to the different parts of the key on which these attributes fully depend , and then decomposed. After decomposition to the second part of the attributes, we will get two schemas: L(Sno,Sname,Sdept,Mname,Sloc), Course(Cno,Cname).

Now, all of the three new schemas are belong to 2NF because there no any non-primary attributes partly depend on the key. And, according to the definition of paradigm, relational schema SC and Course have no any attributes partly or transitively depend on the key ,so they have already belong to BCNF.

But for relation SL, abnormal problem is still existed, such as can not insert the information of a Department that have no students, much redundancy (such as information of a department only need stored just once in normal state, but in SL relation the department information must be stored as much times as the numbers of the students of this department). Following, we first write the functional dependencies of relation SL:Sno→ Sname, Sno→Sdept, Sno→Mname, Sno→Sloc, Sdept→ Sloc, Sdept→Mname.

In relation SL exists non-primary attributes Sloc and Mname transitively depend on the key Sno, decompose the schema using methods described in 3.2 of this thesis:

(1) Decompose the schema attributes into two parts, one part is directly depend on key, the other part is transitively depend on key. the first part of the attributes is (Sname,Sdept), the other is (Mname,Sloc).

(2)Add Sno into first part of the attributes and formed the first relation SD(Sno,Sname,Sdept)， add Sdept into the second part of the attributes and formed the second relation DML(Sdept,Mname,Sloc).

Now, the new relations SD and DML dose not exist the non-primary attributes transitively depend on the key, so they have already belong to 3NF.In fact, relational schema SD and DML have already no any attributes partly or transitive depend on the key , according to the definition of paradigm they have already belong to BCNF.

Ⅴ. CONCLUSION

This article firstly given a detailed analysis and research of normalization theory, then puts forward a relational schema optimization method based on standardized theory, this method proved to be simple and feasible. Realized the relational schema optimization through eliminating the part and the transfer functional dependency in the relational schema.

REFERENCES

[1] WANG Shan, SA Shi-xuan, Database System (Fourth Edition) [M], Beijing, Higher Education Press, Sept 2006.

[2] WANG Shan, CHEN Hong, Database System Principles (second edition) [M], Higher Education Press, July 2004.

[3] MA Yuan, Relational database theory [M], Tsinghua University Press, April 2007.

[4] CHEN Hu, Database technology [M], Beijing, Northern Jiao Tong University Press, May 2002.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 044-047

Application of Simple concept of multi-layer protection in the Security of College Campus Net

Wei Liu[1,2,*], Xianglin Wu [1]
[1] Huazhong University of Science & Technology, Wuhan, China
Email: zslw@163.com
[2] Zhoukou Normal University, Zhoukou, China
Email: xlwu@public.wh.hb.cn

***Abstract*—The concept of simple multi-layer safety protection, which fully dig for the ability of network facilities and multiple utilization of different network security technology, and specific protection plan put forward based on analysis of security of college campus net.**

***Index Terms*—simple multi-layer protection, network, security, college campus net**

I. INTRODUCTION

Along with the rapid development of computer networks, global information has become the main stream of human development, information has become the third of resource with the material and the ability to maintain human society, but the networks can be easily attacked by hacker, malicious software and other illegal acts, many new networks intrusion and attack methods happened because of the network openness, the diversity of forming connection, The uneven dispersion of terminal and Security Vulnerabilities of network communication protocol. So the networks security has also become the most important problem in network construction, and how to ensure security of networks system and build the solid security system, already become the important tasks and duties of designer and attendant of network system[1]. Therefore College campus networks also need to deal with network security as a part of Internet.

II. ANALYSIS OF COLLEGE CAMPUS NETWORKS PROBLEMS

First College campus networks is the Intranet, which could be provided flexible, efficient, relax, fast, cheap and reliable ideal inner environment of information exchanges and sharing, college management, which through several network accessing methods to Internet and share its rich information sources, fully display the whole image of university and offer more service to society. The College campus networks are usually adopted mature or advanced technologies to build the high-bandwidth networks, which provided very favorable conditions for network attack and virus propagation. The main features of its network include the following aspects [2]:

- Complex on structure level and difficult on management level. From the perspective of structure, campus network can be generally divided into three levels: the core, the convergence and the access. From the perspective of subnet type it can be divided into teaching subnet, office subnet, and dormitory subnet. In some universities, broadband access, wireless access, dial-up access and other various forms of access are all available. In addition, the campus network is more often than not with multiple linking ways, except for CERNET, netizens can also link to the internet through China Mobile, China Telecom, China Unicom. This multi-level linking feature makes the campus network faces the problem of more serious management and network security when enjoying the many conveniences brought by high technology and makes network security especially important.
- Diversity of user types and active netizen groups. University network faces a variety of user groups, including students in the dormitory subnet, teachers in the teaching subnet and office personnel in commercial organizations subnet. Thus causes the difficulty in management. Meanwhile, there are a substantial number of students with high computer-related skills, active mind but comparatively weak awareness of law. When trying to use various technologies concerning network, these students are likely to influence and even undermine network security. In this case, university network is expected to provide rich internet resources to meet the demand of teaching and learning, while at the same time its also expected to ensure that the campus network is operated safely, which became a problem relatively difficult to handle.
- Multitudinous application systems with complex functions. Application is the core of establishment of campus network. The network campus should fulfill enormous demands in process of teaching and management which involves learning activities, teaching and research activities, educational administration management, information interchange, etc. Consequently, campus network should set up a great number of application systems, e.g. Web, Ftp, Mail, QA, Network teaching system, etc. While the campus network is provided with abundant functions by means of numerous application systems, there exist innumerable

*Corresponding author
Science and Technology Development of department of science and Technology in Henan Province（102102210265）

potential safety hazards in the web. In other words, campus network has become the target of many attackers. Thus, higher requirements have been raised on web safety administration.

- Inadequate investment for web safety with a lack of administration. The beginning period of web establishment for colleges and universities saw a general tendency that attention had been paid more to technology than safety and administration owing to ideological and funds problems with preference for technology and inadequate operation awareness. To put it another way, the colleges and universities campus network establishing workers usually do not focus on safety, and make do with putting a firewall between intranet and the Internet. The workers for some colleges and universities even avoid doing so, which opens an easy access to viruses and hackers. Meanwhile, most colleges and universities suffer from insufficient workers in web safety administration. One worker generally should take responsibility for web establishment, management and safety, etc. Thus, it is extremely hard to guarantee safe web operation

Based on the above web traits, the major problems in colleges and universities web safety are summarized in the following [3]:

- Fragility of user computer. With the broadening discipline of campus network and more and more nodes to campus network, computer users are diversified in computer level. In addition, the vulnerable nodes of user computers can have some safety holes in their operating systems. What's worse, these operating systems become key targets for viruses and Trojans because of absence of safeguard procedures.
- Transmission of viruses, Trojans and hack tools. Such a problem is caused by consciously or unconsciously insecure visit of internal customers to outer net, e.g. browsing the website of the horse, receiving virus-carrying emails, transmitting viruses by means of chatting tools like QQ and hack tools spreading among the curious students.
- Explosion of net viruses. Owing to advanced bandwidth and sharing characteristics of campus network, infection of one node in intranet becomes source of viruses, and contaminates other user computers in broadcast, multicast and unicast way, resulting in breakdown of intranet. For instance, security incidents, like DOS/DDOS attack and explosion of ARP virus, emerge one after another.

III. NETWORK SECURITY ARCHITECTURE

A. *PDRR model of the network security*

In order to protect your data and network resources and ensure the network availability, confidentiality, integrity, authenticity, non-repudiation, the most common style is the PDRR dynamic network security system model [4].

PDRR model is the combination of the first character of four English words: Protection, Detection, Response, Recovery. These four words constitute a dynamic cycle of information security, as shown in Figure 1. Each part of the security policy includes a group of appropriate safety measures to implement certain security features. The first part of the security policy is prevention, which makes preventive measures according to all known security system, such as the patch, access control, data encryption and so on. Prevention is the first front of security policy and the second one is detection. The detection system can detect the attacker which invades the defense system. This function of security front is designed to test out the identity of invaders, including the attack of resources, system loss, etc. Once being detected, the response system starts responding, include event handling, and other businesses. The last line of the security policy is system restoration. The system can be restored to its original state after the invasion. Every time the invasion occurs, the defense system will be upgraded, which ensures the same type of invasion cannot happen again. Thus the entire security strategies include prevention, detection, response and recovery, four of which compose of an information security cycle.

Figure 2. The secure architecture model of the PDRR network

PDRR model is demonstrated as the existence of the ultimate form of network security, that is to say, a kind of target system and model. It does not concern the engineering process of network security construction and failed to explain the ways and means to achieve the target system. In addition, the model is more focused on technology rather than emphasizes the factors, such as the management. Network security system should be the security architecture fused technology with management, which could solve security issues fully. It should possess the following characteristics: dynamic, progressive, inclusiveness, stratification and balance and so on. It is a true blueprint on which information security activities based.

B. *Computer Network Security Architecture*

Because of the particularity of network security and urgency, IS0 TC97 develops network security architecture NSA (Network Security Architecture) of IS0 7498-2, according to the other models. Based on this standard, a wide range of security systems are designed to meet what different network environment needs for

security codes. In ISO 7498-2, it describes the architecture of open systems interconnection security, and brings forward five categories of security services in the basic structure of information systems security design and eight types of security mechanisms to support these five services [4]. Its security framework structure can be shown in Figure 2.

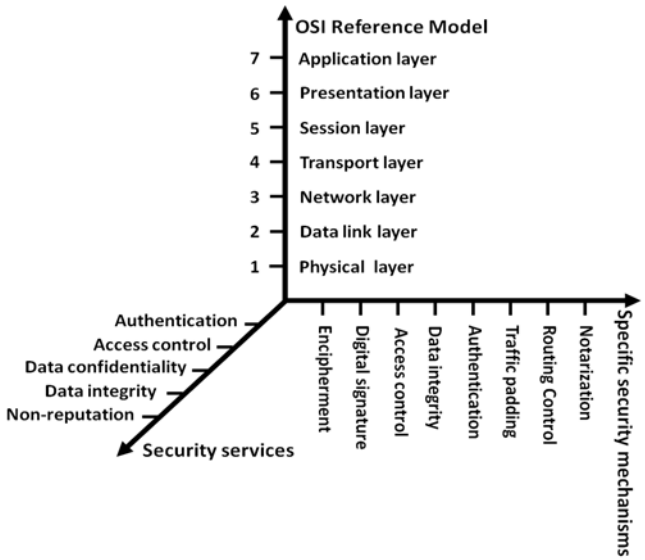


Figure 1. The 3D diagram of framed structure of ISO 7498-2's network security

In addition, this content will be mapped to the 7-layer models of OSI by ISO 7498-2. It should be noted that security services and security is not a one-to-one, and one security service may take one or more security mechanism to perform. In the OSI network model, a security service is provided by a particular layer, that is to say, the security services have the appropriate security mechanisms to support. The relationship between security services and network stratification can refer to table I.

Table I The mapping of security service and OSI network protocol

Security service Network protocol	Authe-ntication	Access control	Data confident-iality	Data-integrity	Non-reputation
Physical layer	N	N	Y	N	N
Data link layer	N	N	Y	N	N
Network layer	Y	Y	Y	Y	N
Transport layer	Y	Y	Y	Y	N
Session layer	N	N	N	N	N
Presentation layer	Y	Y	Y	N	Y
Application layer	Y	Y	Y	Y	Y

IV. CONCEPT AND APPLICATION OF SIMPLE MULTILAYERED PROTECTION

A. *concept of simple multilayered protection*

Model PDRR and IS0 7498-2 NSA depicts a blueprint which can be based on in information security practice. However, in college campus networks, because the level of the importance of data is not high, the shortage of funds, the specialty of the network users, PDRR and part of the NSA is not or needless to realize. For example, the core data (such as financial information, educational information, etc.) need rapid response and recovery when they are attacked, but for the common data, when attacked, the recovery time can be prolonged; we can cut off the network connection once the computers in the protection zone infect the virus, but many common users (such as computers in the computer rooms and dormitories) may continue online after the infection. If we cannot purchase the network security products as the firewall, anti-virus, IDS, encrypted system etc., it will be necessary to fully exploit the potential of existing network equipment, optimize the network to develop a simple multilayered protection.

The simple multilayered protection refers to a protection whose uppermost aim is to promote the network efficiency, fully make use of the existing network equipment, offer a multilayered protection against the threats impacting the campus network security and network operation efficiency, and each level of the protection does not affect the network speed, but ensures the data transmit rapidly.

B. *application of simple multilayered protection*

1) According to the principles of simple multilayered protection, it can offer the following protection

a) In order to protect the server, it is placed behind a firewall. We establish strategies only allowing certain port to be visited. Since there is only one line connecting the firewall and the core equipment, if it protects too many servers, then the firewall will be a bottleneck confining the data transmission. As an alternative, we can set the server directly connected to the core equipment, meanwhile, reasonably set the server's own IP security policies, close the unnecessary port to ensure high-speed service.

b) Reasonably set the server account policies, such as confine the IIS anonymous user IUSR_Computername on the Web server, we can partly control the attack against the port 80; control of upload permissions to decrease the threats of the users' uploading Trojans.

c) Reasonably plan the server's IP address. Those are just visited internally can be set as private IP addresses, which the users outside the campus can't visit; the data servers which allow a single server to visit also can be set as private IP addresses.

d) According to the importance of the data, requirements of the recovery speed, we can grade the methods of backup of different servers: the first-level data for hot backup, the second-level data for incremental backup, the third-level data for sub-period backup.

Of course, the important servers may still need a firewall, IDS, and other equipments to get protected, but the principles of simple multilayered protection aims to adopt different levels of protection according to different servers to ensure the high efficiency of the data services.

2) Under the principle of multi-layer simple protection, the following can be done to ensure the security of network in the structure and equipment of campus network

a) Try best to make the three alternative network structure of "core layer--convergence layer--access layer" replace the two network structure of "core layer--access layer".

In the secondary network structure, "core layer--access layer", the three-forward is fully completed by the core device, connected to the low-end switches in the network edge via a fiber-optic. The access switches are 10/100 M to the desktop. As the core network equipment directly exposed to the secondary network, the device must support large capacity of host and subnet routing table. Such a large secondary network is actually a broadcast domain, which makes most of the equipment assume lot of unnecessary flow, and the core switch is overloaded. Although the appliance of the technology, such as, VLAN isolation, ACL and so on, can limit certain amount of broadcasting, the cost and risk of the network management is increasing because of the multiple program allocations and the complicated allocations. As is known to all, the majority of viruses, Trojans, in a broadcast domain spread quickly. Usually with the virus, the computer in the entire VLAN is easy to be infected, making the efficiency of the core equipment decreased, causing the whole network speeds down. If the three-network structure, "core layer--convergence layer--access layer", is applied, it will transmit the virus to be controlled by VLAN in the convergence layer, simultaneously making the decrease of the network efficiency below the convergence layer. Now, several network companies put forward to make the three-layer access to the network model. It makes the function of the three layers down shift to the access layer, and the transmit of the three layer is efficiently shared by the access layer to reduce the stress of the core equipment. At the same time, the down shift of the broadcast domain forms a protection for the convergence layer and core layer. The stability and robustness of the network in enhanced.

b) Rational use for the ACL function of switch for virus isolation

Access Control List is a list of instructions for router and switch interface to control the data pack from the If Index. ACL is applied to all routing protocols, such as IP, IPX, Apple Table, etc. The list contains the matching, condition and SQL. Through ACL, the non-authorized users can be ensured to access specific network resources, thus to achieve the controlling purpose for the accession.

At present, the switch usually supports the function of ACL. ACL has been applied in many network administrators to control the commonly used ports by virus. However, with the increased virus, network administrators gathered more and more the controlling ports, which make the ACL greater. If the large ACL is applied to every switch interface, it is bound to affect the network speed.

In response to the situation, according to the simple multi- layer protection principle, ACL should be deployed. First, sorting the collected ports, put the most common and most damaging virus on the top and not common and small ones on the below, which forms the ACL sequence, as ACL1, ACL2…ACLn, and make them access the switch from ACL1 to ACLi, ACLi+1…ACLj to the convergence switch, ACLj+1…ACLn to the core switch. So the switch actually applies the ACl controlled by i ports, the data interface into the access switch is "i+1" (suppose the last one is a permit statement), "j-i+1" for convergence switch and "n-j+1" for convergence switch about the controlling ports. According to different processing power, we should adjust the value of i and j to ensure the data through the interface of switch rapidly. For the experience, the general values of i and j can not be too large, such as $i<10$, $j<40$, so we can ensure the access and convergence switch not be cost much by ACL.

After the ACL application, we should always observe the state of each ACL Strategy "hit" in the switch. According to "hit" on the ACL to sort in time, adjust the ACL of each layer. We should abandon the low "hit", while also collect the new port to the ACL.

c) Switch use of new technologies, new features for viruses and attack defense.

New technology will be added in the switch by the network equipment manufacturers, to provide some new ideas for security. Such as the formerly mentioned "three-layer access to the network model", which can effectively control the virus. Also the simple configuration command, can efficiently avoid the attack of ARP virus and other special ACL supported by switch can filter some illegal application. These new technologies and functions need the timely follow-up of the network administrators to apply them to the general network security system.

V. CONCLUSION

With the continuous development of network technology, more and more emphases are put on college campus network security. Any networks security technology can't be provided all the security services for a complex network system, and fend off any attacks. So It is inadvisable to pursuit high investment to ensure the security of network security, the simple multi-layer safety protection, which fully dig for the ability of network facilities and multiple utilization of different network security technology, is a "less input, high-income" security protection technology and fairly effective.

REFERENCES

[1] Wangqiuhua Zhangjianwu Luoyi, "The Design and Implementation of Network Security Architecture," Journal of Hangzhou Dianzi University,vol.25,pp.41-44,oct 2005.

[2] Mawenjie Wangyan, "On the Construction of Security System of campus network," Fujian Computer,vol.1,pp.81-82,2007

[3] Zhengchunxiang Dongjiadong, "the research on Security System of College CampusNetwork," China Education Info,pp.32,jun 2006

[4] Hanxing Chenying, "Architecture & Technology for Security of Computer NetWork," Development & Innovation of Machinery & Electrical Products,vol.19,pp.84-86,sep 2006

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 048-050

Application of Dijkstra Algorithm in Logistics Distribution Lines

Liu Xiao-Yan, Chen Yan-Li
School of computer science and technology, Henan Polytechnic University,JiaoZuo, China
Email: xyanliu@hpu.edu.cn，Email:yanlichen@hpu.edu.cn

***Abstract*—Use heap sort to sort unlabeled nodes in geography network to improve the efficiency of Dijkstra algorithm. Provide separate solutions of path optimization based on Dijkstra algorithm in logistics distribution lines with barriers and no barriers. Propose modified Dijkstra algorithm given vehicle, weather and other factors.**

***Index Terms*—dijkstra algorithm, the shortest path, vehicle routing, logistics distribution**

I. INTRODUCTION

With the rapid development of e-commerce, the improvement and optimization of logistics distribution system has become a research hotspot of many enterprises, experts and scholars. Vehicle routing problem (VRP), as an important component in the optimization of logistics distribution system, has always been one of the most active topics in the field of operations research and combinatorial optimization. With the path optimization, efficiency of transport vehicles can be improved, which will save a lot of manpower and material resources. Currently, viewing from the point of the operation of enterprises in our country, transportation costs account for 30% of the total cost of the national economy, only 10% in developed countries. That is, only from the point of transportation costs, we have "20%" of such a space to develop. As long as we can reduce our transport costs by about 10% of existing cost, a new leap forward of our national economy will come true. Research on route optimization algorithm of logistics distribution can give positive and effective advices and solutions in cost reduction, reduce logistic cost, so as to promote the rapid development of the national economy [1].

As the result of the analogy between the graphs of geographic network and graph theory, graph theory has been widely applied in logistics. Shortest path algorithm in graph theory is considered as a based effective algorithm to solute vehicle routing [2]. Logistics route (whatever by sea, land or air) is similar to the edge of graph, and a series of loading or transporting locations is similar to vertex. The weight of edge can express the distance between two locations、the journey time、traffic expenses and so on [3]. Although shortest path algorithm (this paper we use Dijkstra algorithm) is a traditional and antiquated method, until now it is a hot issue of the optimal path research, because improved and extended algorithms emerge in endlessly. Now there are about 17 kinds of the shortest path algorithm proposed based on graph theory, of which three kinds are tested better by experts, these are TQQ、DKA and DKD. The latter two algorithms are based on Dijkstra algorithm, and the differences between the two different systems just lie on the different implementations of algorithms [4]. In this paper extended Dijkstra algorithm is used to try to solve the hot issues in Logistics route.

II. APPLICATION OF SHORTEST PATH ALGORITHM IN DISTRIBUTION LINES WITHOUT BARRIERS

A. Question

Given starting point and destination, and no barriers at each location in the road, how to obtain the optimal path of the distribution lines? Currently, basic problem that distribution center encounters is how to seek optimal routes, and the core algorithm of which is the shortest path algorithm. For the convenience to solve the problem, we establish a geographic network model that suits for logistics routes planning. The shortest path algorithm based on graph theory has approximately more than ten kinds, and three of which are tested better by experts, they are TQQ、DKA and DKD. TQQ is graph growth theory, and the latter two are based on Dijkstra algorithm [5]. Dijkstra algorithm is a graph search algorithm that solves the single-source shortest path problem for a graph with nonnegative edge path costs, producing a shortest path tree. This algorithm is often used in routing. Most current systems use Dijkstra algorithm as the basic theory to solve the shortest path issue. Different system has different means to realize Dijkstra algorithm. It is generally used in computing the minimum cost path from the source node to all other node. Here the authors use Dijkstra algorithm as the basis of selecting analysis algorithm of logistics routes.

B. Description of Dijkstra algorithm

The basic idea of Dijkstra algorithm is to explore the shortest path from source point (labeled as s) to outside gradually. In execution process assign a number to each point (called the label of this point), which expresses the weight of the shortest path from s to this point(named as P label) or upper bound of the weight of the shortest path from s to this point(named as T label). In each step, modify T label, and alter the point with T label to point with P label, so that the number of vertex with P label in graph G increases one, then we can obtain the shortest path from s to each point only by n-1 steps (n is the number of vertexes in graph G). In order to optimize the algorithm, here we express Dijkstra algorithm in another

way. Suppose each point has a pair of label (d_j, p_j), d_j is the length of the shortest path from the starting point s to the end point j, and p_j is the front point of j in the shortest path from s to j. The basic process of solving the shortest path algorithm from the starting point s to point j is described as follows:

(1) Initialization. Set the starting point as: ①d_s=0, p_s is null; ②all other points: d_i=∞, p_i=?; ③mark the starting point s as k=s and all other points as unlabeled.

(2) Examine the distance between the marked point k and unlabeled point j that is directly connected to k. Set d_j=min[d_j , d_k+l_{kj}], l_{kj} is direct connection distance between k and j.

(3) Choose the next point. Choose the smallest i in d_i from all unlabeled points: if d_i=min[d_j, all unlabeled point j], then i is selected as one point of the shortest path and set as marked.

(4) Find the front point of i. Find j connected directly to i from marked points, make it as front point and set p_i=j.

(5) Mark i. If the target point has been marked or all points have been marked, then the algorithm is finished, otherwise set k=i and turn back to step (2) to continue.

As can be seen from above, in the process of achieving Dijkstra algorithm, the core step is to choose an arc with the shortest weight from unlabeled points. This is a cyclic comparing process. If the unlabeled points are stored in a linked list or array in unordered form, we have to scan all the points to choose an arc with the shortest weight. It will affect computing speed in the case of large amount of data.

C. *Improved Dijkstra algorithm*

Combined with actual situation, in this paper heap sort is used to order the unlabeled points to improve the efficiency and the shortest path of the node as the key word of heap sort. The reason is shown as follows:

(1) Make full use of existing heap data, and greatly reduce the frequency of data comparison. The run-time of heap sort is mainly consumed in constructing the initial heap. For Dijkstra algorithm, we only have to construct one heap in the whole process of seeking the shortest path with n times of heap sort. This can obviously overcome the main drawback of heap sort and show its advantage clearly.

(2) Changes of the shortest path value in Dijkstra algorithm are always smaller than original value. So we use small root heap, namely, the heap root is the node with the smallest keyword value. While refreshing the operation of the heap after modify key word of some node (the shortest path value), we only have to determine whether the node is needed to adjust the position to its parent node. Thus the rearrange operations of heap in Dijkstra algorithm are more simple and efficient than that of traditional heap.

(3) Heap sort requires only an auxiliary space of one record, quick sort requires O(logn) and merge sort requires O(n).

III. OPTIMIZATION OF PATH IN DISTRIBUTION LINES WITH BARRIERS

The above optimal solution is calculated based on the known conditions. Its optimality will be changed when conditions vary. The method to obtain optimal path in changing conditions is designed as follows:

A. *Question*

Suppose logistics distribution center need to deliver goods from initial location O to the destination D. Ideally, logistics distribution center can regard the whole urban traffic network as a plane graph to calculate the shortest path between O and D, so as to save transport costs and time. Actually, there is often such a problem: while a transport truck get to location A in the shortest path and the driver find it can't pass because of road breakdown in the front location B, they must change the route and also ensure the selected path be optimal.

B. *Mathematical model*

Suppose urban traffic network is a plane graph marked as G. Each location in traffic route corresponds to each vertex in graph G. Set G= (V, E) as undirected graph with weighted edge, where V is the set of vertices and E is the set of edges. In plane graph, as we know, the sum of the length of two edges is greater than that of the third edge in a triangle. That is called triangle inequality. Here O represents the starting point of the transport, D represents the destination, SP represents the shortest path without barriers and E(SP) represents transport costs spent on the shortest path. Discussions in the following are all based on two assumptions:

①Graph G is still connected after remove choke point.

②It can be found that the latter point can't be get through because of blockage only when vehicle gets to the previous point.

C. *Algorithm design*

Dijkstra algorithm uses the data structure with two sets to deposit vertices of graph. Set s represents the set of marked nodes, and set (G-S) represents the set of unlabeled nodes. The label of a node shows the shortest distance between this point and the source point. The main idea of this algorithm is: select node W with the smallest label from set G-S, then put node W in the set S. Adjust all the path value of node v (T[v]) which pass through the node W and is connected to the source node in set G-S. If the path value of node v is bigger than the sum of the path value of labeled node W and the distance between v and W, then adjust all the path value of the nodes which are connected to node v. Repeat this process until all the points are entered into set S.

From the above analysis we can see, all the nodes will be labeled after operating Dijkstra algorithm. First, we give symbols and definitions that will be used in this paper. Use WOA to represent the shortest distance between node O and node A, W(SPAD) represent the distance between node A and node D along the optimal path SPAD, W_{ij} represent edge weight from node i to

node j, and T[v] represent the label value of node v. The algorithm is as follows:

①For the given plane graph G, use Dijkstra algorithm to get the optimal path SP from the source node O to the end D.

②When vehicle gets to node A and finds it impossible to get through B because of barriers, compute sub graph G-{B}.

③Reuse Dijkstra algorithm to get the optimal path SPAD from node A to node D.

④The cost from node O to node D is the sum of cost WOA(from node O to node A) and cost W(SPAD)(from node A to node D).

IV. Application of a Case

In accordance with the requirements of customers, a logistics company transports a batch of fresh fruit from A city to B city. To maintain fresh and value, the fruit should be delivered to B city with the shortest time and optimal path. Based on Dijkstra algorithm, we can easily obtain the most convenient and efficient route SP with the cost of W(SP). Staffs start from A city, go along route SP, and get in trouble in intermediate V city because of road congestion caused by floods. In such case, staffs have two choices, one of which is staying put for the bridge repaired and the other is finding another way to go on. This is a comprehensive and complex problem. Sometimes staying put is more appropriate because staffs prefer to walk on according to the optimal path calculated in advance thus distribution team don't have other expenses to choose another path. But its adverse factor is time cost consumption. The value of fresh fruits is their fresh feature, so we must transport them to the destination at the shortest time. If staffs stay put, the value of goods will be lost. Thus staffs reselect path to move on. Now, problems staffs face is how to reselect path to ensure the timely arrival. The road graph is as follows:

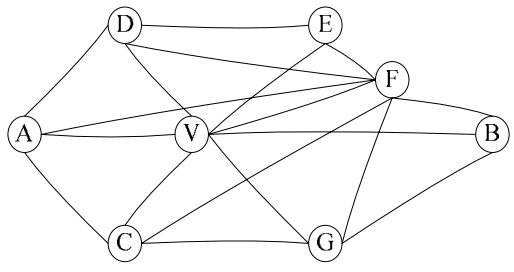


Figure1. Road graph

In this case, we assume that the time from start to finish is proportional to the cost from start to finish. So, we can calculate the optimal path SPAVB from V city to the end city B basing on above algorithm. Namely, at the point V, use Dijkstra algorithm to obtain the optimal path SPVB from V city to B city. Here, WAV+W(SPVB) is the shortest time.

In the era of diversified transportation, logistics company can't control the cost better if simply consider the distance between the two cities. In distribution process, logistics company and staffs need to consider not only the route but also other possible factors, such as road condition, weather, holiday and so on. Reconsider this case. Staffs start from A city, go along SP and get in trouble when get to intermediate city V because floods make bridge break. Staffs need to consider not only the proximity of the road but also the road condition and so on when staffs reselect route. At the same time, staffs can choose railway, waterway to complete distribution task with the lowest cost and the fastest speed.

We follow a more practical significance of this case to modify the algorithm. The factors original algorithm considers is too little, so the solution obtained may be not the optimal. In the case of more selectable variables, it is possible to obtain a better solution. We amend the algorithm designed according to the shortest distance as follows: fully consider convenient traffic conditions, road conditions, weather, distance, transport costs and so on. In the new case, change the weight of the path into comprehensive weight. Namely, W_{ij} shows not simply the distance between i and j but an integrated factor. $W_{ij}=f(a_{ij}, b_{ij}, c_{ij})$, a_{ij} expresses distance between i and j, b_{ij} expresses weather between i and j, c_{ij} expresses vehicle between i and j.

For the case in this paper, staffs need to consider vehicle, weather, distance, cost and other factors when they come across floods in V city and have to reselect other route. When reselect the desired mode of transport, the crucial problem is how to balance speed and cost of transport services. With the full use of road transport which is fast and flexible, carry out short distance railway, waterway and road diversion to increase the transport capacity of the blocked point and achieve more comprehensive and efficient transport means.

References

[1] LI Dong-long, LI Ren-wang,LI Yao-hui, and ZHANG Peng-ju. Improved spanning tree-based genetic algorithm and its application in cost optimization of logistics dispatching system[J]. Mathematics in Practice and Theory. 2009,(21).

[2] HE Xiao-nian, and XIE Xiao-liang. Capacitated logistic distribution vehicle routing optimization[J]. Computer Engineering and Applications. 2009,(34).

[3] LE Yang, and GONG Jian-ya. An efficient implementation of shortest path algorithm based on Dijkstra algorithm[J]. Geomatics and Information Science of Wuhan University. 1999,(9).

[4] HAN Hao, WANG Suling. Modeling and solving for locations of multi-level logistics network[J]. Journal of Shanghai Maritime University. 2009,(4).

[5] WANG Hui, REN Chuan-xiang, YING Chang-chang, and HAO Xin-ga. Study on optimization of logistics distribution route based on niche genetic algorithm. Journal of Computer Applications. 2009,(10).

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15, August 2010, pp. 051-054

Influences of Perceived Risk and System Usability on the Adoption of Mobile Banking Service

Zhihong Li[1], Xue Bai[2]
[1] School of Business Administration, South China University of Technology, Guangzhou, China
Email: bmzhhli@scut.edu.cn
[2] School of Business Administration, South China University of Technology, Guangzhou, China
Email: bai-xve@foxmail.com

***Abstract*—In mobile situations, consumers' perceived risk is the important determinant of whether they adopt the mobile banking service. Meanwhile the root cause of consumers' perceived risk is the existence of system usability of mobile banking service. After reviewing relevant literature, this text pointed out the system usability of mobile banking service is the important factor of consumers' perceived risk, introduced perceived risk into the study of adoption behavior of mobile banking service, and proposed a research model of the adoption willingness of mobile banking service based on perceived risk and system usability. This model framework can help scholars to understand the perceived risk and usability issues of mobile banking service can also help mobile banking provider to improve the product and service quality of mobile banking service, further promote consumers' acceptance and adoption willingness of mobile banking service.**

***Index Terms*—mobile banking service; perceived risk; system usability; adoption**

I. INTRODUCTION

As the development of communication technology in recent years, mobile banking service, as a typical application of mobile commerce, is developing rapidly. Mobile banking service takes advantage of mobile communication technology and equipment to provide various banking and financial services such as account management, remittance transfers, payment services, mobile stock market, foreign exchange operations, etc.[1].

Mobile banking service can provide consumers a variety of banking services anywhere anytime, but at present this service is not adopted by consumers broadly. According to iResearch's lasted investigation, only 14.3% of mobile phone users are using mobile banking service [2]. This adoption rate is much lower than other mobile value-added services. Some scholars attributed the reasons why consumers were not willing to use mobile banking service to some obvious obstacles, such as safety problem, privacy concerns, etc. Most studies on the adoption factors of mobile banking service are based on Theory of Reasoned Action , Theory of Planned Behavior (TPA) or Technology Acceptance model . TAM is the most widely used among them. But, in the framework of consumer behavior, the discussions of the adoption willingness of mobile banking service are rarely seen, and the potential impact of these obstacles on adoption willingness of mobile banking service is unclear.

Perceived risk, or consumers' subjective expectations of the possible loss, provides a convincing analysis framework which can be explained consumers' adoption behavior of mobile banking service. Thus, perceived risk is the important determinant of consumers' adoption willingness, the research on perceived risk of mobile banking service is conductive to understand the determinants of consumers' adoption willingness, gain a more profound grasp of the nature of consumers' behavior. It is noteworthy that in mobile situations mobile banking service is highly dependent on system usability. System usability not only impacts the quality of mobile banking service, but also gives rise to the perceived risk of consumers, further affects the adoption willingness of mobile banking service.

Thus, this paper researched the perceived risk of mobile banking service adoption behavior and its root cause systematically and deeply from the perspective of perceived risk and system usability, proposed a research model of mobile banking service adoption willingness based on perceived risk and system, in order to attract more attentions of academic community on perceived risk and system usability of mobile banking service adoption willingness, further to understand the adoption factors of mobile banking service more comprehensively.

II. LITERATURE REVIEW

A. Perceived risk and its facets

In year 1960, Harvard scholar Bauer introduced the concept of "perceived risk" from psychology into the study of consumer behavior. Bauer professed that all behavior of consumer could result in uncertain consequences which cannot be foreseen by themselves, and some of the consequences is likely to be unpleasant, therefore consumer behavior involves risk from this sense [3]. Bauer specifically pointed out that he was concerned only subjective risk (i.e. perceived risk), not care about the real risk. Soon afterwards this field attracted a large number of the attention of scholars. Cox and Rich said that perceived risk is the perceived nature and quantity of the risk when consumers consider specific purchase decisions [4]. Stone and Gronhaug defined perceived risk as consumers' subjective expectation of loss, the more certain they perceive the subjective expectations of loss, the greater they

perceived the risk [5]. Cunningham divided perceived risk into two factors: uncertainty and consequence. Uncertainty refers to consumers' subjective probability of something occurs or not. Consequence is the hazard of the results after decision-making [6]. This view was endorsed by most scholars.

When Bauer put forward the concept of perceived risk, he did not indicate the specific types of perceived risk. In the following 40 years, a large number of scholars have done systematic studies on the dimensions of perceived risk. Jacoby and Kaplan chose 148 students as research objects, measured perceived risk of 12 different consumer goods, the result showed that economic risk, functional risk, physical risk, psychological risk, social risk these five dimensions explained 61.5% of the variances of the overall risk[7]. In the year 1993, Stone and Gronhaug in their research verified the existence of these six dimensions including economic risk, functional risk, physical risk, psychological risk, social risk and time risk, and found that it was not necessarily independent of each other between the various dimensions. Since all risks are perceived by individuals and perception is concerned with psychological activity, the psychological dimension of perceived risk should be highly correlated with other dimensions [5]. Later, Featherman and Pavlou predicted consumer acceptance level of electronic services from the perspective of perceived risk; their works verified that economic risk, functional risk, psychological risk, social risk, privacy risk and time risk are the six dimensions exist in the internet consumer adoption [8].

B. System usability of mobile banking service

The concept of system usability first appeared in the field of software engineering in 20th 70s, and then it has been widely applied to many other areas. Until now the definition of system usability has not reached a unanimous view among academic community and industry community. According to international standards ISO 9241211[9], system usability is the effectiveness, efficiency and satisfaction of specific products which provide specific services in specific situations. System usability is a basic natural properties expressed in the interaction course between product and its users. It is the quality of products seen from the users, embodies the core competitiveness of products. Through the research of system usability, the final product must be well in line with the target users' cognitive thinking, behavior and demand, improve the system usability level and satisfaction level of products.

At present the studies on system usability of mobile banking are rarely seen, but the studies on system usability of mobile commerce has been mature. Mobile banking as a specific application of mobile commerce, there are many similarities in system usability between mobile banking service and mobile commerce. It is justifiable to take example by existing research to study on the system usability of mobile banking service. In the studies of mobile commerce usability, the most studied conclusions regard to the limitations of mobile devices and the characteristics of WAP site in mobile applications [10]. Later, Olavarr and Maguirem respectively considered mobile communication network and situational factors into the system usability factors of mobile commerce [11][12]. Qingfei Min summarized the previous research results, proposed an integrated model combined with the four factors mentioned above [13].

III. ADOPTION MODEL OF MOBILE BANKING SERVICE BASED ON PERCEIVED RISK AND SYSTEM USABILITY

After systematically reviewed the literature about perceived risk and usability of mobile banking service adoption, it is discovered that the studies on the impact of perceived risk on mobile banking service are rarely seen. In reality, perceived risk is indeed existed when users decide whether to accept mobile banking service. Therefore, on the basis of Featherman and Pavlou's research result about the dimensions of perceived risk, this paper introduced perceived risk from the fields of traditional offline shopping and online shopping into the field of adoption factors of mobile banking service, with a view to recognize the impact factors of mobile banking service adoption more comprehensively and more systemically.

Meanwhile, owing to the differences between mobile banking service and traditional online and offline shopping, the usability of mobile banking service not only affects the service quality level, but also directly impacts on all dimensions of perceived risk of mobile banking service. In a sense, it is rational to believe that the usability of mobile banking service is the root cause of consumers' perceived risk, and usability holds an overall influence on consumers' perceived risk. Therefore, summing the analysis mentioned above, considered the perceived risk factor and usability factor of mobile banking service adoption, this text proposed an adoption model of mobile banking service based on perceived risk and usability

A. Perceived risk of mobile banking service

Perceived risk is the psychological feeling and subjective understanding of various objective risks, it derives from the objective risk but distinct from the objective risk. The research on the dimensions of perceived risk is the basis of research on perceived risk. At present, perceived risk has been used to explain the risk of offline and online shopping behavior. In these related studies, the study conclusion made by Featherman and pavlou has been widely recognized. They researched the buying behavior of consumers who accept electronic service provided in the internet, they concluded that perceived risk includes economic risk, functional risk, psychological risk, social risk and time risk these six dimensions [8]. In mobile situations, consumers' perceived risk also exists the six dimensions mentioned above, but its specific meanings are inevitably different. Therefore, on the basis of Featherman and Pavlou's research result about the

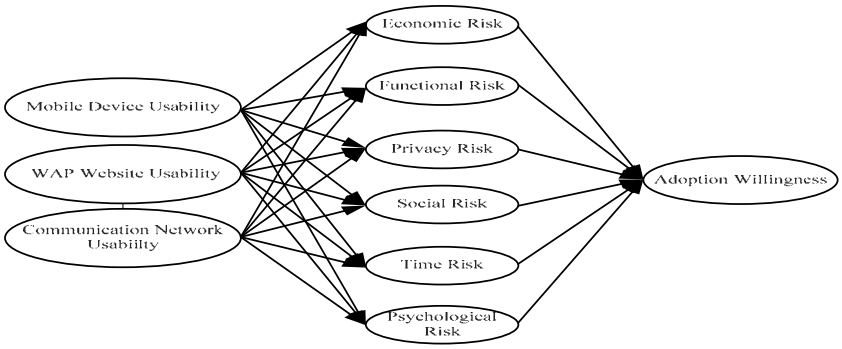


Figure 1. The research model of the adoption willingness of mobile banking service based on perceived risk and usability

dimensions of perceived risk, this text made the following new interpretation of the six dimensions.

Economic risk is the possibility of financial losses due to the use of mobile banking service, such as financial losses caused by password theft or wrong operation etc. functional risk is the possibility of unavailable service or the service cannot satisfy users' needs which provided by mobile banking service. Privacy risk is the occurrence possibility of loss of personal information in the course of using mobile banking service. Social risk is the possibility of users who are not accepted or agreed by other people due to the use of mobile banking service. Time risk is the possibility of the loss of time due to the use of mobile banking service, such as the long period of transaction processing, the long latency of customer service, etc. Psychological risk is the possibility of mental stress of the users due to the use of mobile banking service. This spirit pressure may come from the outside world such as the non-recognition of their friends and family, and may also come from themselves such as the irritable mood of financial losses when the response time of a certain type is too long.

B. System usability of mobile banking service

The system usability of mobile banking service is the effectiveness and satisfaction degree of the users who accept the mobile banking service. The mobile banking service is provided by mobile device terminals, the stability of mobile communication network and the convenience and security of WAP site will be also impact on the usability of mobile banking service. According to the previous research results of system usability, this text divided the usability of mobile banking service into three dimensions: the usability of mobile devices, the usability of WAP site, the usability of mobile communication network.

The usability of mobile devices is the usability of the input-output devices, computing speed, storage space and response speed, etc. [14]. The usability of WAP site is the usability of content display, content navigation, and human-computer interaction of the WAP Site. The usability of mobile communication network is the stability of network access, network coverage and data transfer rates [15].

C. Influences of system usability on perceived risk

Mobile devices are the carrier of mobile banking service which can provides high-quality service; the usability of mobile devices has an important impact on consumers' adoption of mobile banking service. Compared with traditional desktop computer and laptop, the usability of mobile devices exist obvious defects, these defects will affect consumers' perceived risk. For example, the screen size of mobile device and key region are small, this brings inconvenience when users make input-output operations. The limitation of battery capacity may shut down the service in the course of using mobile banking service. Besides, due to the continuous spread of mobile viruses, the stability of mobile devices system are suffered grave threat, this bring users privacy risk and economic risk. To sum up, it is convincing to say that the usability of mobile devices inevitably affect all dimensions of consumers' perceived risk.

The technology achievement of mobile banking service has a variety of solutions; among them the solution based on WAP is more convenient and practical than others. At present, the usability of WAP site includes following several aspects: first, it is inconvenient for users to view information and easily overlook the key messages because of the limitation of screen size; second, the poorness of human-computer interaction. The buttons of mobile device are small, and the amount of information displayed is limited, when consumers are making a certain type of operations, it is likely to cause the puzzlement of consumers operations if WAP site cannot provide necessary tips and navigations. WAP site is the interface of high-quality mobile banking service provided for consumers; the operability, reliability and convenience of WAP site directly affect all dimensions of consumers' perceived risk.

Mobile communication network is the foundation and platform of mobile commerce. The quality of mobile communication network is another key factor of the usability of mobile banking service. In mobile situations, the consumers' locations are constantly changing at any time, if consumers choose the use of mobile banking service, it must be ensured the connection of communication network, and this puts forward a higher requirement of the stability and coverage of mobile network signal. At the same time, data transfer rates will also be a great impact on the consumers' satisfaction sense of mobile banking service.

IV. ENLIGHTMENT AND DISCUSSION

To take effective measures to reduce consumers' perceived risk not only attracts widely attentions of the experts, but also has a positive role in promoting the development of mobile banking service, it is of great practical significances. Perceived risk is the important determinant of consumers' adoption willingness of mobile banking service, and the usability of mobile banking service is the root cause of consumers' perceived risk. Therefore, to promote the adoption of mobile banking service must get start from improving the usability and reducing the perceived risk.

The improvement of the usability of mobile banking service. Because the differences of consumers' handheld mobile devices, the improvement of the improvement of the usability of mobile banking service becomes extremely important. In the design and development of WAP site, it should be simplify consumer operation as far as possible, provide popular concise content navigation, enhance the readability of web content and must ensure the safety and reliability of WAP site. Second, the mobile banking service providers should choose mature technical standards and safe, reliable communication network to ensure the stability and coverage of communication network.

The decrease of consumers perceived risk. Perceived risk is consumers' subjective feeling in the course of using mobile banking service. In order to effectively reduce consumers' perceived risk, besides improving the usability of mobile banking service, the following aspects should also be noted: first, mobile banking service provider must establish a sound credit system and security system in order to make consumers produce a trust sense of accepted products and services and ensure the safety and integrity of consumers' information; second, to strengthen the positive publicity guide, take necessary marketing tools and popularize the mobile banking service knowledge are helpful to dispel misunderstanding and enhance the awareness of mobile banking service; third, to introduce the third-party certification, enhance the trust sense of transactions and website security are good to reduce consumers' perceived risk.

V. CONCLUSIONS AND LIMITATION

In mobile situations, consumers' perceived risk is the important determinant of whether they adopt the mobile banking service. Meanwhile the root cause of consumers' perceived risk is the existence of system usability of mobile banking service. After reviewing relevant literature, this text introduced perceived risk into the study of adoption behavior of mobile banking service, and pointed out that the usability of mobile banking service is the important factor of consumers' perceived risk. On the basis of previous research results of perceived risk and system usability, this text proposed a research model of the adoption willingness of mobile banking service based on perceived risk and system usability. This model framework can help scholars to understand the perceived risk and system usability issues of mobile banking service can also help mobile banking provider to improve the product and service quality of mobile banking service, further promote consumers' acceptance and adoption willingness of mobile banking service.

The limitation of this text is the proposed preliminary model needs to be further improved and empirical test. However, it is believed that this model has certain reference significance for the future research on the adoption willingness of mobile banking service, and the research from the perspective of perceived risk and system usability will bring new impetus to the adoption of mobile banking service.

ACKNOWLEDGMENT

The authors wish to thank the anonymous reviewers for their work.

REFERENCES

[1] Tiwari. R, "Mobile Banking as Business Strategy: Impact of Mobile Technologies on Customer Behavior and Its Implications for Banks," PICMET 2006 Proceedings, 2006, pp. 1935-1946.

[2] iResearch, "China mobile internet behavior research report," 2003, http://news.iresearch.cn/Zt/83611.shtml.

[3] Bauer, "Consumer Behavior as Risk Taking," Proc AmerMarkAssoc, 1960, pp. 389-398.

[4] Cox. D. F and Rich. S. J, "Perceived risk and consumer decision making," Mark Res, vol. 1, 1964, pp. 32-39.

[5] Stone, Robert. N, Gronhaug, Kjell, "Perceived Risk: Further Considerations for the Marketing Discipline," European Journal of Marketing, vol. 27,1993, pp. 39-51.

[6] Cunningham. S.M, "The Major Dimensions of Perceived Risk, In F.C.Donald (Ed.), Risk Taking and Information Handling In Consumer Behavior," Boston: Harvard University Press, 1967, pp. 82-108.

[7] Jacoby. J and Kaplan. L, "The components of perceived risk ,in Venkatesan,M. (Ed.),"Proceeding of the 3rd Annual Conference, Association for Consumer Research,Chicago,IL, 1972, pp. 382-393.

[8] Featherman Mauricio S and Pavlou Paul A, "Predicting e-services adoption: a perceived risk facets perspective," Human-Computer Studies , vol. 59, 2003, pp. 451-474.

[9] Nielsen. J, "Usability engineering," San Francisco: MorganKaufmann, 1994, pp. 875.

[10] Wade. V. P and Ashman. H and Barry. S, "Adaptive hypermedia and adaptiveWebbased systems," Proc of the 4 th International Conference on AH, 2004, pp. 732.

[11] Olavarr Ietald and Navaa. A, "Wireless communications: a bird's eye view of an emerging technology," Proc of International Symposium on Communications and Information Technology, 2004, pp. 541-546.

[12] Maguirem, "Context of use within usability activities," Int J Human-Computer Studies, vol. 55 , Otc 2001, pp. 453-483.

[13] Qingfei Min and Shuangming Li, "From usability to adoption: new m-commerce adoption study framework," Application Research of Computer, vol. 5, 2009, pp. 1799-1082.

[14] Mennecke. B. E and Strader. T. J, "Mobile commerce: technology, theory, and applications," Hershey PA: Idea Group Publishing, 2002, pp. 849-856.

[15] Leuven. B, "Wireless communications," Piscataway, 2006, pp. 934-943.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 055-057

A Rate Control Scheme of the Even Low Bit-rate Video Encoder

Gan Yong [1], Zhang Li [2], and Liu Yingfei [2]
[1] Depa. of Computer Science Zhengzhou University of light industry Zhengzhou China
Email: Ganyong@zzuli.edu.cn
[2] Depa. of Electronic Science Henan Information Engineering School Zheng zhou China
Email: {zhangli, Liuyingfei}@zzuli.edu.cn

***Abstract*—According to the application demand of remote surveillance system based on PSTN, the paper puts forward a rate control strategy of very low bandwidth. The strategy presents the rate control arithmetic of I frame and introduces quadratic rate distortion mode, which solves the problem of the rate control used at very low bandwidth preferably. The results of experiment show that the arithmetic reduces the delay of encoder buffer greatly, and improves the quality of reconstructed images clearly at the same time.**

***Index Terms*— bit rate control, video encoder，even low bit-rate, verification model**

I. INTRODUCTION

Virtual code rate control strategy is the key for low delay and high Quality video under narrow bandwidth conditionCode rate control has frame level and macroblock level for code rate control. Code rate control in frame level uses one and the same quantized value within a frame, macroblock is opposite.

This article provided a new approach of video encoder for long range video transmission based on extreme low bandwidth line(like PSTN)and this strategy leads in two steps: code rate control model and realization of I frame code rate control, and improve on P frame code rate control algorithm， effectively increase precision and code rate of image transmission for video encoder.

II. EVEN LOW CODE RATE CONTROL ALGORITHM

Code rate control algorithm of VM8[1] only provide control model for P frame，not I frame. The reason is that the algorithm hypotheses the first coded frame for I frame， the rest are P frames. In practice, it need a certainty number of I frame to adapt transmission fault toleration and information retrieval coding. In long range video frequency supervisory control application， we increase I frame code rate control in order to gain optimized output code stream.

If then the second order R- D Model[2] can be expressed as：

$$Q_t^2 + a_1 r Q_t + a_2 r = 0 \quad (1)$$

where a_1, a_2 are the first and second order coefficient respectively, r is related to the digit of present frame object, occupied digit of basic information and image absolute dispersion, derived as below:

$$r = \frac{M_t}{H_t - B_t} \quad (2)$$

Where the digit of present frame object， t t Code rate is primarily controlled by adjusting Q_t which can be obtained by solving second order equation (2) below:

$$Q_t = \frac{M_t \cdot a_1 + \sqrt{(M_t \cdot a_1)^2 + 4 \cdot (B_t - H_t) \cdot M_t \cdot a_2}}{2 \cdot (B_t - H_t)} \quad (3)$$

In equation (3), and linear regression technology according to actual quantized value and coding digit of encoded frame. After each encryption, and In addition, as encryption of first frame proceed, and are usually unknown, Equation(3) can not be used for calculating , here we assume and assign it to be a certain value (typical 15) ,then again use coded real quantized value and its digit to obtain and by linear regression technique.

Code rate control strategy of video encoder falls into four phases： initialization phase , precoding phase , encrypt phase as well as aftertreatment Phase[3-5]. The core of code rate control is quantized parameter and model renewal.

A. Code Rrate Control Algorithm for I Frame

I frame is established by a second order distortion model similar as P frame, and its code rate control involved with primal problem as following ： determination of initial quantized value, distribution of target digit and determination of quantized value.

a. Determination of initial quantized value

Both initial quantized value for I and P frames are constant (typically), since the image content and transmission bandwidth are different, filling ratio and image quality in buffer are varied significantly at the beginning of the first several frames. Shown in equation (1), when ， code rate control model degrades into a first order model (seen equation 1)，this model only has one parameter with empirical value (typical value 130)，then using equation(3)to acquire initial quantized value.

by equation(3), target digit of I frame (method of computation is described in details in subsequent chapter), MAD value , the head digit model coefficient(empirical value)could work out initial quantized value.The first quantized value of P frame after I frame, has access the same value as

b. Distribution of target Digit

I, P, and B are frames of distributed target digit provided by TM5 of MPEG-2 , but this calculation is complex and closely related to other R-D models. This text offered a target digit distribution method for combining the second order rate distortion model

In order to simply equation (1)，I and P frame has:

$$\frac{B_t^I}{M_t^I} = \frac{a_1^I}{Q_t^I} \tag{4}$$

$$\frac{B_t^P}{M_t^P} = \frac{a_1^P}{Q_t^P} \tag{5}$$

The notation I and P in the right upper corner of every variable in the two equations above are only used to distinguish which belongs to I or Q parameter. In one video frequency sequence, it has:

$$B_t^I \cdot N^I + B_t^I \cdot N^P = T \cdot r_t \tag{6}$$

Where N^I is number of I frames. N^P is number of P frames, T is successive period of video sequence (second), r_t is output bit rate(bit per second).

Assume image quality is constant in one video sequence (quantized value of I and P frames are the same), then power.

$$Q_t^I = Q_t^P \tag{7}$$

Combine equation (4),(5),(6),(7):

$$B_t^I = \frac{T \cdot r_t}{N^I + \frac{M_t^P \cdot a_t^P}{M_t^I \cdot a_t^I} \cdot N^P} \tag{8}$$

$$B_t^P = \frac{T \cdot r_t}{N^P + \frac{M_t^I \cdot a_t^I}{M_t^P \cdot a_t^P} \cdot N^I} \tag{9}$$

Target digit of I frame B_t^I can be calculated by equation (8), where M_t^I and M_t^P are the same type frame of MAD value, then a_1^I 、 a_2^I and a_1^P 、 a_2^P are calculated by linear regression technique.

c. Quantized Value Calculation

Calculation for quantized value Q_t of I frame has the similar method as P frame.

With respect of given values, target digit B_t , current frame MAD value M_t , a_1 and a_2 ,then Q_t is obtained by solving the quadric equation in this variable.

B. Improved Code Rate of P Frame

a. Target digit distribution

Target digit distribution of P frame has access to equation(9)，when only I frame in sequence and the rest are equal to P frame，equation(6)can be transferred into simplified form:

$B_t^P = \frac{T \cdot r_t - A_t^I}{N^P}$, A_t^I is actual coding digit of I frame.

Compared with results B_t^P , we need filling ratio of buffer zone to do the adjustment which is described in literature [2] and [6].

b. Quantized value adjustment

In order to stay picture quality stable and limit regulatory amplitude peak that is not exceeding 0.25 times of previous frame(Q_{t-1}).In this way，if quantized value of previous frame is small and current frame calculated is large，after the adjustment, the current one become smaller, that results in over size of actual coding digit of current frame，increase transmission delay.

This article improved algorithm of quantized value of P frame in order to solve this problem:

If $Q_t < 0.75Q_{t-1}$, $Q_t = 0.75Q_{t-1}$

If $Q_t > 1.25Q_{t-1}$, three steps are taken as follows:

a)Firstly, assign $Q_t = 1.25Q_{t-1}$, use equation (1) to work out code digit B_t of quantized Q_t .

b)If $B_t > 2R_P$ (R_P is deleted digit from coded frame in buffer zone), assign $B_t = 2R_P$, then turn to c, otherwise, assign $Q_t = 1.25Q_{t-1}$ to directly finish this process.

c)According to B_t , use equation (1) to calculate new Q_t as quantized value of current frame.

III. EXPERIMENTAL RESULTS AND ANALYSIS

A. Experiment Instruction

Sequences are tested as five different code rates at 9.6kbps, 28.8kbps, 33.6kbps, 44.8kbps, and 56kbps respectively. The benchmark: GOP height is 15, frame rate 3fps, and GOP height is 30 at 9.6 kbps of code rate. The diagram displayed below shows the testing results between the new algorithm and VM8 algorithm in MPEG-4. Compared with curve 3, obviously, new algorithm takes on distinct dominance in code rate line.

Figure 1. Image Quality of VM8

Figure 2. Image Quality of improved algorithm at 33.6 kbps

In figure 2, it displays that basketball images received in terminal and was very obvious that the transmission quality in improved algorithm is much smarter than that in VM8 at 33.6 kbps (typical PSTN bandwidth).

B. Result Analysis

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Do not use abbreviations in the title unless they are unavoidable.

Experiment results could be concluded from figure 1:

TABLE I. EXPERIMENT RESULTS

Test code rate (kbps)	Algorithm	Actual code rate (kbps)	Average PSNR-Y (dB)	Maximum buffer delay (s)	Code flow deviation(%)	Number of overflow
9.6	Initial	10.85	31.62	21.845	65.75	75
	Improved	9.49	32.15	2.673	31.23	0
28.8	Initial	32.53	31.54	6.924	62.58	75
	Improved	29.07	31.64	1.151	43.00	0
33.6	Initial	36.21	31.83	6.356	68.20	75
	Improved	33.92	32.42	1.249	41.38	0
44.8	Initial	44.96	32.93	3.272	55.42	10
	Improved	44.47	33.65	0.506	39.12	0
56	Initial	56.09	33.76	2.603	50.04	1
	Improved	54.88	34.59	0.55	46.21	0

a. Some improvement are taken into account such as appropriate arrangement of quantized value of I frame, adjustable amplitude for P frame and dynamic adjustment of GOP length to largely decrease delay of maximum buffer and avoid overflow in buffer zone.

b. PSNR values in improved algorithm are all higher than those in initial one in all different bandwidths. Furthermore, seen from figure 1 it is observed that in identical bandwidth, PSNR of reestablish video in improved algorithm is higher and code digit is less in improved algorithm in the same quality of picture.

c. Analyzed from the view of output code stream deviation(D_b), mean deviation of output code rate in improved algorithm is even smaller.

IV. CONCLUSION

This article is based on the application of low bandwidth line，I frame code rate control algorithm is raised by VM8 code rate control algorithm and discussed in selection of initial quantized value , distribution of target digit, determination of quantized value. Meanwhile the strategy for the improvement of initial algorithm assigns appropriate quantized value with its change. Experimental results indicate that the new method improved control precision and whole reestablish video quality, and decrease the delay of output buffer.

REFERENCES

[1] “Information technology – coding of audio-visual objects, part 1: systems, part 2: visual, part 3: audio,” ISO/IEC JTC1/SC29/WG11, FCD 14496, Dec. 1998.J. Clerk Maxwell, *A Treatise on Electricity and Magnetism*, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] T. Chiang and Y. Q. Zhang, “A new rate control scheme using quadratic rate distortion model,” IEEE Trans. On Circuits and Systems for Video Technology, vol. 7, no. 1, pp. 246-250, Feb. 1997.

[3] Dapeng Wu, Y. Thomas Hou, Wenwu Zhu, et al, “On End-to-End Transport Architecture for MPEG-4 Video Streaming over the Internet,” IEEE Trans On Circuits and Systems for Video Technology, vol. 10, no. 6, pp. 923-941, Sep. 2000.

[4] Sun Yu, and Ishfaq AHMAD, “A new rate control algorithm for MPEG-4 Video Coding,”Accepted by *Visual Communication and Image Processing*, SPIE, San Jose, Jan. 2002..

[5] “MPEG-2 Video Test Model 5,” ISO/IEC JTC1/SC29/WG11 MPEG93/457, Apr 1993.

[6] B Zhou, X Li, “Bit rate control strategy based on MPEG-4 standard ”, Computer Science,Vol.30,No.10,2003.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 058-060

A New Penalty Function Algorithm in Constraints Posynomial Geometric Programming

Jing Shujie[1] , Han Yanli[2] , Han Xuefeng[2]
[1]Institute of Mathematics and Information Science, Henan Polytechnic University, Jiaozuo,China
Email: jsj_jjj@hpu.edu.cn
[2]Institute of Mathematics and Information Science, Henan Polytechnic University, Jiaozuo,China
Email: hanyl@hpu.edu.cn,hanxuefeng@hpu.edu.cn

***Abstract*—Geometric programming is a special nonlinear programming ,it's application is very extensive.Using the existing results and characteristics of constraints posynomial geometric programming and penalty function technique, the author designs a new algorithm for constraints posynomial geometric programming and proves the convergence of the algorithm.**

***Index Terms*—constraints posynomial,geometric programming,exponent penalty function,convergent theorem**

I. INTRODUCTION

Geometric programming is a special nonlinear programming, it's application is very extensive. There are many applied examples in economic analysis and other economic activities, chemical balance, the engineering analysis and engineering design.And the application examples of geometric programming are also increasing. However, the development of it's theory and algorithm has been slow, it is because that the difficulty of posynomial geometric programming are mostly greater than zero,at the same time the theoretical and practical calculations of generalized geometric programming who belongs to DC programming and whose duality programming belongs to non-smooth optimization are very difficult. Therefore, researching the theory of geometric programming algorithm has great significance.

Consider the following geometric programming:

$$(SGP)\begin{cases}\min y_0(t)=\sum_{j=1}^{T_0}C_{oj}\prod_{i=1}^{m-1}t_i^{d_{0ij}} \\ s.t\begin{cases}y_l(t)=\sum_{j=1}^{T_l}C_{lj}\prod_{i=1}^{m-1}t_i^{d_{lij}}\le 1, l=1,2,\cdots,L \\ t=(t_1,t_2,\cdots,t_{m-1})^T>0\end{cases}\end{cases}$$

where $C_{ij}>0, d_{tij}$ are any real numbers, T_l are non-negative integer.

In [3] and [4], the author transformed the general constraints posynomial geometric programming problem into the problem (P) using the duality theory. In this paper, the author attempts to find a new algorithm for the geometric programming (SGP) using penalty function. Based on [3] and [4] ,the author constructs a new algorithm for geometric programming through the problem (P).

Penalty function methods are important and more practical methods for solving constrained optimization problems. Its basic idea is transforming a constrained problem into a single unconstrained problem or into a sequence of unconstrained problem and by solving these unconstrained problems to solve the constrained problem. To use unconstrained optimization problem instead of constrained optimization problem, the objective function of the unconstrained optimization problem must be a proper combination of the objective function of the constrained optimization problem and constraint functions. Usually the constraint functions who construct a penalty item are placed into the objective function via a penalty parameter in a way that penalizes any violation of the constraints. The construction principle of penalty items are: if the current iteration point is not feasible, it is necessary for its implementation of punishment, and the punishment value is increasing with the improving of the infeasibility of the point; no penalty for feasible points. The role of penalty items is to force the iterative point closer and closer and finally in the feasible domain with the progress of iteration. Constructing different penalty items corresponds different penalty function methods. Therefore, research on the different penalty items has important theoretical and practical value.

Consider the following constraints positive define geometric programming P:

Minimize $f(x)$

subject to $Ax=b, x>0$,

$$f(x)=C^Tx+\sum_{i=1}^{n}x_i\ln x_i-\sum_{l=0}^{L}e_l^Tx\ln e_l^Tx=-\ln d(p)$$

where

$$x=(x_1,x_2,\cdots,x_n)^T=(p_{01},\cdots,p_{0T_0},\cdots,p_{L1},\cdots,p_{LT_L})^T$$

$$C=(c_1,c_2,\cdots,c_n)^T=-(\ln c_{01},\cdots,\ln c_{0T_0},\cdots,\ln c_{L1},\cdots,\ln c_{LT_L})^T$$

$$A=\begin{bmatrix}1 & 1 & \cdots & 1 & 0 & \cdots & 0 & \cdots & 0\\ d_{0,1,1} & d_{0,1,2} & \cdots & d_{0,1,T_0} & d_{1,1,1} & \cdots & d_{1,1,T_1} & \cdots & d_{L,1,T_L}\\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots\\ d_{0,m-1,1} & d_{0,m-1,2} & \cdots & d_{0,m-1,T_0} & d_{1,m-1,1} & \cdots & d_{1,m-1,T_1} & \cdots & d_{L,m-1,T_L}\end{bmatrix}$$

where $m-1$ is the dimension of the optimization variable t , $b=(1,0,\cdots,0)^T\in R^m$, $e_l=(0,\cdots,0,1,\cdots,1,0,\cdots,0)^T\in R^n$, $l=0,1,\cdots,L$,

Corresponding author: Han Yanli,E-mail:hanyl@hpu.edu.cn

that is the T_l components of e_l are 1 from the $1+\sum_{k=0}^{t-1}T_k$ component to $\sum_{k=0}^{t}T_k$ component , the remaining components are 0.

In [1],the author refer to the function $L(x,\mu)=f(x)+\mu\alpha(x)$ as the auxiliary function, where penalty item α is of the form $\alpha(x)=\sum_{i=1}^{m}\left|\sum_{j=1}^{T_0+T_1+\cdots+T_L}a_{ij}x_j-b_i\right|^p$ and p is a positive integer and $\mu>0$ is a pe0ired of references [5] and [6], we use good natures of the exponent functions to construct a class of new penalty function. In section 2 , we constructs the new penalty function and gives the corresponding algorithm. In section 3, in the weaker conditions convergence theorem of the algorithm and its proof are given.

II. New penalty function and algorithm

A. New penalty function

We select exponent function as penalty function, as follows:

$$\alpha(x)=\sum_{i=}^{m}\left|\exp\left|\sum_{j=1}^{T_0+T_1+\cdots+T_L}a_{ij}x_j-b_i\right|-1\right|^2$$

Then auxiliary function is

$$L(x,\mu)=f(x)+\mu\sum_{i=1}^{m}\left|\exp\left|\sum_{j=1}^{T_0+T_1+\cdots+T_L}a_{ij}x_j-b_i\right|-1\right|^2$$

B. Algorithm

- To figure axis labels, Initialization Step Let $\varepsilon>0$ be a termination scalar. Choose an initial point x_1 ,a penalty parameter $\mu_1>0$,and a scalar $\beta>1$. Let $k=1$, and go to the Main Step.
- Main Step

a) Starting with x_k ,solve the following problem: Minimize $L(x,\mu_k)=f(x)+\mu_k\alpha(x)$

Let x_{k+1} be an optimal solution and go to Step b.

b) If $\mu_k\alpha(x_{k+1})<\varepsilon$ or $\nabla_x L(x_{k+1},\mu_k)<\varepsilon$, stop: otherwise, let $\mu_{k+1}=\beta\mu_k$, replace k by $k+1$,and go to Step a.

III. Convergent theorem

A. Lemma

Suppose that f,h are continuous functions on R^n .Let $\alpha(x)=\sum_{i=1}^{l}\left[\exp|h_i(x)|-1\right]^2$,and suppose that for each μ ,there is an $x_\mu\in R^n$ such that $\theta(\mu)=f(x_\mu)+\mu\alpha(x_\mu)$ .Then, the following statements hold true:

1. $Inf\{f(x):h(x)=0\}\geq\sup_{\mu\geq0}\theta(\mu)$, where $\theta(\mu)=\inf\{f(x)+\mu\alpha(x)\}$.
2. $f(x_\mu)$ is a nondecreasing function of $\mu>0$, $\theta(\mu)$ is a nondecreasing function of μ ,and $\alpha(x_\mu)$ is a nonincreasing function of μ.

Proof

Consider feasible point x , then $\alpha(x)=0$.

Let $\mu\geq0$,then $f(x)=f(x)+\mu\alpha(x)\geq\inf\{f(y)+\mu\alpha(y):y\in X\}=\theta(\mu)$.

Thus, $Inf\{f(x):h(x)=0\}\geq\sup_{\mu\geq0}\theta(\mu)$. Statement 1 follows.

To establish Statement 2, let $\lambda<\mu$.

By the definition of $\theta(\lambda)$ and $\theta(\mu)$, we have

$$f(x_\mu)+\lambda\alpha(x_\mu)\geq f(x_\lambda)+\lambda\alpha(x_\lambda) \quad (1)$$

$$f(x_\lambda)+\mu\alpha(x_\lambda)\geq f(x_\mu)+\mu\alpha(x_\mu) \quad (2)$$

Adding these two inequalities and simplifying, we get $(\mu-\lambda)[\alpha(x_\lambda)-\alpha(x_\mu)]\geq0$.

Since $\lambda<\mu$,

then $\alpha(x_\lambda)\geq\alpha(x_\mu)$.

It then follows from (1) that $f(x_\mu)\geq f(x_\lambda)$ for $\lambda\geq0$.

By adding and subtracting $\mu\alpha(x_\mu)$ to the left-hand side of (1), we get

$$f(x_\mu)+\mu\alpha(x_\mu)+(\lambda-\mu)\alpha(x_\mu)\geq\theta(\lambda).$$

Since $\lambda<\mu$ and $\alpha(x_\mu)\geq0$, the above inequality implies that $\theta(\mu)\geq\theta(\lambda)$.This completes the proof.

B. Theorem

Consider Problem P: where f,h are continuous functions on R^n. Suppose that the problem has a feasible solution.Furthermore, suppose that for each μ there exists a solution $x_\mu\in R^n$ to the problem to minimize $L(x,\mu)$,and that $\{x_\mu\}$ is a contained in a compact subset of R^n.Then

$$Inf\{f(x):h(x)=0\}=\sup_{\mu\geq0}\theta(\mu)=\lim_{\mu\to\infty}\theta(\mu)$$ Further more, the limit $\bar{x}$ of any convergent subsequence of

$\{x_\mu\}$ is an optimal solution to the Problem P, and

$$\mu\alpha(x_\mu) \to 0 \text{ as } \mu \to \infty .$$

Proof

By Part 2 of Lemma, $\theta(\mu)$ is monotone, so that $\sup_{\mu\geq 0} \theta(\mu) = \lim_{\mu\to\infty} \theta(\mu)$.

We first show that $\alpha(x_\mu) \to 0$ as $\mu \to \infty$. Let x_1 be an optimal solution to the problem to minimize $L(x,\mu)$ for $\mu = 1$.

If $\mu \geq (1/\varepsilon)[f(y) - f(x_1)] + 2$, then we have $f(x_\mu) \geq f(x_1)$.

We now show that $\alpha(x_\mu) \leq \varepsilon$ .By contradiction, suppose that $\alpha(x_\mu) > \varepsilon$. Noting Part 1 of Lemma , we get

$$Inf\{f(x): h(x) = 0\} \geq \theta(\mu) = f(x_\mu) + \mu\alpha(x_\mu)$$

The

$$\geq f(x_1) + \mu\alpha(x_\mu) > f(x_1) + |f(y) - f(x_1)| + 2\varepsilon > f(y)$$

above inequality is not possible in view of the feasibility of y .Thus $\alpha(x_\mu) \leq \varepsilon$ for $\mu \geq (1/\varepsilon)[f(y) - f(x_1)] + 2$.

Since $\varepsilon > 0$ is arbitrary, $\alpha(x_\mu) \to 0$ as $\mu \to \infty$.

Now let $\{x_{\mu_k}\}$ be any convergent subsequence of $\{x_\mu\}$, and let $\bar{x}$ be its limit.

Then

$$\sup_{\mu\geq 0} \theta(\mu) \geq \theta(\mu_k) = f(x_{\mu_k}) + \mu_k \alpha(x_{\mu_k}) \geq f(x_{\mu_k})$$ Since

$x_{\mu_k} \to \bar{x}$ and f is continuous, hence $\sup_{\mu\geq 0} \theta(\mu) \geq f(\bar{x})$ (3)

Since $\alpha(x_\mu) \to 0$ as $\mu \to \infty$, $\alpha(\bar{x}) = 0$; that is $\bar{x}$ is a feasible solution to the Problem P. In view of (3) and Part 1 of Lemma, it follows that $\bar{x}$ is an optimal solution to Problem P and that $\sup_{\mu\geq 0} \theta(\mu) = f(\bar{x})$.

Note that $\mu\alpha(x_\mu) = \theta(\mu) - f(x_\mu)$ as $\mu \to \infty$, $\theta(\mu) \to f(\bar{x})$ and $f(x_\mu) \to f(\bar{x})$ Then $\mu\alpha(x_\mu) \to 0$ as $\mu \to \infty$. This completes the proof.

REFERENCES

[1] Mokhtar S Bazaraa ,Hanif D Sherali and C M Shetty,Nonlinear Programming Theory and Algorithms Third Edition,Published by John Wiley & Sons,Inc.,Hoboken,New Jersey.

[2] YUAN Ya-xiang and SUN Wen-yu, Optimization Theory and Methods,Shanghai：Science Press ,2001.

[3] G Shujie and ZHANG Kecun,Generalized Projecting Variable Metric Algorithm for Geometric Program,JOURNAL OF ENGINEERING MATHEMATICS:Vol.15 No.1,1999.

[4] JING Shujie,BE Xiaosan and ZHANG Kecun,A Polynomial Time Algorithm in Positive Define Geometric Programming with Constraints, JOURNAL OF ENGINEERING MATHEMATICS:Vol.19 No.2,2002.

[5] LIU Fang, SHAN Rui , WANG Fang and CHEN Yah-de, A Kind of Revised Penalty Function Algorithm, Journal of Gansu Lianhe University :Natural Sciences 2008，1（3）：40-42.

[6] ChENG Guixiang and ChEN Lanping, Hyperbolic Penalty Function Multiplier Method, Hyperbolic Penalty Function Multiplier Method 2007，5（5）：6-10.

[7] GONG Chun and WANG Zhenglin, Proficient in MATLAB Optimization Calculation,Beijing:Electronic Industry Press，2009.

[8] YANG Mingsheng,XIONG Xiwen and LIN Jianhua, MATLAB-based and mathematical software,Dalian：Dalian University of Technology Press，2003.

[9] G.DI PILLO and L.GRIPPO,A New Augmented Lagrangian Function for Inequality Constraints in Nonlinear Programming Problems, JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS,Vol.36,No.4,APRIL,495-519 1982.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology (ISCSCT '10)
Jiaozuo, P. R. China, 14-15, August 2010, pp. 061-065

Research on control strategy of a novel stand-alone photovoltaic system

Liu Jie[1], Liu Sanjun[2]
[1] Department of Computer Science and Technology Henan Polytechnic University
Jiaozuo, China
Email: liujie@hpu.edu.cn
[2] Department of Computer engineering Jiaozuo University
Jiaozuo, China
Email: liusanjun1975@163.com

***Abstract*—Solar power has become the fastest growing and most widely application in the view of solar energy application. A novel stand-alone photovoltaic system topology is proposed in this paper, in which push-pull output circuit is adopted. The system consists of battery array, CUK charger, battery, boost transformation and inverter. Charging strategy based on the battery current regulation principle is adopted in CUK charger, by the way which proposed, a 100% battery state of charge is reached in shorter time. Boost topology using push-pull structure transformation. Closed-loop control method is simple and effective; it is conducive to reducing the system volume and further improves efficiency.**

***Index Terms*—Solar, stand-alone photovoltaic, inverter, topology**

I. INTRODUCTION

With the increasing human demand for energy, fossil energy reserves are becoming exhausted, while the use of fossil fuels has brought serious consequences to the human environment. In the world today, with the energy crisis and all kinds of soaring energy prices, every country have looked into the renewable energy. Solar is undoubtedly the most impressive in all kinds of renewable energy resources. Solar energy shows superiority on many aspects, such as a reserve of "infinity," the universality of existence and utilization, the economic efficiency revealed gradually and so on. Its' exploitation is a effective approach to resolve the energy shortage, the environmental pollution , the greenhouse effect and other problems ultimately caused by the energy of conventional energy, especially fossil energy .It is a ideal substitute energy of human ,as in [1].

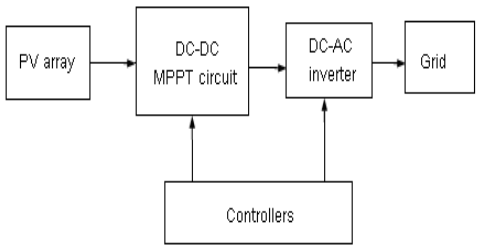


Figure1. Stand-alone photovoltaic power generation system

Fund project: sponsor by Henan Province key scientific and technological project (082102240008)

At present, in the stand-alone photovoltaic power generation system, the universal adoption of the structure is shown in Fig. 1. First, collect solar energy using solar cells, then charge battery through the DC / DC converter. As the voltage of storage battery is too low, which often can not meet the requirements of inverter, so it still needs a boost converter to raise the DC voltage .And finally transform the DC into the 220V/50Hz AC through the inverter for users.

II. SYSTEM DESIGN AND WORKING PRINCIPLE

As the design is independent of type of photovoltaic conversion system, it is essential part of the battery by adding a DC / DC converter (charger) to achieve the maximum power point tracking and battery charge and discharge management between the solar cells and batteries. As the battery voltage is low, can not meet the requirements of the DC bus voltage inverter, need to join a boost circuit to increase the DC voltage between the input of the battery and inverter. Therefore, with the final inverter, the output from the solar cell output to the system, solar energy through the three transformation, namely DC / DC conversion, boost conversion, inverter, as in [1-3].

Therefore, the design use this more traditional three-tier systems architecture, system block diagram shown in Fig. 2. This architecture is characterized by reliable, independent control of a simple, easy system of modular software and hardware design.

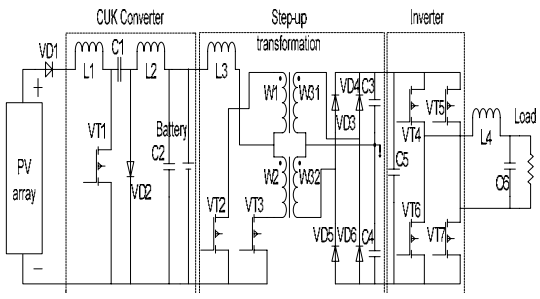


Figure2. Stand-alone photovoltaic power generation system

A. Topological charge

In this paper the design of the independent operation of photovoltaic conversion system, with the need for battery

storage of solar energy to prepare for the case without the use of sunlight, the battery becomes an essential system component. As the battery voltage of the solar cell when maximum power is greater than the selected voltage, so this paper based on CUK circuit (shown in Fig. 2) for photovoltaic power generation system MPPT charge control strategy, and the solar cell board output in series a diode VD1, to prevent the battery's energy to the solar cell anti-irrigation and damage PV panels.

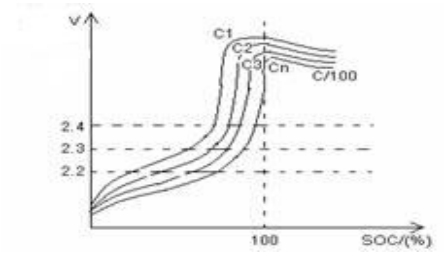


Figure3. Battery charging strategy principle

Fig. 3 shows the photovoltaic battery charging strategy of basic idea, which C1, C2, C3, Cn for the battery charging rate; C for the battery capacity (Ah); V for the battery charge voltage； SOC for the battery factory real-time capacity and the percentage of rated capacity, It reflects the relationship between the battery charge voltage and battery level status of the capacity, as in [4-5]. Charge control process is as follows.

1） Use current sensors continuously detect the actual battery charge current, if the current is less than the maximum allowable charge current setting (initial setting value C/100, C is battery capacity), then call the MPPT process to achieve the maximum use of solar energy; or by adjusting the DC / DC open circuit switch duty cycle to limit it is not greater than the maximum allowed charging current set value.

2) Using voltage sensors continuously detect the battery voltage at both ends and compared with the overshoot voltage, when it is greater than the set value of overcharge voltage, reduce the maximum allowable charge current set value, and repeat the process 1).

3) When setting the maximum allowable charge current decreases from C/10 to C/100, and reached the overcharge voltage point, it shows that the batteries is full and should end of the process.

4) When the battery charge current to C/100, maintain a small current charging the battery in order to compensate for battery self-discharge losses. When the detected power is re-allowed to the maximum current to the battery is charging.

The traditional control strategy charged photovoltaic system, only used in the fast-charge stage MPPT control, but the system is, whether for which charge stage, as long as the actual charge current is not greater than the maximum allowable charge current setting, that is, can be used MPPT charge control, which makes the utilization of solar array output power can be greatly enhanced. The charge control can not only make full use of the PV array output power, and combined with MPPT technology allows short period of time can make the battery fully charged state to prolong battery life.

B. Boost topology change

Boost voltage can used by a Boost converter, push-pull converters, full bridge converter, half-bridge converter and two-transistor forward converter and so on. Boost is not an input-output isolation and the other five were isolated in the converter, so the Boost converter is excluded in order to achieve step-up and isolation. Full bridge, half bridge, two-transistor forward converter are relatively the occasion high pressure to low pressure change, but the system input is the battery voltage while the output is AC 380V, the input side of the lower voltage and current greater than some converters they are not suitable. So only the most suitable for such low voltage high current input and large output and can play an electrical isolation of the push-pull converter.

Transform part of the step-up in Fig. 2 shows, the push-pull side of the current, input from the battery supply voltage full-bridge side as the output DC high voltage bus connected. At the boost mode, VT2, VT3 work as a switch, VD3 to VD6 work as a rectifier worked for the whole bridge, because the presence of L, VT2, VT3 duty cycle must be bigger than 0.5, that is, VT2, VT3 can overlap the work of conduction. Assume that all switches and tubes, diodes are ideal devices, all the inductors, capacitors, transformers are ideal components, and the transformer secondary winding turns of the two equal, that is, W31 = W32.

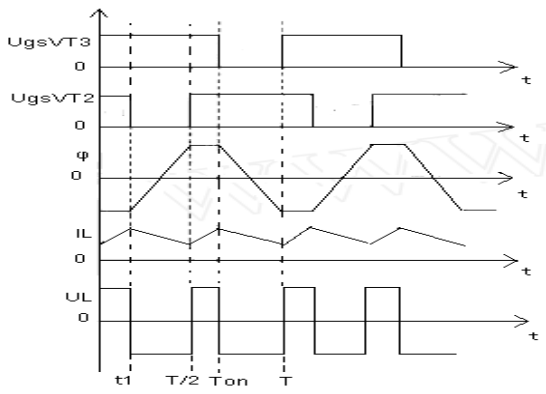


Figure4. Boost working waveform

Fig. 4 shows how it works in a switching cycle T.

1） 0 ~ t1 stage

At time 0, VT3 and VT2 turning, transformer primary being in short-circuit condition, the inductor current flow through W2, VT3, and W1, VT2 is turning at the same time due to the coupling end of W2 and W1 of the non-coupling side flows to the coil, when the synthesis of the two primary coil magnetic potential is zero, core magnetic state of the same, no induction coil potential, push-pull input voltage Uin added to the L3, the inductor current rise, inductive energy storage, the inductor current reaches the maximum when t1 while output voltage decline.

2） t1 ~ T/2 stage

At time t1, VT2, the inductor current constitute a loop through W2, VT3, then the core is magnetized, induction EMF Ew2 generated by W2 coupling is positive, as a result, VD4 and VD5 is turning and the output voltage rise.

3） T/2 ~ Ton stage

At time T/2, VT2 and VT3 still turning, transformer primary is short circuit and inductance store energy, the working state of this period is similar to the 0 ~ t1 stage.

4） Ton ~ T stage

At time Ton, cutoff VT3 but VT2 still turn on, the inductor current flow through W1, VT2 , the core is demagnetized, inductor current down, the working state of this period is similar to the t1 ~ T / 2 stage.

By V second of the inductor in a half cycle points to zero, the circuit's input-output relationship are as follows:

$$\begin{cases} U_{in}t1+(U_{in}-\dfrac{W2}{W31+W32}U_{in})(1-D)T=0 \\ \dfrac{U_{out}}{U_{in}}=\dfrac{W31+W32}{W2}\bullet\dfrac{1}{2(1-D)} \end{cases}$$

D is VT2 and VT3's duty cycle.

D. Inverter topology options

Common topologies of inverter are full bridge and half bridge. Full-bridge inverter circuit characteristics are suitable for high-power, high-voltage input places, while it also has the advantage of high DC voltage utilization. Considering a higher utilization ratio of DC voltage, so this system uses the full-bridge inverter circuit. As a classic inverter circuit, the principle of the full-bridge inverter circuit had detailed in a number of specialized books, so this paper will not repeat them.

In addition, the capacitor C5 as the connection with the inverter and DC step-up transformation (Dclink link), whose main role is to maintain the input voltage of inverter is 380V approximately, in order to ensure the output is 220V/50HZ AC.

III. CONTROL STRATEGY

The topology mainly included two parts, the former DC-DC converters and the latter DC-AC inverter, the part of the DC-DC including CUK transform and step-up transform while the CUK part makes maximum power point tracking (MPPT). A detailed analysis of two-tier control theory is as follows:

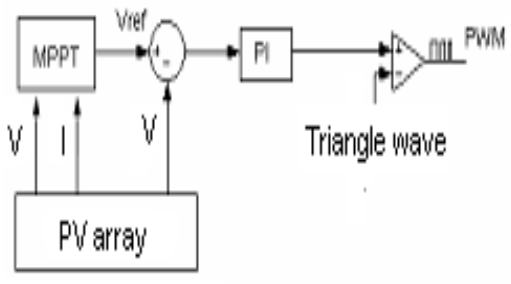


Figure5. DC-DC control flow chart

The flow chart of CUK converter control section is shown in Fig. 5. The A / D sampling by the solar array output voltage and current compare with previous voltage and current then according to MPPT control algorithm get the reference voltage at optimum operating point, then subtract the reference voltage and output voltage from A / D sampling of solar array, let the output through a proportional integral part, and then compare with the output and a fixed frequency triangle wave to get a PWM control signals, and finally the PWM control signals control CUK converter switch state through the drive circuit; while the step-up change is also controlled by adjusting the duty cycle of switch to achieve, so it is not discussed here.

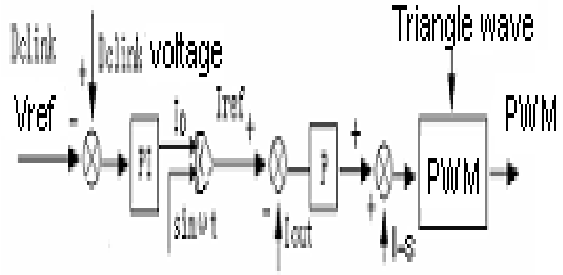


Figure6. DC-AC control flow chart

Fig. 6 is for DC - AC inverter control part of control chart. DC-AC tracking control DClink to maintain a constant output voltage, output voltage of DClink set here constant value is 380V. Subtract Dclink reference voltage and Dclink actual voltage get from the A / D sampling, and then get through a proportional integral part get DC-AC inverter output current amplitude I0 , and then inverter output current vector Iref will be obtained by multiplying a given reference unit sinusoidal signal sinωt and I0, then subtract the Iref and Iout from A / D sampling ,let the output through a proportional integral part, then add to the given reference voltage signal Vs, then compare with the output and a fixed frequency triangle wave to get a PWM control signals, and finally the PWM control signals control the DC-AC inverter switch state through the drive circuit, as in [6-8].

IV. SIMULATION AND EXPERIMENTAL VERIFICATION

In order to verify the proposed topology and control strategy is effective, according to Fig. 2 the main circuit topology and Fig. 5 and Fig. 6 Control system block diagram ,we can simulate and experimental study the system.

Experimental test system is shown in Fig. 7, solar cell array consists of eight 50W polysilicon solar array in series, the open circuit voltage of which is about 170V or so, and it's rated input power is 400W. On the input side , we use an ammeter and a voltmeter to measure the input voltage and current of solar cell; on the output side we use FLUKE 43B power quality analyzer to detect parameters and waveforms of the output AC voltage and current of the inverter. Since the output AC current value

is too small, we adopt the measurement with the current probe around 8 turns.

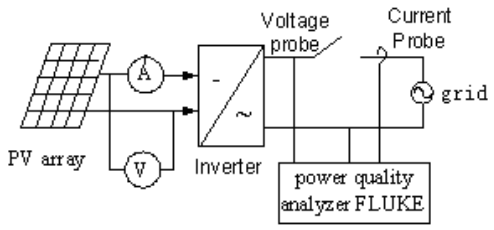


Figure7. Sketch map of testing

At 11:00 am the inverter output experiment waveform:

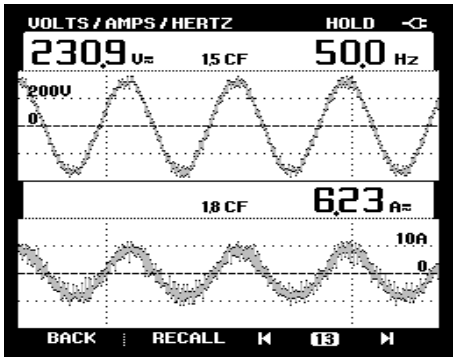


Figure8. Output voltage and current waveforms

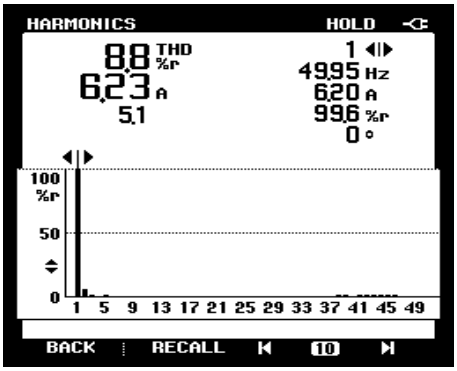


Figure9. Stand-alone photovoltaic power generation system

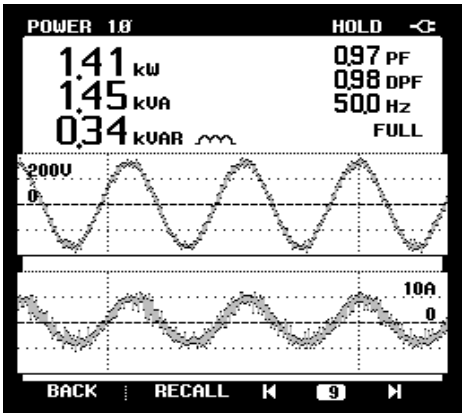


Figure10. Harmonic analysis of output current

Test results are that the inverter's input voltage is 122V and input current is 2.0A, and its input power is:

$$P_{in} = 122 \times 2 = 244W$$

Fig. 8, Fig. 9 and Fig. 10 respectively are the output voltage and current waveforms, the output current harmonic analysis of maps and the power factor. Seen from the Fig. 8, the inverter's output voltage is 230.9V,output current is 6.23/8A and output power is:

$$P_{out} = \frac{1.45 \times 10^3}{8} = 181.2W$$

Therefore, the inverter efficiency is:

$$\eta_1 = P_{in} / P_{out} = 181.2 / 244 = 0.74$$

As can be seen from Fig. 9 and Fig. 10, inverter output power factor is 0.97 and the fundamental component of output current take the total current 99.6%. It can be said inverter output power quality is satisfactory.

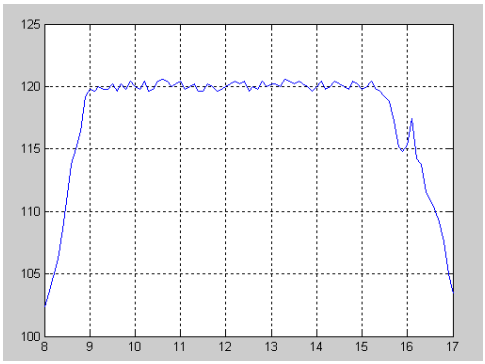


Figure11. Changing course of PV voltage with MPPT control

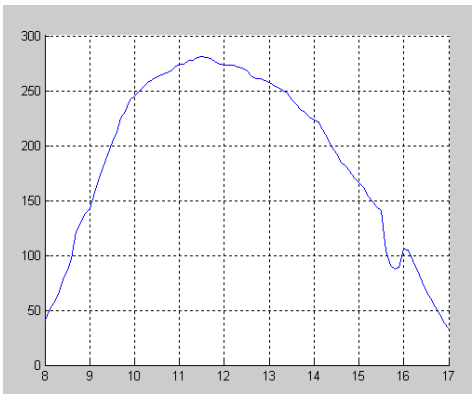


Figure12. Changing course of PV power with MPPT control

MPPT control results are shown in Fig. 11 and Fig. 12. Fig. 11 shows the voltage tracking control results of the solar array at the control of MPPT from 8 am to 17 pm .As is shown in the chart, most of the time the best operating point voltage of the solar cell has no obviously change. Fig. 12 shows the tracking control results of output power of the solar array. We can be see that the output power varies equably with time (light intensity) , and the maximum output power occurs at about 12:00.

Comparing the maximum power tracking control results described above with the measurement data of output characteristics of solar arrays to, the results shows that the output power of the solar array with MPPT control and the maximum power voltage by measuring volt-ampere characteristics of solar arrays is equal basically. we use FLUKE 43B power quality analyzer to

detect parameters and waveforms of the output AC voltage and current of the inverter. And the result shows that the output power quality is good.

Ⅴ. CONCLUSION

This paper presents a new type of push-pull output independent PV power circuit topology, which uses SPWM modulation and closed loop control strategies. Through modeling and simulation and experimental study, the results have shown that the theoretical analysis is correct, and verified that we can get the 220V/50HZ stable power by this method.

ACKNOWLEDGMENT

The authors would like to express their gratitude to the managers and teachers of Department of Computer Science and Technology for kindly supporting this research. The research are sponsored by Henan Province key scientific and technological project (082102240008), Doctor Fund of Henan Polytechnic University (648193) and Young Backbone Teachers Fund of Henan Polytechnic University (649093).

REFERENCES

[1] LI Wei, “Present situation and developing tendency of solar energy industry in China [J] ,”.Chinese Journal of Power Sources，2009，(8)：742-743.

[2] Chen Jinmei and Chen Lan, “Study of Photovoltaic Maximum power point tracking Techniques[J],” Science Technology and Engineering, 2009, (17): 4940－4945.

[3] Bai Lianping and Bai Shi, “Design of Solar Controller with Maximum Power Tracking [J],” Electrical Technology, 2009, No.8 pp: 104-108.

[4] Xiao Peng,Chen Chengguo and Wu Chunhua, “A new type of solar independent power system[J] ,”.Electric Transmission, 2008, 38 (8)

[5] Ma Lan,Qian Li and Xiao Lang, “Current Push-Pull two-way Full Bridge Converter [J] ,”.Power Electronics，2008，42（1）

[6] Wang Feng, Zhang Miao and Hu Xiaowei, “New Cuk circuit and its application in photovoltaic systems [J] ,” Power Electronics，2009，43 (.5)

[7] Yang Haizhu, Jin Xinming and Liu Jie, “500 Photovoltaic Inverter design[J].,” Electronic Elements，2006, (3)

[8] Tang Youhuai, Zhang Haitao, Luo Shan and Jiang Zhe, “Study of a new stand-alone PV system inverter [J] ,” Practical World，2008，27 (4)

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 066-069

A Model for Uncertainty Interval Matrix of Security Assessment

Bing Xia, FengJun Miao, Qiusheng Zheng
School of Computer Science, Zhongyuan University of Technology, Zhengzhou, China
Zhengzhou Key Lab of Computer Network Security Assessment, Zhengzhou, China
Email: xiabing@zzti.edu.cn

***Abstract*—Interval matrix based on AHP is typical security assessment model and wide utilization in all fields. How to generate interval reasonable and confidence interval is scarcely. With the help of expectation and entropy of the backward cloud to construct a certain confidence interval matrix [Ex-3En, Ex+3En], the paper proposed a model of fuzzy comprehensive evaluation matrix with confidence interval. This model has been applied to our design security assessment tools. Application shows that the model is effective generation interval matrix, and can avoid the uncertainty of interval matrix and theoretical proof that interval matrix with the 99.74% confidence.**

***Index Terms*—Security Assessment, AHP, interval matrix, confidence interval, backward cloud**

I. INTRODUCTION

AHP (Analytic Hierarchy Process) is mainly qualitative and quantitative evaluation model for multi-objective, multi-criteria security assessment, not only to provide a simple and effective decision making, but also is currently the major traditional security/risk assessment methods [1]. Traditional AHP method is concerned with the experts, in the judge matrix generation process, uncertainly of expert parameter seriously and expert subjective assessment affected the accuracy of the conclusions, credibility of the results. Scholars to carry out a series of studies on shortage of judge matrix. [2, 3] put forward interval matrix to improve Matrix defects caused by human factors. [4] developed seven kinds of evaluation criteria to increase objectivity of judge matrix. [5] using the accumulation factor gives the weight of each index, from a purely quantitative point of view sort the results are given, better to avoid bias caused by subjective factors. With the help of fuzzy assessment method is good at handling imprecise and ambiguous information, [6] proposed information security risk assessment method based on AHP and fuzzy comprehensive evaluation. [7] pointed out that the judgments given by experts to determine the value of the table is not linear, but the relative importance of any two interval-valued form which on this basis by Sugihara, Maeda and Tanaka's interval model.

Although the AHP has done a lot to improve, but the establishment of interval matrix is still constrained by the expert subjectivity. In short, there are still plenty of shortcomings in AHP and its improved algorithm.

1) Credible of interval matrix. Most of the algorithms used nonlinear interval matrix to describe the uncertainty of expert, but in how to generate interval matrix and interval matrix on the credibility of quantitative research does not give the corresponding results.

2) The random and fuzzy of assessment parameters. Although the above algorithm to a certain extent make up for shortage of AHP, but judge matrix parameters and the interval matrix is still man-given, there is still considerable ambiguity and randomness, and thus influence the calculation results.

How to resolve the uncertainty caused by experts assessment and how to generate a credibility interval matrix, based on backward cloud , this paper proposed a confidence interval of fuzzy comprehensive evaluation matrix model to achieve above two shortage.

II. SCHEME

A. Concept

Cloud [8] is an effective tool in uncertain transforming between qualitative concepts and their quantitative expressions.

Cloud , set U is a value space that expressed in precise quantitative, $X \subseteq U$, T is qualitative spatial concepts on space U, if the element x ($x \subseteq X$) there exists a stable tendency of random numbers $C_T(x) \in [0,1]$, called the X is degree of membership on T, denote $C_T(x): U \to [0,1], \forall x \in X(X \subseteq U), x \to C_T(x)$, the concept of T from the space U to the [0,1] mapping the distribution of the data interval, called the cloud.

Cloud (Ex, En, He) is one-dimensional cloud which representation with the expected value *Ex*, entropy *En*, hyper entropy *He*, reflects the qualitative features of the concept of quantitative. Backward cloud is an algorithm which can convert qualitative concepts into quantitative values.

confidence interval, set $F(x,\theta)$ is the distribution function of X, set θ is a location parameter on X, $\theta \in \Theta$ (Θ is all possible values), for a given value $\alpha(0 < \alpha < 1)$, $\underline{\theta} = \underline{\theta}(X_1, X_2, \cdots, X_n)$ and $\bar{\theta} = \bar{\theta}(X_1, X_2, \cdots, X_n)(\underline{\theta} < \bar{\theta})$ are two sample of X , if the two statistics meet $P\{\underline{\theta}(X_1, X_2, \cdots, X_n) < \theta < \bar{\theta}(X_1, X_2, \cdots, X_n)\} \geq 1-\alpha$, so $[\underline{\theta}, \bar{\theta}]$ is interval

This research was supported by the Key Technologies R&D Program of Henan Province of China（No. 092102310038、092102210029）.

which the confidence level is $1-\alpha$,for short confidence interval.

Confidence interval remedies for the accuracy of deficiencies that concept of a single qualitative convert into the concept of a single quantitative. From the point of view to judge, if a given parameter is a confidence interval on the parameters, judge matrix is made more objective and credible. From the number of cloud features known that the concept of entropy reflects the uncertainty of qualitative, which is a range of acceptable size that qualitative concept can be described in domain space. Statistics show that interval $[E_x-3E_n,E_x+3E_n]$ is best description to qualitative concepts, in this paper also uses $[E_x-3E_n,E_x+3E_n]$ as confidence interval.

B. Idea and soluation

Use all experts assessment parameter as cloud droplets, based on backward cloud generator; we can get three figures to describe the characteristics of clouds. Then use $[E_x-3E_n,E_x+3E_n]$ we can get a confidence interval .Based on confidence interval, and use fuzzy comprehensive assessment model, we can get a judge evaluation interval matrix. Make this judge parameter is not a linear, certainty but reflects range of qualitative concept and nonlinear interval parameters.

After get interval comparison matrix, we should converter interval comparison matrix to the general comparison matrix. Use paper [2] method, this paper proposed a digital approximation program to create judge matrix. At the same time, if judge matrix consistence can not be satisfy ,then need two strategies to adjust matrix with maximum value of ***CI***(Consistency Index) based on single order and total order.

Therefore, based on confidence interval of fuzzy comprehensive assessment matrix model, digital approximation program and automatically adjust the Matrix program together constitute the qualitative and quantitive model. Show in Figure 1.

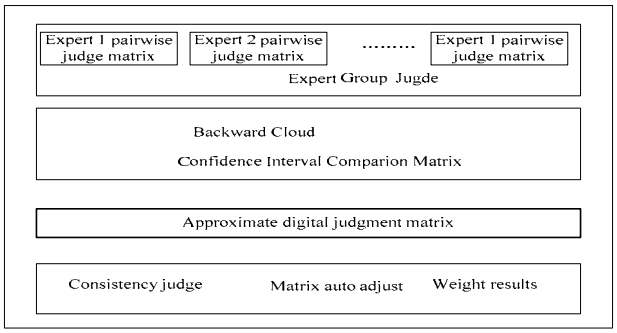


Figure1：Backward cloud assessment model

C. Fuzzy Comprehensive Assessment Matrix Model

In the risk assessment based on backward cloud algorithm, how to generate fuzzy comprehensive assessment matrix with confidence interval is a key algorithm of the model. At the comparison matrix is constructed, the first thing is to collect experts, administrators, technicians judge parameters based on two objects comparison. Second, human factors in order to minimize the uncertainty, using reverse cloud generator for Cloud (*Ex*, *En*, *He*). Then introduce confidence intervals to enhance the credibility of the

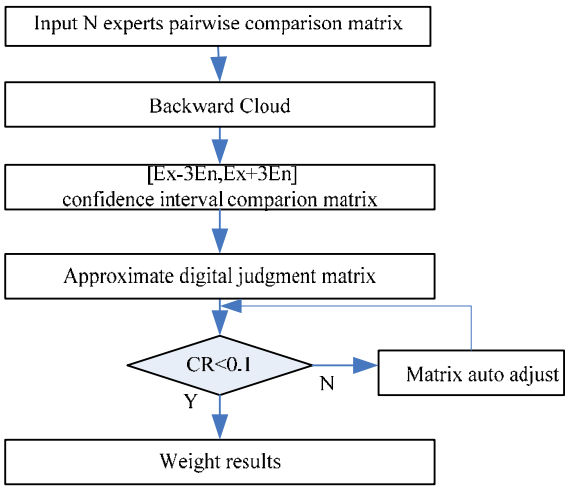


Figure 2. Backward cloud assessment model .

judge matrix. As shown in Figure 2.

Based on interval number of fuzzy comprehensive assessment matrix model, making quantitative assessment of experts on indicators is not a specific score, but the non-linear range interval, such as the form $a_{ij}=[l_{ij},u_{ij}]$. In this solution, compared parameters falls within the range of *99.73%* confidence level interval, namely: $a_{ij}=[l_{ij},u_{ij}]=[E_x-3E_n,E_x+3E_n]$.Algorithm flow is as follows:

Algorithm, Fuzzy Comprehensive Assessment matrix generate algorithm based on Confidence intervals

Input *N* experts have given judge matrix samples

Output a confidence interval of the fuzzy comprehensive assessment matrix

1) First obtain ***N*** experts' judge matrix samples $x_i \quad (i=1,2,...n)$;

2) Use corresponding parameters sample as cloud droplets, and calculate average value uses formula (1)

$$E_x = \bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad (1)$$

The value is ***Ex*** that expectation of convert qualitative concepts into quantitative values.

3) Calculate first order central moment and the sample variance use formula (*2*), formula (*3*).

$$B = \frac{1}{N}\sum_{i=1}^{N}\left|x_i-\bar{x}\right| \qquad (2)$$

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x})^2 \qquad (3)$$

4) The calculation reflects the fuzziness and randomness of entropy uses formula (*4*).

$$E_n=(\pi/2)^2\times B \qquad (4)$$

5) Calculate entropy uses formula (*5*).

$$H_e = (S^2 - E_n^2)^{1/2} \quad (5)$$

6) Use $a_{ij} = [l_{ij}, u_{ij}] = [E_x - 3E_n, E_x + 3E_n]$ generates interval matrix parameters. .

So, based on backward cloud, we can generate a confidence interval of the fuzzy comprehensive assessment matrix.

III. APPLICATION AND PROOF

A. Application

Zhengzhou Key Laboratory of Computer Network Security Assessment proposed a multi-level data fusion and analysis of the hierarchical assessment model, from 11 categories and 36 items, to assess host security situation [9]. To the common service as an example, using this algorithm to construct fuzzy comprehensive evaluation matrix based on confidence interval. Suppose three experts give follow judge matrix based on the impact of the service version, service hole and security configuration three aspects, As follows:

$$\begin{pmatrix} 1 & 1/5 & 1/3 \\ 5 & 1 & 3 \\ 3 & 1/3 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1/4 & 1/2 \\ 4 & 1 & 4 \\ 2 & 1/4 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1/4 & 1/4 \\ 4 & 1 & 3 \\ 4 & 1/3 & 1 \end{pmatrix}$$

Though above algorithm, we can obtain *En* matrix and *Ex* matrix. As follows:

$$\text{Ex}=\begin{pmatrix} 1 & 0.233 & 0.361 \\ 4.333 & 1 & 3.333 \\ 3 & 0.305 & 1 \end{pmatrix} \qquad \text{En}=\begin{pmatrix} 0 & 0.028 & 0.116 \\ 0.557 & 0 & 0.557 \\ 0.835 & 0.046 & 0 \end{pmatrix}$$

Based on *Ex* and *En*, using $[E_x - 3E_n, E_x + 3E_n]$ interval, we can construct a fuzzy comprehensive evaluation matrix with confidence interval, as follows:

$$\begin{pmatrix} [1\ 1] & [0.150\ 0.317] & [0.013\ 0.709] \\ [2.663\ 6.004] & [1\ 1] & [1.663\ \ 5.004] \\ [0.494\ \ 5.506] & [0.167\ \ 0.444] & [1\ 1] \end{pmatrix}$$

Then with reference [3] to make a consistently approximate general digital judgment matrix, as follows:

$$\begin{pmatrix} 1 & 0.121 & 0.359 \\ 8.23 & 1 & 2.95 \\ 2.79 & 0.339 & 1 \end{pmatrix}$$

Finally, uses AHP process ,do not adjust matrix as *CR= 1.32009E-8 <0.1*, we gain approximate weight of every elements on common service as follows table 1：

Table1. shows the weight result

common service weight distribution		
version	hole	configuration
0.083	0.685	0.232

Examples show distribution results are conform to reality. So, the algorithm can effectively generate an interval matrix with certain degree possible, and can reflect uncertainty degree of expert parameter subjective.

B. Proof

Set $(1-\alpha)$ as an uncertainty degree of confidence, due to cloud obey normal distribution of $N(E_x, (\sqrt{n}E_n)^2)$, so with the central limit theorem can be obtained:

$$\lim p\left\{\left|\frac{\bar{X} - E_x}{E_n}\right| \le Z_{\frac{\alpha}{2}}\right\} = \frac{1}{\sqrt{2\pi}}\int_{-Z\%}^{Z\%} e^{-\frac{t^2}{2}} dt = 1 - \alpha$$

E_x , where $\bar{X}$ is the unbiased estimate, and there is $\frac{\bar{X} - E_x}{E_n} \sim N(0,1)$. The $\frac{\bar{X} - E_x}{E_n}$ obey distribution $N(0,1)$ does not depend on any unknown number, according to the definition of the standard normal distribution $1-\alpha$ as follow:

$$p\left\{\left|\frac{\bar{X} - E_x}{E_n}\right| \le Z_{\frac{\alpha}{2}}\right\} = 1 - \alpha$$

Where $Z_{\frac{\alpha}{2}}$ is the standard normal distribution of the bilateral point position.

So we get a confidence interval $[\bar{X} - Z_{\frac{\alpha}{2}}E_n, \bar{X} + Z_{\frac{\alpha}{2}}E_n]$ of E_x with $1-\alpha$ confidence level, in order to facilitate data processing, making $Z_{\frac{\alpha}{2}}$ as integer value, by the look-up table, we can get $Z_{\frac{\alpha}{2}} = Z_{0.0026} = 3$. So get a confidence interval with *0.9974* confidence level. Also because $\bar{X}$ is an unbiased estimate of E_x , the interval can be changed $[E_x - 3E_n, E_x + 3E_n]$.Proof is completed.

IV.CONCLUSION

Due to subjective of human structure matrix parameters and thinking of expert's random, resulting in assessing parameter uncertainty. Therefore, this paper by the help of expert group judge and cloud theory, from confidence interval view, proposed a cloud-based inverse matrix generation program, and gives the key to confidence intervals based on fuzzy comprehensive evaluation matrix generation algorithm. Greatest contribution of this Paper is that how to generate interval comparison matrix with a degree of confidence .Application and proof show that this program is feasible, and the interval matrix with *99.73%* confidence level, thus better solve the credibility problem of interval matrix.

REFERENCES

[1] Bing Xia,Lei Pan, Feixian Sun,Assessment model based on multivariate data fusion and hierarchical analysis.Computer Engineering, vol.36(No.10), May 2010. *(in chinese)*

[2] Baiyuan LONG, Dong WANG, Dong-qing XIE, et al.Hierarchy risk evaluation method based on interval judgment matrix.Computer Engineering and Applications, Vol 44,pp.127-130,October 2008(In Chinese).

[3] Yuzhong Zhang,Cuiping Wei.On the properties and priprity method of a consistent interval comparison matrix.OR Transactions.Vol.11(No.3).pp.113-120,Sep,2007(In Chinese)

[4] Yan Zhu, YongTian Yang, YuQing Zhang , DengGuo Feng.Research on Information Security Evaluation Model Based on Hierarchy Structure.Computer Engineering and Applications. Vol 40,pp.40-43, June 2004(In Chinese)

[5] FangGe LI , JiYun BAI , HongJie ZHAO.The Study of Solving Weight by AHP's Accumulation Factor Sequence Evaluating Data.Operations Research and Management Science. Vol 14,pp. 60-63,June 2006(In Chinese).

[6] Long Xiao,Yong Qi,Qianmu Li.Information security risk assessment based on AHP and fuzzy comprehensive evaluation.Computer Engineering and Applications .Vol.45(No.22):pp82-85,2009. (In Chinese).

[7] Fei Cheng,Jian Luo.Improved uncertain type of AHP algorithms.Joural of Xiamen Univesity(Natural Science).Vol.45(No.2):pp186-190.Mar,2006. (In Chinese).

[8] DeYi Li, ChangYu Liu, Yi Du, Xu Han. Artificial intelligence with uncertainty.Journal of Software,Vol 11, p1583-1594, November 2004.

[9] Bing Xia,Fei Pei,Qiusheng Zheng.Research of Implement of Security Assessment system based on Policy and Management.Joural of Zhongyuan University of Technology.Vol.20(No.6):pp29-34,Dec,2009.(In Chinaese).

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 070-073

Performance Test and Optimization Study of High Performance Parallel Cluster System

Li-hong Wang[1], Wei Wu[2]
[1]Department of computer science and Technology, Henan Polytechnic University, Jiaozuo, China
Email: wlh@hpu.edu.cn
[2]Moder Education Technology Center, Henan Polytechnic University, Jiaozuo, China
Email: win@hpu.edu.cn

***Abstract*—This article studies the performance optimization of the large-scale cluster system by the performance test for Dawning Tiankuo high performance cluster system. The efficiency of this system is exhibited through the test and analysis for this cluster system by running the test software in different parallel environment. Results prove that the high performance computer cluster has acceleration function and stability. These results offer a foundation for the exploiture and study.**

***Index Terms*—-performance, PCG, parallel computing, MPI, Open MP, PGI**

I. INTRODUCTION

With the development of scientific, research, researching objects are becoming ever more complicated. Especially in that numerical simulation, because of researching model becoming more complicated, emulation of data and calculating quantity is becoming more and more. In the face of enormous calculated load, make use of cluster system and parallelism resolving the calculating time is a ideal schemes [1].

For large-scale numerical calculation in the project and shorten scientific period, we purchase Dawning series large-scale parallel calculate cluster system. For testing work calculating capacity of this system, implementing frequently-used large scale linear equations parallel algorithm, testing and analyzing the calculated time of different parallel models and its functions, adopting optimize measure.

II. SYSTEM HARDWARE ARCHITECTURE

The function of computer system lies on the distribution of its software and hardware. Before target-oriented optimizing system, we need understand the composition of computer's software and hardware.

A Distribution of Hardware

At present there are many kinds of large scale parallel computer system, many Dawning high performance computer have sky-high calculating speed. but these computer hardware are expensive, managing and maintenance cost much, ordinary institution is hard to bear all the costs. With the development of network technique, many computers are connected together through network, the methods of parallel calculation consisting with cluster system are becoming a development tendency of parallel computer. There are two ways to Construct cluster system :one is to make use of LAN technology connecting many computers and servers together, this way may make the best of computing resource in LAN (Local Area Network) and decreasing cluster system cost, Another way is to make use of professional Net and connect server optimized to cluster system. The cost of this way is too high, but compare with before, it is possessed of higher computational efficiency and larger communications network and lower time delay of data communication and more stable system service. Dawning Tiankuo is a kind of high-end product in Tiankuo series Server. Its VLAN (Virtual Local Area Network) configuration is shown in table I.

This system has 35 compute nodes in number including two I/O nodes, and its theoretical compute peak value of floating-point arithmetic can come up to 2.9 trillion times per second, Internal memory of this system is up to 560GB, total capacity of storage is up to 10 TB.

B Software Platform

Operating system of every node in this system adopt Suse Enterprise Server 10, making use of Gridview 2.0 serve as job scheduling system, and in order to carry on convenient scientific calculation, this system install current math library for example ACML、LAPACK、ScaLAPACK、BLAS、GOTO、Atlas、FFTW and so on. In order to improve performance about concurrent compiler, this system install the compiler as Intel C++、Fortran etc. In addition this system also install current frequently-used parallel environment in parallel computing fields like OpenMp and MPI.

TABLE I. CLUSTER SYSTEM HARDWARE CONFIGURATION

Node type	Node	CPU type	CPU numbers	memory	Hard disk
compute node	30	Intel Xeon E5 530 four nuclear 64bit processor	2	16G	146GB 15000R PM hot plugSAS
management node	1	Intel Xeon E5 530 four nuclear 64bit processor	2	16G	146GB 15000R PM hot plugSAS
SMP big node	2	Intel Xeon 74 40 four nuclear 64bit processor	2	32G	300GB×2 15000 RPM hot plugSAS

III. PARALLEL COMPUTING CAPABILITY JUDGE STANDARD

Want to test the capability of the cluster system, it needs to judge standard. Universal method is to record response time α of running program in different environment and compute the speed-up ratio S_p and parallel efficiency E_p and obtain a testing result.

A Response Time

The main standard of testing computing capability is time. Response time is also called turnaround time, it means spending the full time about completing a task. Its formula is

$$\alpha = t_c + t_{I/o} \tag{1}$$

In this formula, t_c is the time about a program of CPU. It contains user CPU t_e executive time and system CPU time t_s operating system's spending. $t_{I/o}$ is I/O time of system, it contains the I/O time of input /output unit and the exchange time about the page of auxiliary memory and main memory as well as the time of internetwork communication. Because of continually exchanging data between different nodes in the process of executing the task of cluster system, internetwork communication usually is the most important factor in program efficiency. For this reason, if it wants to cut down the response time, it must cut down to the utmost the time of communication and to use to the greatest advantage of CPU.

B Speed-Up Ratio S_p and Parallel Efficiency E_p

Speed-up ratio S_p is the main testing target about efficiency of parallel processing system. it formula is

$$S_p = \alpha_s / \alpha_p \tag{2}$$

In this formula, α_s is the response time about giving program on the single processor, α_p is the response time about the same program on parallel operating system containing the many processors. Speed-up ratio S_p reflect the speedup times being obtained by computing speed on parallel. In the case of theory, if program are executed completely, the same p processors can reach Speed-up ratio p, but in the case of reality, Speed-up ratio S_p usually less than p [2][3].

In order to reflecting the parallel efficiency, define and come into being parallel efficiency E_p, the formula as below

$$E_p = S_p / p \tag{3}$$

As a general rule, range of E_p value is 0 to 1. If E_p is nearer to 1, the parallel efficiency of algorithm is much higher.

IV. PARALLEL TESTING PROGRAM

For effectively testing the efficiency of parallel system and can choice typical testing program. Because of most scientific calculation as finite element calculation, finite difference calculation and so on, they need to solve large linear equations, making use of solving the large linear equation is the wonderful representativeness. It adopt basing on domain decomposition parallel preprocessing conjugate gradient method to solve linear equation set [4][5].

Frequently-used parallel schema have three kinds about MPI, OpenMp and MPI+OpenMP. MPI is a distributed storage model basing on message passing and has favorable communication capability; OpenMP is a memory parallel storage model basing on sharing and can make best of computing resource of single node many computing core; MPI+OpenMP is a blended parallel model and can achieve two-stage of distributed and memory sharing model [6].

Preconditioned Conjugate Gradient method (PCG) is a well-rounded and higher parallel efficiency iteration arithmetic solving linear equation set. So PCG is a basic arithmetic about solving linear equation set, the parallel arithmetic of solving equation about $Ax = f$ and its arithmetic describe as follows.

Step1: initialization:

$$x^{(i,0)} = 0, r^{(i,0)} = f^{(i)} \tag{4}$$

$$M^{(i)} = \varepsilon \sum_{j \in \phi} m^{(j)} \tag{5}$$

Step2:

$$z^{(i,0)} = (M^{(i)})^{-1} r^{(i,0)}, s^{(i,0)} = \varepsilon \sum_{j \in \phi} z^{(j,0)} \tag{6}$$

$$\alpha_1^{(i,0)} = r^{(i,0)}.s^{(i,0)}, \alpha_1^{(1)} = \sum_{i \in \Omega} \alpha_1^{(i,0)} \tag{7}$$

$$\beta_2^{(1)} = \alpha_1^{(1)}, p^{(i,1)} = s^{(i,0)} \tag{8}$$

Step3: For the k time iteration (k =1,2,3…)

$$u^{(1,k)} = A^{(i)} p^{(i,k)} \tag{9}$$

$$\alpha_2^{(i,k)} = p^{(i,k)} u^{(i,k)},$$

$$\alpha_2^{(k)} = \sum_{i \in \Omega} \alpha_2^{(i,0)}, \tag{10}$$

$$\alpha = \alpha_1^{(k)} / \alpha_2^{(k)}$$

$$x^{(i,k)} = x^{(i,k-1)} + \alpha p^{(i,k)}, \; r^{(i,k)} = r^{(i,k-1)} - \alpha u^{(i,k)} \tag{11}$$

$$z^{(i,k)} = (M^{(i)})^{-1} r^{(i,k)}, \; s^{(i,k)} = \varepsilon \sum_{j \in \phi} z^{(j,k)} \tag{12}$$

$$\beta_1^{(i,k)} = p^{(i,k)} s^{(i,k)}, \beta_1^{(k)} = \sum_{i \in \Omega} \beta_1^{(i,k)} \tag{13}$$

If $\beta_1^{(k)} < \varepsilon \beta_1^{(0)}$, iteration is over and input x, ε is a quantity controlling residual error.

Otherwise:

$$\beta \beta_1^{(k)} \beta_2^{(k)} p^{(i,k+1)} = s^{(i,k)} + \beta p^{(i,k)} \tag{14}$$

$$\beta_2^{(k+1)} = \beta_1^{(k)} \alpha_1^{(k+1)} = \beta_1^k \tag{15}$$

In case $k = k + 1$, repeat step 3.

In this formula: the right mark (i,k) of Variable stand for the k times iteration of the i proceeding; The sign of $\varepsilon\sum_{j\in\phi}$ show boundary local communication summation; The sign of $\sum_{i\in\Omega}$ show global communication summation;

For insure that coefficient array of linear equations is nonsingular, it contains package process of coefficient array in the program.

V. Parallel Cluster System Test and Optimization

A Test Scheme

MPI parallel model need input some configuration item when program is running and allocate corresponding environment, so then realize normal deserialize. In which "-np" option behind number appoint scale of program parallel, that is how much proceeding this system contain; "-machinefile" option behind file appoint a configuration file used to allocation system calculate resource, this file contains calculate node where every process is going to be running. The same program can be tested in different environment through smoothly making use of two kinds of configuration properties [8][9].

Beneath MPI+OpenMP mixed parallel model, if parallel on 2n calculate unit and use "-machinefile" collocate n nodes, every node using two threads put in to effect parallel calculate.

For knowing the specific properties of this high performance parallel computer platform at a different angle, test program is compiled to four kinds of versions. Every version specific compile condition is shown table II .

B Performance Test

To length running time and adopt different order of coefficient matrix to solve as testing, f of right side of equation $Ax = f$ is divide into ten times to load and contrast their performance in different condition.

(2) PGI compiler performance optimize test

The version program of Serial and Optimize Serial are running respectively, testing result is shown to table III.

(2) Sharing memory model parallel performance test

Compared running response time through adopting optimize version and OpenMP parallel version on a compute node, obtaining the histogram is shown figure1.

TABLE II. Every version this program compile option

	PGI	MPI	OpenMP
Serial	no	no	no
Optimize Serial	Yes	no	no
Pure OpenMP	yes	no	yes
Pure MPI	yes	yes	no
MPI+OpenMP	yes	yes	yes

TABLE III. PGI Optimize Contrast

	1500 order array	1800order arrays	2500order array	3500order array
Serial	60.49s	84.06s	109.21s	171.59s
Optimize Serial	23.01	31.25s	41.84s	66.25s
Running velocity ratir	2.63	2.69	2.61	2.51

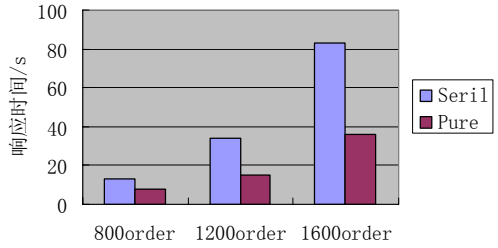


Figure 1. Single node OpenMP parallel respone contrast

(3) Contrast performance of parallel model

Making use of MPI and MPI+OpenMP mixed parallel model and respectively computing 1500order, 1800order, 2500order, 3000order and 3500order arrays on 4, 10, 20 compute units, the speed-up ratio are obtained being shown in table III .

C Analyzing parallel performance

Through the data of table III, and we can know influence of program running efficiency on the PGI compiler optimized. The efficiency of optimized program is raised 15 times over, when keeping watch on system process, we can obviously find that program without optimizing engrosses 50% resources of CPU and optimized program engrosses 100% resources of CPU, it can be seen from this that PGI can Optimize running codes and improve availability of CPU.

Through figure 1, we can see that OpenMP can dig computing potential of many core CPU and enhance running speed.

Through figure 2, we can see that parallel program can enormously enhance computing speed. As nodes reduce to pure and simple MPI parallel, program can obtain speed-up ratio near node number and make best use of computing resource of many modes, but joining OpenMP mixed programming model, speed efficiency of parallel computing is weaker than pure MPI. When node number is proved to 20 nodes over, efficiency of MPI parallel program gradual decline but mixed model program is gradually advancing. Thus it is clear that node number is less and MPI communication burden is less, efficiency of MPI is better than mixed method, but node is more and MPI communication burden is more, using mixed method again, making use of OpenMP to decrease the number of node, reducing communication burden between nodes and make use of thread to advance efficiency go a step.

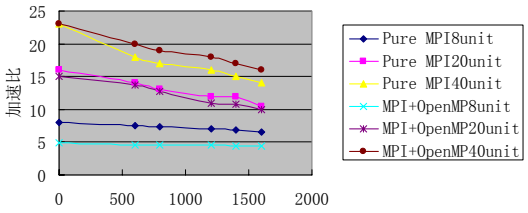


Figure 2. Two kinds of parallel model response time contrast t

VI. TAG

Parallel testing program is developed based on parallel preprocessing conjugate grad arithmetic and on different parallel model, and test parallel efficiency about high performance parallel cluster ,and study optimized method of high performance, parallel calculate test result comes up to fact, it takes on reference to some extent .

Through testing, we can get below some understands about high performance computer cluster:

(1) Speed-up radio of cluster and node numbers of calculates are kept with regulation relations, that is that speed-up radio increased with node numbers increasing .

(2) Communication time delay is one of main factors that influence speed-up radio, data communication should be decreased between nodes in order to obtain better speed-up radio, thereby decreasing communication time to obtain better speed-up radio.

(3) AS too many compute units are used, we can use OpenMP coordinate MPI to lower communication between processes and advance the speed-up radio, but degree of advancing is not obvious.

(4) Cluster computer can shorten time of processing information in the large data processing fields and advance work efficiency go a step and possess large processing technology and method to better use at one time.

ACKNOWLEDGMENT

I would like to thank my colleagues on the HPU-HPC team for their contributions, insights, and support.

This paper is supported by the high-performance grid computing platform of Henan Polytechnic University.

REFERENCES

[1] Fengyun, Zhoushuqiu, MPI+OpenMP mixed parallel program model application research. *Computer system application*, 2006(2): 86-89.

[2] Zhuyongzhi, Libingfeng, Weironghui. Bewulf-Tcluster system high scalability research. *Computer Science*, 2008, 35(2): 298-300.

[3] Shameem Akhter, Jason Roberts. *Multi-core Programming*. Beijing: Electronic Industry Publishing House. 2007.

[4] Graham Glass, King Ables. *Linux for Programmers and users*. Bejing: Tsinghua University Publishing House. 2006.

[5] Chenguoliang. *Parallel Arithmetic Put Into Practice*. Beijing: Higher Education Publishing House. 2004.

[6] Wenxiaofei, Zhuzongbai, Hucunzhi, etc. *Efficiency Evaluate High Performance Computer Cluster*. Wuhan Science and Technology University Academic Journal: Information and Administrative Engineering Version, 2005, 27(4): 19-22.

[7] Meijerink J A. *Guidelines for the Usage of Incomplete Decomposition in Solving sets of Linear Eqution as they Occur in Practical Problems*, Comp. Phys., 1981.

[8] Majingyan, Yushuangyuan. *Based on analysis parallel environment MPI of MPICH*. Scientific and Technical Information, 2006(30): 6-7.

[9] Cao zhennan, Feng shenzhong, Wang qin. *In IA cluster node parallel program model efficiency analysis*. Computer Engineering and Application, 2004(20): 84-86.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 074-076

The Research of Coal-mining Control Configuration Software's Real-Time Database

An Weipeng[1], Li Miao[2]
[1]School of Computer Science & Technology, Henan Polytechnic University, Jiaozuo Henan, China
Email: awp@hpu.edu.cn
[2]School of Computer Science & Technology, Henan Polytechnic University, Jiaozuo Henan, China
Email: limiao2048@126.com

***Abstract*—Real-time is the core of configuration software. It is the prerequisite for the normal operation of the coal-mining control configuration. The article describes the configuration software for mining real-time database system's design and implementation methods, provides the method for using the dynamic-link library to build real-time database system, and gives the way of the data model and interface. Using this method to establish real-time database system is full of openness and versatility.**

***Index Terms*—Real-Time Database, Data Model, Dynamic Link Library**

I. INTRODUCTION

Present, there is more function which the common configuration software have. For coal enterprises, it is high degree of redundancy, expensive and the focal point which is not protrusion. So it is necessary to research the Mining configuration software. Because the real-time database is the key to the configuration software and it is a direct impact on the performance level of success or failure of the configuration software. So this article describes the real-time database system's design and implementation, which is for mining configuration software.

II. THE INTRODUCTION OF REAL-TIME DATABASE SYSTEM

Real-time database system is the core configuration software. The data Defined in the configuration software is different from the traditional data or variables, it not only contains the variable value characteristics, but also packages the data together with the data's attributes and data related to the operation method as a whole, providing services as the form of an object. The values, attributes and methods defined as one of the data call data objects. In the design, it uses the data objects to express the system in real-time data and uses the object variable to replace the traditional sense of the value of the variable. We use the database management of all data objects as a collection of real-time database.

III. THE DESIGN OF REAL-TIME DATABASE SYSTEM

A. The Storage Strategies Of Real-Time Database

Due to taking full account of the requirements of real-time system among the design of system, establishing the data storage strategies should be bases of the Different
types of data which are required to respond to the speed and the size of the amount of data. Therefore, the SQL database system, the Memory Database and the Document Management System are used to conceive the Construction of Real-Time Database System in the text.

1) Using the SQL database save those shared data which are largeness and no special requirements, meanwhile it is operated by the Interface functions which is provied by the Real-time database interface functions.

2) The memory database stores the data which access to high frequency data. As the memory access time than disk access time in multiples of the number of low-level 105-fold, So taking memory database to deal with real-time collection of data can eliminate I / O bottlenecks.The document
management system save those data which is in need of long-term preservation of shared.

3) The data,which requires long-term preservation and share, will be saved with document management systems.

B. The Functional Modules Of Real-time Database

The text uses object-oriented technology and defines the Real-time database as the form of class. The function module achieves the association with the real-time database by calling the Real-time database interface functions. The design of Real-Time Database class includes to the initialization module, the object search module, the Content modify and update module, the Alarm Modules, the Calculation and shows module and so on.

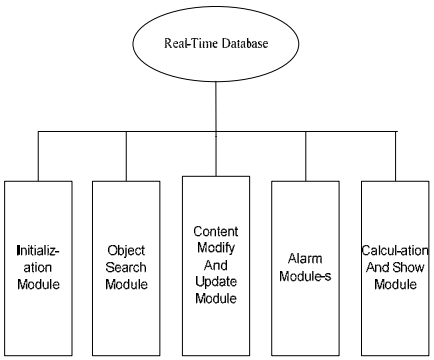


Figure 1. The Real-Time Database function module

The Real-Time Database function module

C. The Structure Of Real-time Database

The memory real-time database is full advantage of database and relational database, the Memory Database

processes on-site real-time data, viewing, alarming and analysising the data on this basis. The relational database deals with the historical data and realizes the retrieval of historical data and online analysis capabilities, the specific real-time database structure diagram is as follows:

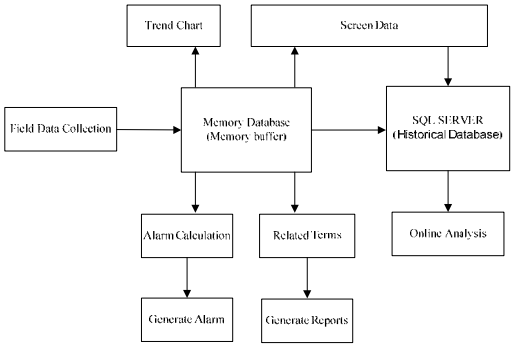


Figure 2. The Schematic diagram of real-time database

IV. THE REALIZATION OF REAL-TIME DATABASE SYSTEM

A. *Real-Time Database Model*

It is necessary of considering the characteristics of the field of coal mine monitoring in this paper, the field data of mining companies are mainly gas concentration, Carbon monoxide, Dust Concentration, temperature and Device Status so on. The Database records are base on measurement points. The unit which Real-time database stores is not only the Variable values, but also Including the variable attributes and variable operation methods. The real-time data type as the underlying class structure:

```
Class Conter
{ Public:
Cstring point_kind;
// point type
    Cstring point_name;
// point name
    Int point_index;
// dot
    Char DeviceName;
//device Name
    Bool StoreMark;
// the sing of storing in the database
    ……
  Public:
    Bool CopyToFile(point_index);
 // add a point to the file
    Bool LoadFromFile(point_index);
 //load a pint from the file
    Bool IsStore();
  //store the data
    Void Alarm(float Value, bool AlarmMark);
//alarm
};
```

B. *The Use Of DLL Build System Is Running Real-time Database*

Dynamic Link Library (DLL) is a Windows program in a special unit and is referred to as non-mission-oriented executable module, which is by the caller of the task-driven. In order to improve the system's real-time performance, the run-time and real-time database is created by DLL in this text. The Real-time database stored in a DLL-owned global memory, then it provides the interface functions to achieve the database read-write, query and management functions. This system has a comprehensive open and high Real-time.

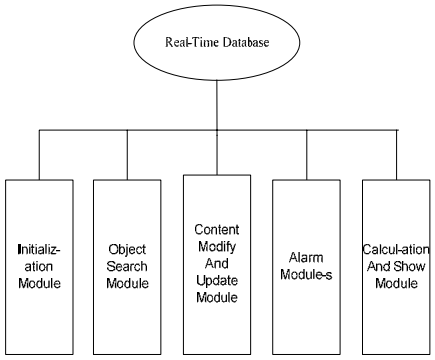


Figure 3. Real-Time Database Dynamic Link technology

The Real-Time Database function module

C. *The Interface Mechanisms For Real-time Database*

Real-time database is based on dynamic link library in the global memory, it accesses to the real-time database by a set of API interface functions which is provided by dynamic link library. The operational program is responsible for starting the dynamic link library through the system running. Other applications can access real-time database through the interface function. The real-time database interface is an open interface specification for users and it allows users to make use of the interface to directly access the database, so it provides a convenient method for users to develop input-output interface driver and the user module. The Interface type system is filled with openness and secondary development function.

TABLE I.
FUNCTION FEATURES

Function Prototype	Function Description
RTData ReadDate();	Read real-time data object values
RTData CollectData();	Get the device data collected
Bool LoadFromFile(point_index);	Read point information from the file
Int GetIndex(Cstting IndexName);	Through the data object's name to obtain the serial number

V. SUMMARY

In this paper, Coal-mining real-time database configuration software has done a detailed study of and proposed the method using the dynamic-link library to build real-time database system. Using this way establishing real-time database system has a comprehensive open. At the same time the text gives the data model and interface mechanisms for the

implementation method, which is relatively strong versatility.

REFERENCES

[1] Song E M, Kim Y K, Ryu C H, et al. No-Log Recovery Mechanism Using Stable Memory For Real-Time Main Memory Database Systems, RTCSA '99, IEEE CS, Dec. 1999

[2] Zheng Zong-han, WEI Hai-kun, LI Qi. Design and application of algorithm engine of real time database[J]. Process Automation Instumentation, 2004, 25(6): 4-7.

[3] Fang Lai-hua, WU Ai-guo, HE Yi. Research on the key technology of configuration software[J].Control and Instruments in Chemical Industry, 2004, 31(1): 33-35.

[4] Shu L C, StankovicJ. A. Achieving Bounded and Predictable Recovery Using Real-time Logging. Proceedings of the Eighth IEEE Real-Time and Embedded Technology and Applications Symposium, Sept. 2002, pages 269-285

[5] Xu Yu-ru. Real-time Data Acquiring System Research, Machinery & Electrouics, 2004

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 077-081

Design of the mine gas sensor based on Zigbee

Su Baishun, Pang Zhengduo, Meng Guoying
School of Electromechanical and Information, China University of Mining and Technology (Beijing), Beijing, China
subaishun@163.com

***Abstract*—To detect mine gas concentration effectively in the temporary working location of underground, This paper introduces a mine gas sensor based on ZigBee which is adapting to environment of coal mine underground, then gives the overall design of physical structure of system, network topology and its hardware and software designed of the composed module.**

***Index terms* — mine gas sensor;Atmega128L; CC2420; ZigBee stack**

I. INTRODUCTION

At present the mine gas concentration is detected by fixed gas sensor which mounted at fixed locations in China coal mine, and then connect to the working station through the underground cable ,at last connect to the monitoring center. With the extension of the mining face, The distance between the main roadway and the mining face can stretch to several hundred meters or several kilometers, a large number of gas emission will cause gas overrunning and abdominal mass near the mining face in the process of mining, The gas concentration can not be detected effectively in the conditions of movement and the on-site maintenance of big mechanism equipments in the temporary working location, including laying the communication lines out of time, sensor can not meet the requirements of dynamic detection, real-time transmission and rapid deployment[1][2].

We propose a wireless mine gas sensor based on ZigBee which mainly monitor mine gas concentration where the staff and machines operating location is and the exploitation of the monitoring of surface, can be installed in the miner lamp, mine car, excavators and other machinery and equipment, or the position of gas emission to make up for the lack of wired communication systems [3].

At present ZigBee technology is more and more widely used in industrial and agricultural production, ZigBee(IEEE 802.15.4 standard) is a rising wireless network technology which is of short space, low complicacy, low power consumption, low data rate and low cost.

The ZigBee stack architecture is made up of a set of block called layers. Each layer performs a specific set of services for the layers. The IEEE802.15.4 standard defines the two lower layers: the physical (PHY) layer and the medium access control (MAC) sub-layer. The ZigBee Alliance builds on this foundation by providing the network (NWK) and the framework for the application layer. Which includes the application support sub-layer (APS), the ZigBee device objects (ZDO) and the manufacture defined application objects [4]. The outline ZigBee stack architecture can be shown in fig. 1.

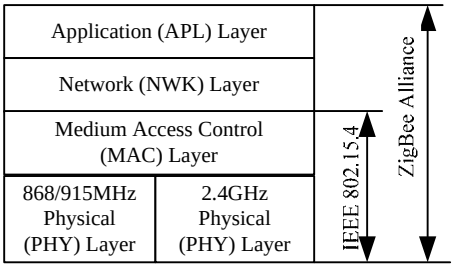


Figure 1. Outline ZigBee stack architecture

II. SYSTEM OVERVIEW

Two different type devices are a full-function device (FFD) and a reduced-function device (RFD) in ZigBee networks. The FFD can operate in three modes serving as a personal area network (PAN) coordinator, a coordinator or a device. An FFD can talk to RFDs or other FFDs, while an RFD can talk only to an FFD. An RFD is intended for end device that are extremely simple such as sensor. They do not have the need to send large amounts of data and may only associate with a single FFD at a time [5].

The arrangement and the scope of the wireless sensor nodes constantly change on the march of coal mining and advancement, which can cause serious power consumption of node in long-distance data transmission. In order to ensure the network data transmission efficiently and save energy consumption, we use the cluster tree network is a special case of a peer-to-peer network in which most devices are FFDs. An RFD may connect to a cluster tree network as a leave node at the end of a branch, because it may only associate with one FFD at a time. Any of the FFDs may act as a coordinator and provide synchronization services to other devices or other coordinators[6].The network topology is can be shown in fig. 2.

A number of wireless gas detection nodes spread over the entire monitoring area by the cluster tree network, System will class as a cluster in a certain region of the nodes, the cluster head is elected in the cluster by the clustering algorithm, receives the gas sensor information from each node in the cluster t and send information to the coordinator through a hierarchical routing protocol. Meanwhile it can receive the command of the coordinator and send it to the other node in the cluster [7].

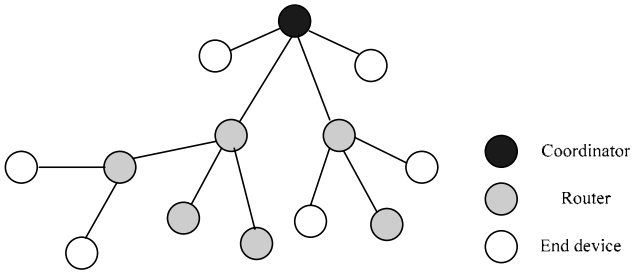


Figure 2. cluster tree network topology

III. SYSTEM HARDWARE DESIGN

The system can be divided into two parts: the sensor nodes and the coordinator node [8]. Hardware design should consider carefully several factors such as reliability, energy and cost. The specific designs are in detail as follows.

A. Sensor node

Mine gas sensor consists of electrical bridge, signal conditioning circuit, Alarm circuit. Block diagram of mine gas sensor hardware architecture can be shown in Fig. 3.

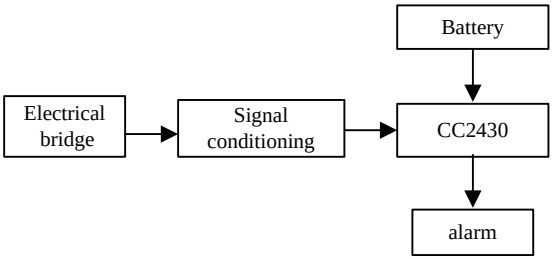


Figure 3. Block diagram of mine gas sensor hardware architecture

In fig.3, the CC2430 is used for controller of mine gas sensor. The CC2430 is a true system-on-chip (Soc) solution specifically tailored for IEEE 802.15.4 fully compatible with the hardware layer and physical layer. And ZigBee application produced by TI company [9].The CC2430 combines the excellent performance of the leading CC2420 RF transceiver with an industry-standard enhanced 8051MCU.Combined with the industry leading Zigbee protocol stack.

1) Collecting and conditioning module

The mine gas sensor is very important of detecting the mine gas concentration. We choose MJC4/3.0J sensor with supporter catalyst filled element which can detect coal mine methane with 3V power supply and can change physical quantity to electrical quantity [10]. The collecting and conditioning circuit can be shown in fig.3. The measure bridge consists of D2 (black component, also called catalytic component), D1 (white component, also called compensation component), resistors R1 and R2. The variable resistor RW can be adjusted to ensure the bridge is in a state of equilibrium. When the gas is zero, the voltage outputs zero. When there is gas, the electrical bridge breaks the balance to produce a differential output signal which is proportional to gas concentration. Differential output signal is relatively week, so we can constitute a differential input to amplify the signal by LM324 operational amplifier, direct be linked to voltage differential input. The fig.4 can be shown as follows.

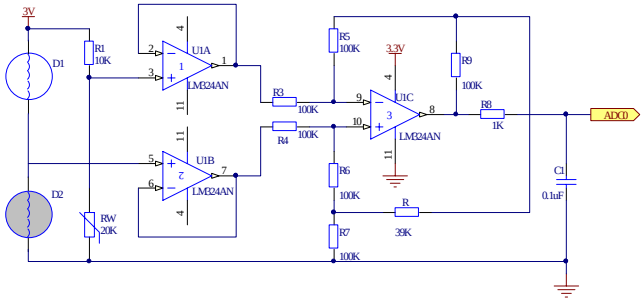


Figure 4. Block diagram of coordinator Hardware architecture

In fig.4, RW is the zero potentiometer which can realize zero point correction. The different amplificationfactor can be attained by adjusting R. The CC2430 has an internal 10 bit A /D converter, Voltage output signal directly connect to the CC2430 pin to execute internal A/D conversion, which can fully meet the precision requirements.

B. Coordinator node

The coordinator use ATmega128L as the controller. ATmega128L is a low-power CMOS based on the AVR enhanced RISC architecture 8 bit microcontroller [11]. The block diagram of hardware structure can be shown in fig.5.

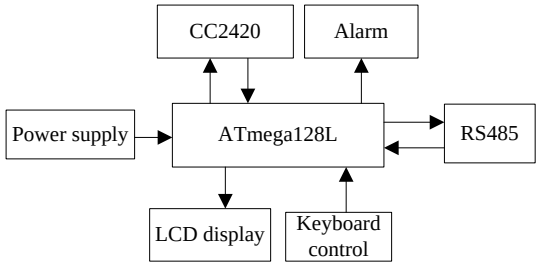


Figure 5. Block diagram of coordinator hardware architecture

1) CC2420 module

The CC2420 is a true single-chip 2.4GHz IEEE 802.15.4 compliant RF transceiver with baseband modem and MAC support designed for low-power and low-voltage wireless application produced by Ti company which is suitable for both RFD and FFD [12]. CC2420 application circuit can be shown in fig.6.

CC2420 is configured via a simple 4-wire SPI-compatible interface (pins SI, SO, SCLK and CSn) which is used to read, write buffered data, and read back status information.

CC2420 is connected with the Atmega128L with SFD, FIFO, FIFOP and CCA pins which can indicate state of sending or receive data. RESETn pin can make CC2420 reset, VREG_EN pin can start up voltage comparator of CC2420 and generate 1.8V voltage so as to put it into proper condition.

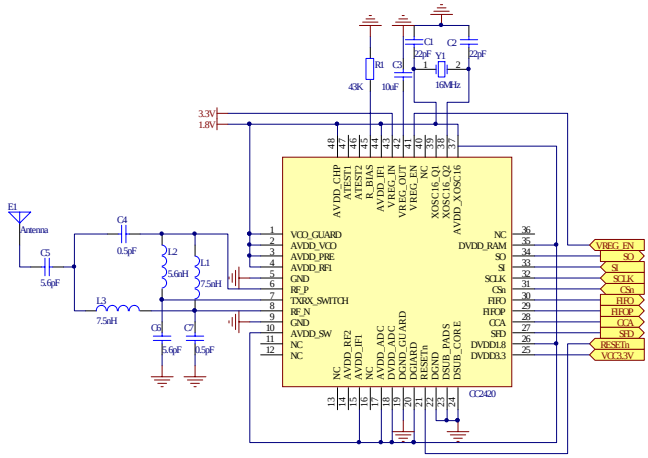


Figure 6. Block diagram of coordinator Hardware architecture

2) Power supply

Power supply voltage of ATmega128L is 3.3V, and we can convert 5V to 3.3V by LM1117 which is a low dropout voltage regulators features with 3.3V voltage output. It can be shown in Fig. 7.

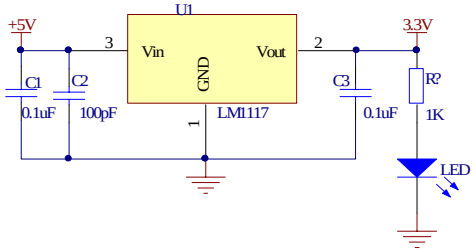


Figure 7. power supply

3) RS485 communication

The coordinator communicates the data with the wired workstation by RS485 interface. So we choose MAX485 to realize communication. MAX485 is a 5V low power the RS-485 transceiver and can meet the RS-485 serial protocol requirements [13]. The RS485 communication can be shown in fig.8.

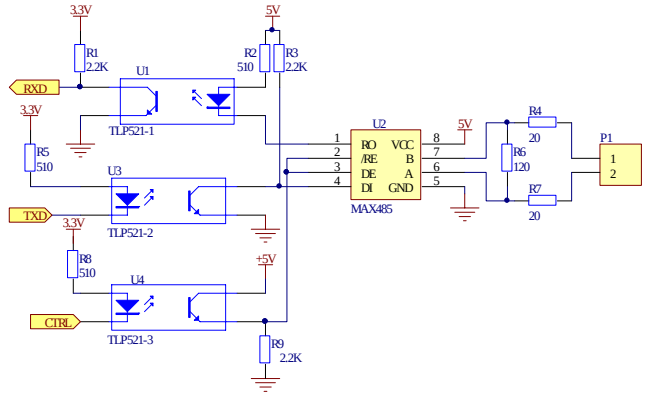


Figure 8. RS485 communication

RXD0 and TXD0 Pins of ATmega128L are connected to MAX485's RO and DI Pins through TLP521-4 which is the phototransistor optically coupled isolators. I /O pin of ATmega128L is connected to DE and /RE pins of MAX485 to control data receiving and sending to improve the immunity from interference.also we should place a termination resistors in the two point to improve the reliability of the RS485 communication.

4) LCD display

LCD display is the interactive platform between the user and the coordinator, which can display function menu by the key-press option. OCM12864-9 is a 128×64 dot-matrix liquid crystal display modules with controller by ST7565P produced by Gold Palm Electronics CO.,Ltd, which can show the current terminal node parameters from data collection terminal, such as device type, network ID. LED+ connecting to ATmega128L can control shading value of OCM1284-9 by connecting with 9015.The LCD display module can be shown in fig.9.

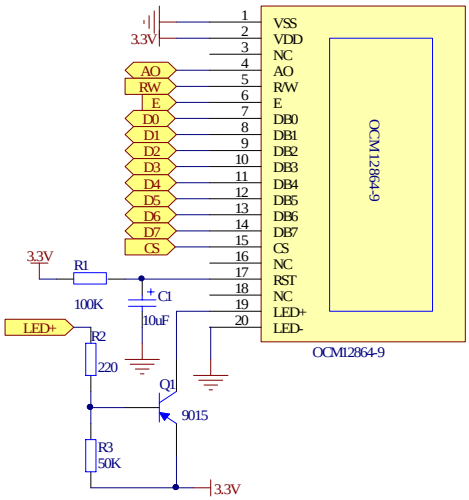


Figure 9. LCD display

5) Keyboard control

Keyboard control is designed by 4×4 matrix key array including number key and function key. Number key can set the environmental parameters upper and lower limits of the end device, group number while functional key can provide configuration and inquiring information of data collection terminal. Key control module can be shown in fig. 10.

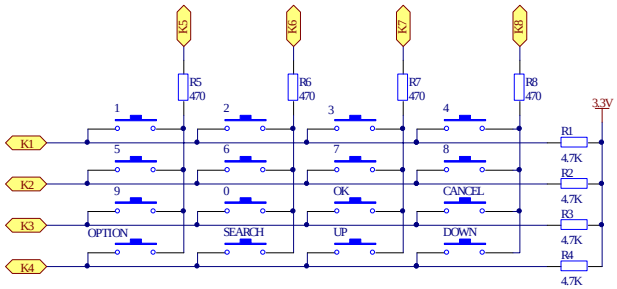


Figure 10. Key control module

IV. SOFTWARE DESIGN

Software of system uses IAR Embedded Workbench （EW）for MCS-51 produced by IAR System company which is a set of high sophisticated and easy-to-use development tools for programming embedded

application[15]. Software system use ZigBee 2007/PRO Zstack-2.0.0 of TI Company, Which can be managed by adding the operating system (OS).

The OS Abstraction Layer (OSAL) API allows the software components in the Z-stack to be written independently of the specifics of the operating system, kernel or tasking environment. OSAL is independent of ZigBee stack. But it whole stack can run based on OSAL System build a task and allocate task ID and functions.

Software design consists of sensor node and coordinator node. The specific design is in detail as follows.

A. Mine gas sensor softwre design

After power up, Mine gas sensor first initialize ZigBee stack including designated device type and network parameter configuration, become a beaconless terminal, it can search coordinator of designated channel and request to join, send its network address which is a unique 64 bit extended addresses used for direct communication with the coordinator when joining successfully, and is exchanged for a short address allocated by the coordinator. Collect the mine gas sensor concentration data every one second. The sensor node can shut off when there is no data transmission and go into the sleep mode so as to save the power consumption [14]. The flow chart of its software design can be shown in fig.11.

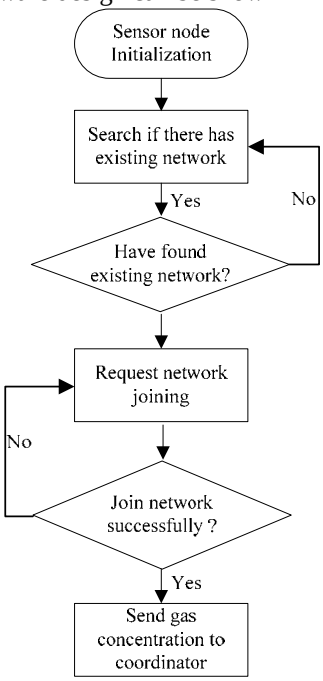


Figure 11. Flow chart of sensor node software design

B. Coordinator software design

The coordinator automatically build network after initialization and allows end devices to join the network. After end devices successfully join the network, it boots binding by key and waits for end device binding request. The key can configure network node and functions, such as network joining, address binding, routing, data collection, encryption selection node increasing, decreasing and disconnection from the network [15]. The flow chart of its software design can be shown in fig.12.

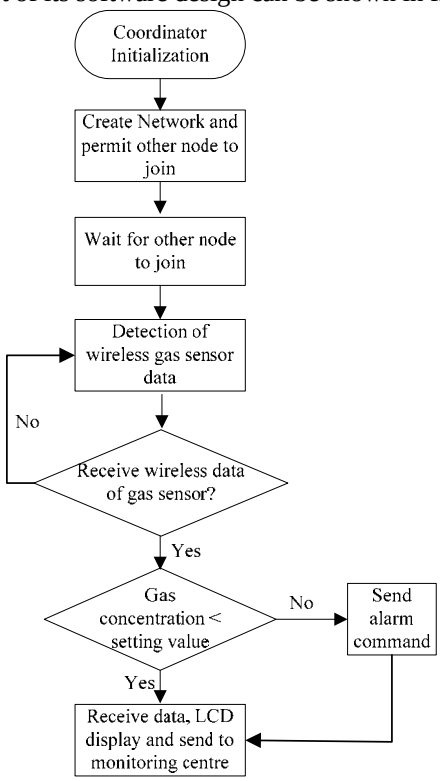


Figure 12. Flow chart of the coordinator software design

V. CONCLUSION

The underground mine gas concentration detection based on the ZigBee network can realize the wireless data transmission and greatly improve the intrinsic safety of the mine gas detection system with the advantage of the low cost and flexibility. It will play a great role in the Coal Mine Safety Monitoring systems as a supplement to the present wire transmission [16].

ACKNOWLEDGMENT

This paper was supported by Henan Education Department National Science Research Project under Grant 2010B510013.

REFERENCES

[1] M.Li and H.Guo, “Smart Sensor for Underground Coal Mine Based on ZigBee Protocol,”Instrumnet Technique and Sensor, pp.66-68,August 2007.

[2] Q.Wang,J.B.D,P.Wang,and C.Liu, “Design amd Implementation of Wireless Gas Sensor,” Instrumnet Technique and Sensor,pp.53-55,May 2009.

[3] Q.Z.Wang, R.C.Liu, Y.Q.Ma, J.C.Zhao, L.Z.feng, and S.G.Liu, “Application Study of Mine Alarm System based on ZigBee Technology,”Proceeding of the IEEE International Conference on Automation and Logistics,”(ICAL2008),Qiaodao,China,pp.2637-2540, 2008.

[4] IEEE.IEEE standard 802.15.4-Wireless Medium Access Control(MAC) and Physical Layer(PHY) Specifications for Low-Rate Wireless Personal Area Networks(LR-WPANs),IEEE ,

2003,Online,Available:http://standards.ieee.org/getieee802/download/802.15.4-2003.pdf

[5] ZigBee Specification,v1.0,ZigBee Alliance,December 2006.

[6] X.Li, X.F.Yan,Y.G.Sun,and T.Yang, "Energy-Aware Hierarchical Clustering Algorithm For Wireless Sensor Networks," Vol.19,No.4,pp.1279-1283,August 2006.

[7] X.Q.Zhang and L.Y.Li, "Design and Simulation for WSN Based on Layered Clustered Tree,"Journal of Wuhan University of Technology (Transportation Science&Engineering),Vol.32, No.6, pp.1137-1140.December 2008

[8] X.X. Cun, Z.Zhao, and C.Wang. "Field Applications and Design Technologies of Wireless Sensor Networks,"[M].Beijing:National Defence Industrial Press. pp.116-119,2009,5.

[9] CC2430 A True System-On-Chip solution for 2.4GHz IEEE 802.15.4 /ZigBee[OL].http://www.chipcon.com,www.TI.com

[10] B.lin, "Research on the Dedection of Methane Concentration Using sensor with Supporter Catalyst Filled Element on Constant Temperature,"Journal of University of Electronic Science and Technology of China. Vol .35, No.4, pp.521-523,August 2006.

[11] ATmega128 datasheet. www.atmel.com

[12] CC2420 2.4GHz IEEE 802.15.4/ZigBee-ready RF Transceiver.http://www.chipcon.com.

[13] H.J.lin, H.F.wang, N.X.Xiao,C.X.liu,and P.F.Ch, "Research on Coal Mine Personal Positioning System Based on ZigBee and CAN," Proceedings of the 2009 International Conference on new Trends in Information and Service Science(NISS2009), Beijing,China, pp.749–754, 2009.

[14] L.J.Zhou,Z.D.Li, and Y.P.Luo, "Hardware Platform of Gas Density Monitor System Based on Wireless Sensor Network,"Chinese Journal of Sensor and actuators.Vol.20,No.11,pp.2522-2525,Novmber 2007.

[15] D.X.Ruan,D.F.Tang,X.G.Zhang,and X.D.Liu, "Application of ZigBee Based Wireless Sensor Network in Underground Coal Mine Environmental Monitoring,"Coal Mine Machinery, Vol.29,No.6,pp.163-164,June 2008.

[16] W.Shi and L,L,Li. "Multi-parameter Monitoring System for Coal Mine based on Wireless Sensor Network Technology," Proceedings of the 2009 International Conference on Industrial Mechastronics and Automation (ICIMA2009),chengdu,China, pp.225-227,2009.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 082-085

Spreading Cycle Model of Emergency Events on Internet

Xiqiong Wan[1], Qi Zhu[2],Weihui Dai[3], Xiaoyi Liu[2], Diefei Sun[3]
[1] School of Mathematical Sciences, Fudan University, Shanghai 200433, P. R. China
Email: xqwan@fudan.edu.cn
[2] School of Information, Fudan University, Shanghai 200433, P.R.China
Email: angel811109@163.com, liuxiaoyi92@gmail.com
[3]School of Management, Fudan University, Shanghai 200433, P.R.China
Email: whdai@fudan.edu.cn, 0525058@fudan.edu.cn

***Abstract*—Emergency events have been known more quickly and attained more attention than they did ever before. However, the public opinion on Internet usually exerts a heavy impact on the development of the emergency events as well as the social psychology. Aiming to provide a practical methodology for the management of Internet spread of emergency events, this paper presented a cycle model to describe the spreading process, and applied the Tobit model to research the influence factors in that process. Further, a life-cycle emergency management strategy was discussed for the achievement of healthy spreading environment on Internet.**

***Index Terms*—emergency event, Internet spread, emergency management**

I. INTRODUCTION

The popularization of Internet has led to the fact that news can be published and shared at any time and place, and has become the first choice of channel for the public to have discussions and make comments. Emergency events have been known more quickly and attained more attention than they did ever before. However, the public opinion on Internet usually exerts a heavy impact on the development of the emergency events as well as the social psychology.

With these considerations, scholars have made a large number of researches on the Internet spreading mechanism and netizens' psychology and behavior.

On the aspect of Internet spreading research, J. Zhang[1] and L. J. Jiang analyzed the formation process of network opinion. W.H.Wei[2] and L. Wang[3] illustrated the impacts of Internet opinion from both the positive and negative side. X.F.Hu[4] established the small world network model that realized the regional simulation of network opinion. And L.Y. Li[5] presented the cellular automata model to analyze the specific factors' effect on Internet spread. Y.C. Liu[6] later built an agent-based Internet model, working out the different reactions of individuals that play the different roles in Internet spread of emergency events.

On the side of the social psychological impacts of emergency events, G.H. Guo, H.Y. Bi, Y. Hu, etc[7]-[12] have made some significant contributions to some special phenomena on netizens' behavior in emergency events, such as the opinion leader phenomenon, the group polarization phenomenon, the network violence phenomenon, etc. Y.M. Liu, Z.R.,etc[13]-[15] have concentrated on the public psychological impacts of emergency events and achieved some results in social psychology formation mechanism, changing features, impact factors, etc. Q.L. He, A.B.Zhou, etc[16]-[18] pointed out the psychological impact mechanisms and several psychological intervention modes.

Aiming to provide a practical methodology for the management of Internet spread of emergency events, this paper presented a cycle model to describe the spreading process, and applied the Tobit model to research the influence factors in that process. Further, a life-cycle emergency management strategy was discussed for the achievement of healthy spreading environment on Internet.

II. INTERNET SPREADING CYCLE MODEL OF EMERGENCY EVENTS

A. Internet Spreading Cycle Model

The basic composition of Internet spread of emergency events are the information source, spreading platform, and the participants. The information source can either be the Internet itself or be the traditional channel, but both spread through the network platform. The spreading platform ranges from the major media network, such as portals, network forums, to blogs and various social media networks, etc. The participants of Internet spread are usually the netizens, which can be furtherly divided into controller, opinion leader, follower and observer. They play the different roles in that spread.

The Internet spread of emergency events is featured to be anonymous, technically versatile, and very fast. It is the major reason that netizens' identification is unknown, which leads to the wide popularization and heavy impacts of the emergency events on Internet. The technical versatility, consisting of words, pictures, video and other forms, is another feature of the Internet spread that makes public favor of such way of taking in information and

This research was supported by the National Social Science Foundation of China (No.06BJL043), National Natural Science Foundation of China (No. 90924013), and Shanghai Leading Academic Discipline Project (No.B210).
Corresponding author: Qi Zhu

expressing opinions. Besides, the Internet spreading scope and influences are used to being magnified in a shortest time because of the technological reason. Last, opinions appear complicated and hard to distinguish on the Internet because its spread lacks a regulated spreading procedure.

Concerning to the Internet spreading process, this paper has referred to the corporation crisis life-cycle theory presented by Steven Fink[19], who divided the corporation crisis into five stages of incubation stage, outbreak stage, diffusion stage, decaying stage, and aftermath stage, and formed an Internet spreading cycle model of emergency events as figure 1.

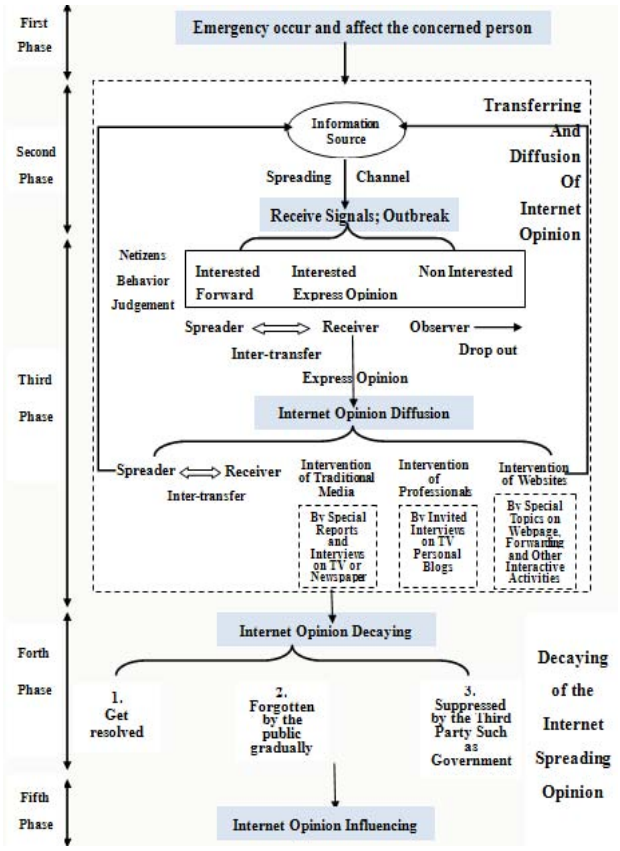


Figure 1. The cycle model of Internet spreading of emergency events

As the model shows, the whole Internet spreading process of emergency events can be divided into five stages: the incubation stage, the outbreak stage, the diffusion stage, the decaying stage, and the aftermath stage. The spreading scale and influences are different among each stage. In the first stage of incubation, the emergency occurs and affects only the related persons because it has not been spread around. In the second stage of outbreak, the information concerning the emergency is published on the Internet and gets some netizens' attention. In the next stage of diffusion, the information about the emergency gets forwarded and commented in large number of times, accelerating the spread of opinions. In the forth stage of decaying, the emergency comes to the end and the opinions decrease to the least. In the final stage of aftermath, some long-term social effects still exist, for example, the Internet popular phrases, the new regulations, etc.

B. Influence Factors

As it has been pointed out in the cycle model of the Internet spread, the prevention of information spread should be started from the period of opinion outbreak. In order to take the Internet spread into control, we use the Tobit model, which was presented by the American economist James Tobin in 1985, to calculate the extent of different key factors of the emergency events that may cause the netizens' attention. In Tobit model, the explaining variables are observable while the explained variables are able to be observed in a restricted range. With this model, we firstly selected some variables and quantitative measuring method, and collected the sample data from major websites and forums, calculating their involvement values. After that, we found out the relevant factors of netizens' involvement level, using a group of influencing variables to make regression, and figured out the major influence factors of Internet spread.

The model variables we selected can be listed as table 1.

TABLE I.
. Model Variables

Variables Name	Definition
BBS Type	1.hot forums 2.potal sites 3.major media network version
News Type	1.fact 2.commentary 3.Inquiry 4.joking
Visit Quantity	number of post clicks
News Word Quantity	number of post characters
Opinion Leader	1 for existing, 0 for none
Stage of Spreading Cycle	1.incubation stage 2.outbreak stage 3.diffusion stage 4.decaying stage 5.aftermath influencing stage

With these variables, we set the model quotation as following:

$$y_i = \begin{cases} \beta^T X_i + \varepsilon_i, \beta^T X_i + \varepsilon_i > 0 \\ 0, \beta^T X_i + \varepsilon_i \le 0 \end{cases} \quad (1)$$

$$y_i = \beta_0 + \beta_1 BBS + \beta_2 News + \beta_3 Visit + \beta_4 Newsword + \beta_5 Opileader + \beta_6 Stage + \varepsilon_i \quad (2)$$

In the quotation, β_0 refers to the intercept, β_i refers to the coefficient to be estimated, ε_i refers to the differential, y_i refers to the replying number of the ith post.

With the software of Stata 9.0, we calculated the relevant coefficients of the influencing factors. The results are showed in table 2.

TABLE II.
. The Relevant Coefficients of Variables

	BBS	News	Visit	News word	Opinion leader	Stage
BBS	1					
News	-0.0885	1				
Visit	-0.1706	-0.1107	1			
News word	0.0005	-0.1091	0.1841	1		
Opinion leader	-0.391	0.0505	0.0207	-0.0799	1	
Stage	0.0173	-0.1039	-0.1114	-0.2512	-0.0741	1

According to the results, all the absolute values of the coefficients in the table are below 0.6, meaning that the explaining variables are in small relevance.

Further with the State 9.0 and Tobit regression method, we made regression analysis of all 1896 sample data. Table 3 shows the regression results.

TABLE III.
. Tobit Regression Results

Tobit regression Number of obs 1896
LR chi2(6) 40.72
Prob > chi2 0
Log likelihood -239.61507 Pseudo R2 0.0783

reply	Coef.	Std. Err.	t	P>t	[95% Conf.	Interval]
BBS	-7.506056	2.13949	-3.51	0.001	-11.75922	-3.252891
News	-1.83664	1.173084	-1.57	0.121	-4.168654	0.4953736
Visit	0.0001976	0.0000454	4.35	0.000	0.0001073	0.0002878
News word	-0.13193	0.303517	-0.43	0.665	-0.7353	0.471445
Opinion leader	8.508498	3.097694	2.75	0.007	2.350486	14.66651
Stage	2.379038	1.410971	1.69	0.095	-0.42588	5.183955
/sigma	10.91701	1.01422			8.900804	12.93321

Obs. summary: 93 left-censored observations at reply<=0
1803 uncensored observations

According to the results, the factors that pass the T-test are the BBS type, the visit quantity and the opinion leader. For the BBS type factor, the result shows that the netizens' involvement extent is relavant to the website type. The netizens are more likely to participate in the forums that are closer to the public forum. For the visit quantity factor, its coefficient is close to 0, indicating that the post click numbers have no relationship with the reply numbers. This is because the fact that the overload of Internet information let the netizens pay more attention to the quality of the information. The numbers of post click do not affect the netizens' involvement level, either. For the opinion leader factor, the result proves that the opinion leader effect is obvious in promoting the Internet spread.

Besides, factors of the news type, of the news word quantity and of the spreading stage do not pass the T-test, indicating that these factors do not have much contribution to the Internet spread of emergency events.

III. THE EMERGENCY MANAGEMENT OF INTERNET SPREAD

Since the Internet spread has both the positive and negative impacts on the society, it is essential to build a correspondent emergency management mechanism. As is talked about above, the process of the Internet spread of emergency events can be concluded into a cycle model, which contains five phases. In this section we will present a life cycle emergency management strategy for the achievement of healthy spreading environment on Internet.

To treat the Internet spread as a special emergency, its management can be divided into the same five phases as the ones of the Internet spreading life-cycle model. In the first incubation stage, monitoring and warning of the Internet information about the emergency events is the major task for the management department. In the second outbreak stage, it should be put in the primary place to publish and to get the feedback of the emergency event information. In the stage of the opinion diffusion, to collect information and analyze the emergency caution constitutes the main responsibility of the management. The related department of government should be involved in controlling and guiding the opinion directions. In the following opinion decaying stage, emergency events that caused Internet spread of opinions should get emergency processing, and mass media starts to make conclusion and review of the whole process. In the last aftermath stage, the emergency management effect is evaluated and the governmental regulations are revised to improve.

In this life-cycle emergency management of Internet spread, corresponding measures and intelligence information system procedures are presented with each phase of the Internet spread of emergency events in details.

V. CONCLUSION

This paper researched the Internet spreading mechanism, presenting an Internet spreading cycle model and working out the influencing factors. On such basis, a life-cycle emergency management strategy was presented.

How to make use of the Internet opinion power to improve the governmental policy is our future research direction.

ACKNOWLEDGEMENT

This research was supported by the National Social Science Foundation of China (No.06BJL043), National Natural Science Foundation of China (No. 90924013), and Shanghai Leading Academic Discipline Project (No.B210).

REFERENCES

[1] J. Zhang, "On the general discipline of the formation of Internet opinion," *Marketing Modernization*, vol.3, pp.189, 2005.

[2] W. H. Wei, "On the influence of Internet opinion," *Journal of Adult Education of Gansu Political Science and Law Institute*, vol.6, pp.170-171, 2006.

[3] L. Wang, "On the Internet opinion of emergency event," *Youth Journalist*, vol.15, pp.89, 2008.

[4] X. F. Hu, "Public opinion propagation model based on Small world networks," *Journal of System Simulation*, vol.18(12), pp.3608-3610, 2006.

[5] L. Y. Li, "Research of network transmission factory about public opinion based on cellular automata," *Science Technology and Engineering*, vol.8(22), pp. 6179-6183, 2008.

[6] C. Y. Liu, "Study on agent-based communication network model of public opinion on Internet," *Computer Simulation*, vol.26(1), pp.20-23, 2009.

[7] G. H. Guo, "On the new subject of public opinion: Internet citizen," *Journal of Social Science of Hunan Normal University*, vol.33(6), pp.110-113, 2004.

[8] E. H. Wang, "The composition and activities of Internet users of china," *Statistical Research*, vol.7, pp.55, 2005.

[9] H. Y. Bi, "Analysis on the netizen behavior in Internet intelligence," *Guangxi Social Sciences*, Vo.4, pp.157, 2007.

[10] Y. Hu, "Leader formation model during public opinion formation in Internet," *Journal of Sichuan University (Natural Science Edition)*, Vol. 45(2), pp. 347-351, 2008.

[11] X. Luo, "Research on the formation mechanism of Internet opinion violence," *Contemporary Communications*, vol.4, 2008.

[12] Q. Li, *Development of the Internet Opinion Violence and Its Response*, Xiamen: Xiamen University, 2008.

[13] Y. M. Liu, "Research on the effect factors of the psychological carrying ability in the emergency event," *Journal of Inner Mongolia Agricultural University*, vol.6(4), pp. 150-151, 2004.

[14] Z. R. Liu, "Review of the netizens' psychology From the irrational Internet opinion," *Modern Communication (Journal of Communication University of China)*, vol.2007(3), pp.167, 2007.

[15] Y. Liu, "The simulation of the public psychology in SARS on complexity adaptation System," *New Research in Management Science and System Science*, pp.721, 1995.

[16] Q. L. He, "Psychological stress of critical incident and crisis intervention," *Occupational Health and Emergency Rescue*, vol.26(5), pp. 252-254, 2008.

[17] A. B. Zhou, "The psychological impact mechanism of critical incident and individual responding strategy," *Journal of Hexi University*, vol.21(1), pp.106, 2005.

[18] H. J. Dong, "On psychological influence and individual coping with vital emergent events: a case of India Ocean tsunami," *Journal of Natural Disasters*, vol.15(4), pp. 88-91, 2006.

[19] S. Fink: *Crisis Management: Planning for the Inevitable*, New York: American Management Association, 1986.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 086-090

Research on Handoff for Mobile IPv6

Jia Zong-pu,Wang Gao-lei
School of Computer Science and technology/Henan Polytechnic University, Jiaozuo 454003, China
jiazp@hpu.edu.cn, wglwgl123@sina.cn

***Abstract*—The handoff delay of Mobile IPv6 seriously affected the real-time communication service quality, therefore various improvement methods based on the basic Mobile IPv6 protocol are proposed. The working principle of Mobile IPv6 is described , the current main switch methods are summarized and the typical methods are detailed and compared in this paper. And at last the future research hot-spots are proposed.**

***Index Terms*—Mobile IPv6, FMIPv6, HMIPv6, FHMIPv6, FHMIPv6, W-MPLS**

I. INTRODUCTION

With the rapid development of multimedia technology, the demands of video, audio and other information are also increasing. But higher requirements are needed by Qos mobile user's frequent handoff, and limited bandwidth and other factors. So, Mobile IPv6 switching schemes are eager to study further.

In November 1996, IETF announced a draft agreement on Mobile IPv6. After 24 versions of the improvement, it was submitted as the standard Mobile IPv6 protocol, which still has more problems to be solved, such as security, AAA, handoff delay, multicast and so on. In order to achieve seamless roaming, switching performance, the basic Mobile IPv6 protocol switch method had to be improved.

Currently, Mobile IPv6 switching research is very active. IETF has also set up the MIPSHOP working team, who is responsible for solving the standard Mobile IPv6 protocol switching delays and developing appropriate standards. Fast handovers for Mobile IPv6 (FMIPv6) scheme and Hierarchical Mobile IPv6 management (HMIPv6) scheme have been made with some improved, combined and other switching schemes. The main Mobile IPv6 switching is summarized and other schemes is analyzed and compared.

II. THE PRINCIPLE OF MOBILE IPV6

In the basic Mobile IPv6 protocol, when switching between different subnets, Mobile Node (MN) needs to go through mobile testing, the new address configuration, duplicate address detection, binding the registration process, resulting in a lot of switching delay. The handover operation is showed as Figure 1.

The mobile node's address obtained in the hometown network is called the hometown address, obtained in the outside areas network called the transmission address, the MN in the hometown network correspondence through the hometown address. When the MN roams to other networks or switches during the different network, after it examines already being in the new network, it configures the current network of transmission-address, through duplicate address detection to validate the uniqueness of the addresses, it sends a bind update to Home Agent (HA). After receiving updated information, the HA returns a confirmation information, in the future, the HA can easily establish a connection with the MN directly.

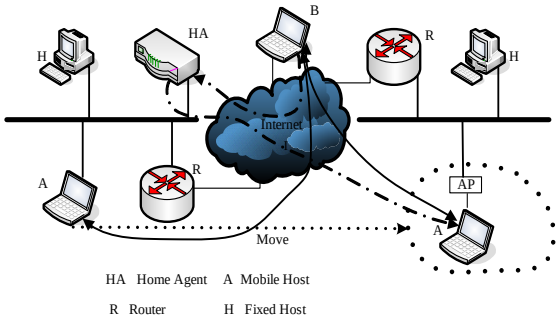


Figure 1. The working mechanism of the Mobile IPv6

In the basic Mobile IPv6 protocol, problems of big switching delay, high data loss rate and heavy signal load etc are appeared in the switching process. In view of this, many improvement schemes are made from all angles of the research. Today, the most typical schemes are as the following categories: fast handover scheme, hierarchical management scheme, fast handover for hierarchical scheme, the mechanism based on MPLS and flow label.

The above plans make the improvement separately from the different angles, and also have their own advantages and disadvantages, so we evaluate whether a program is good or bad, from a comprehensive consideration contains the switching delay, packet loss, signal load and design complexity. Under different demand, we should choose programs based on specific circumstances and as far as possible meet our needs.

III. FAST HANDOVER SCHEME

The main idea of Fast handover scheme improves the switch performance by introducing link layer mobility prediction or link layer trigger mechanisms, through altering the basic Mobile IPv6 protocol in motion detection, new care-of address configuration, and duplicate address detection process. The benefit of such programs is that it effectively reduces handoff latency and data loss rate, while increasing a new signal load.

[1] Supported by the Open Foundation of the Key National Defense Science and Technology Laboratory of Education Ministry in JiLin University (No. 421060701421).

At present, representative scheme in fleetness eager exchange is Mobile IPv6 fleetness eager exchange for scheme (FMIPv6) [2]. To further enhance the cutting performance, Enhanced Forwarding from the previous care-of address (EFWD) program [10], and Router-based switching program [4] has been proposed, the former design is which through introducing link layer trigger mechanism and establishing the tunnel between the old and new networks to reduce switching delay and data loss. The latter is designed to access the router instead of the use of the MN to do motion detection, care-of address configuration and duplicate address detection work.

FMIPv6 basic operation as shown in Figure 2 :(1) When the MN as the second level trigger being aware of the need to enter a new subnet, the router will send a request broker news RtSolPr to the old router. (2)The old router receives and later returns a cut initiate news HI to the new access router. (3)The new access router receives a message, and then it sends a confirmation message HACK. (4)As an agent on a router solicitation message response, old access router will send a Proxy Router Advertisement (PrRtAdv) message to MN and MN will get the care-of address. (5)MN receives PrRtAdv message sent by the old router and gives a fast binding acknowledgment message F-BACK to MN and to the network in which the old router is located and also to the new access router network through the tunnel. (6)When the MN get to a new subnet and has worked with a new subnet established after the second layer connection, MN issued a fast neighbor advertisement messages F-NA, then new access router can forward data to MN.

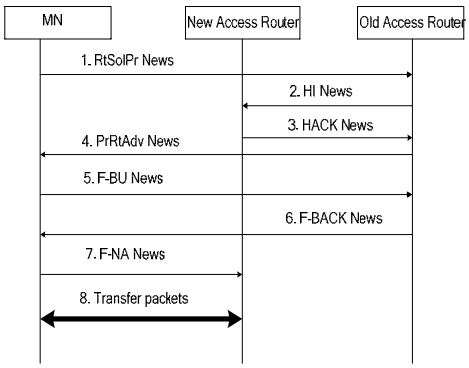


Figure 2. FMIPv6 switching process

Through analysis of the switching process, we can find FMIPv6's handover delay is smaller than the basic Mobile IPv6, data loss is low, but it increase the load of signaling interaction and also the complexity of protocol design and implementation.

IV. HIERARCHICAL MANAGEMENT SCHEME

Hierarchical Management is registered through the introduction of local management mechanism to amend the basic Mobile IPv6 protocol in the binding registration process, reduce the registration frequency of the MN to the remote HA and CN, thereby reducing switching delay. The advantage of such schemes is effectively reduces signal load.

At present, the level switch scheme (HMIPv6) [5] become a classic hierarchical management class switch program. It is better than FMIPv6 in improving the overall performance. After years of hard work, so many this kind of methods have been proposed based on different aspects. For example: Care-of address pool based on Hierarchical Mobile IPv6 handoff scheme [12], the main idea of this scheme is to introduce address pool in the access router and Mobile Anchor Point(MAP), the address used in this network is stored in the address pool, eliminating the need for care-of address of the DAD operation. Adaptive Active forecast neighbor unicast handoff scheme for Mobile IPv6, it uses adaptive and active prediction algorithm, and full account of special circumstances such as ping-pang effect, it combines hierarchical mobility management approach, while uses predictive neighbor unicast, in order to reduce the network load, and reach fast smooth handoff. Hierarchical management of MAP discovery protocol [8], this agreement improves MAP protocol of the HMIPv6, achieves the function that the MN chooses the MAP agency intelligently and selects the MAP discovery protocol on the router to make transparent, which is easy to promote. Switch based on PMIPv6 domain management method [11], this method updates message by sending messages in PMIPv6 inter-domain binding PBU, which makes switching objectives PMIPv6 domain ahead of time that the home network prefix of MN, and avoids the participation of the MN mobility management and re-configured care-of address, which is effective to reduce the handover latency. Switching HMIPv6 of improved DAD policy [3], this scheme uses link layer to assist network layer switching, and adopts distribution of effective management of care-of address new strategies, in order to avoid the DAD operation during switching process.

HMIPv6 handover process is as follows: (1)If moving locally, the MN only needs to register a new care-of address to MAP, MAP as a local home agent at this time, it accepts all the packets which are sent to the home agent , and sends to the MN's current address. (2)When the MN moves to a new MAP management area, MAP discovery operation will be conducted, and gets two care-of addresses through the access router uses stateless way: regional care-of address and link care-of address. (3)MN sends a binding and updating message to the MAP, the message forms new regional care-of address in the field of home address, link care-of address as source address binding update message, and binding update message is to bind the MN's region address and link care-of address. (4)MAP returns the binding confirmation message to the MN, to show the success or failure of the registration. (5)After MN receives the confirmation message, MN will bind the regional care-of address and MN's home address, notices HA and CN, and write to the home agent and the binding cache of the CN.

It can be seen from the above process, HMIPv6 in signal load increased significantly and in the switching delay and data loss slightly higher than the MIPv6. In the implementation complexity aspects, HMIPv6 introduced the MAP agency, an increase of implementation

complexity and easily form a network bottleneck when there are too many Mobile Nodes.

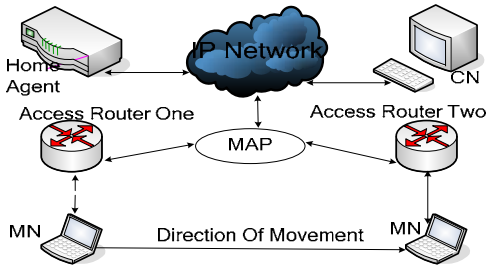


Figure 3. The structure of HMIPv6

V. FAST HANDOVER FOR HIERARCHICAL SCHEME

Program combines the former two types of such switching programs and optimizes them to achieve in the switching delay, data loss and signaling aspects of the performance of the combination of increased load.

At present, more extensive application of such programs, Mobile IPv6 application layer management structure use the fast switching scheme which effectively combines the fast switching scheme and hierarchical management program that FHMIPv6 (Fast Handover Support in Hierarchical Mobile IPv6), and shows good switching performance. To further improve the switching performance, it made for Mobile IPv6 handover latency, high packet loss problem, a hierarchical information exchange based Mobile IPv6 fast switching (IFHMIPv6) mechanism[7], an information exchange mechanism designed to enable mobile node can predict the areas of access routers within the relationship between the neighbors and the corresponding second and third tier information; combination of layered switching and fast switching, hierarchical mobile IPv6 in Mobile IPv6 adjust the speed of signaling processes, simplifying preparation phase switching operation; by setting the tunnel timer, remain in the original router to establish the tunnel. The results show that: IFHMIPv6 radio access network discovery in reducing delays and the candidate routers based on the discovery delay, further reducing the overall switching latency and pack loss.

The main principle of FHMIPv6 is to apply both FMIPv6 and HMIPv6 in the basic Mobile IPv6 protocol at the same time, but is not a simple combination of the two, which will cause triangular routing problem.

As is shown in Figure 4, the data packet sent to MN will be transferred to the pre-access router through MAP agent, the pre-access router then transfer the packet to new access router, in the hierarchical network topologies, the data packet will go through the MAP agent again to form a triangle routing, FHMIPv6 chooses MAP agent instead of pre-access router to realize the optimization of data flow, in Figure 5, the data packet sent to the MN is sent to new access router directly through MAP agent other than pass the pre-access router, thus avoid the triangle routing.

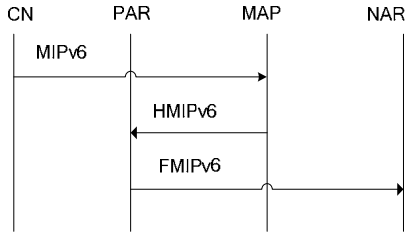


Figure 4. Data flow of simple combination of HMIPv6 and FMIPv6

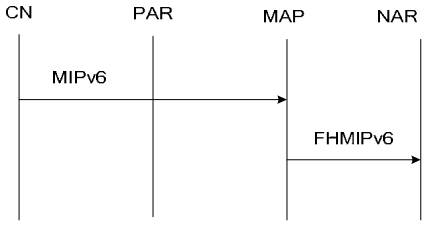


Figure 5. Effective data flow in FHMIPv6

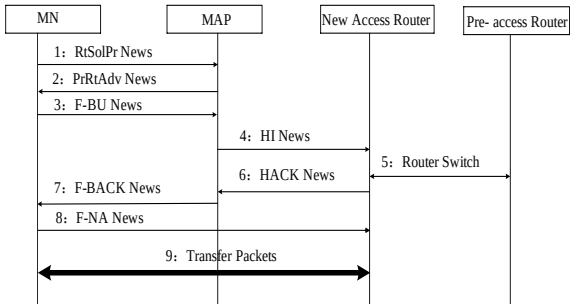


Figure 6. The process of FHMIPv6

The process of fast level switch is shown in Figure 6.(1)Because the trigger in level two indicates that MN will move to a new subnet in the same area, MN will sent a router agent request message to MAP to get the information about the new access router and the new forward address.(2)MAP will return a router agent announcement to MN as soon as it receives the message, then MN will configure a new transfer address and send a update information about the fast binding to MAP.(3)After receives it, MAP begins the switching process between the access routers through a initial message to the new access router.(4)The new access router receives the switching initial message, examines effectiveness of the new forward address, and sends a acknowledge information to MAP, the two-way tunnel between the new access router and the MAP is set up.(5)After receiving the information, MAP sends a acknowledge message of fast binding to MN.(6)As soon as MN knows the connection information, it sends a updated fast binding information to the new access router. The new access router can transfer data to MN then.

The fact that FHMIPv6 combines the advantages of FMIPv6 and HMIPv6 can be learned above, it works very well in the aspects of reducing the switch delay, data loss, and the signal load, also avoids the triangle routing

problem, but increases the complexity of designing a MAP agent and the burden of MAP agent.

VI. The Mechanism Based on MPLS and Flow Label

This method has two switching strategy: the integration of MPLS and Mobile IP, insert flow label to Mobile IP. The previous one takes advantage of MPLS, it does not need tunnel package, and it improves the performance of switch by using MPLS label to achieve fast switch. The latter one uses the flow label mechanism to realize redirection of data packet and reduce the data loss.

Recently, with more and more ISP and users moving to the MPLS network, many suggestions on the integration of MPLS and Mobile IP are proposed, such as [13], [14], [15], [16], the integration plan does not need the tunnel to transfer the data, instead, it takes advantage of the fact that search the label takes less time than search IP address, to reduce the network burden and improve the switching performance. Micro-mobile plan for MPLS is proposed by people, such as [17], [18], [19], [20], [21].

At present, many improvements have been made to this plan, a 2 5 mobility scheme for fast handover based on MPLS forwarding mechanism and a virtual interface architecture (W-MPLS) is detailed and discussion as follows [9]. In addition, in order to achieve seamlessly switching, the seamless switching scheme of Mobile IPv6 network relying on MPLS networks is proposed [22], the scheme accomplishes switching of low delay and grouping loss by limiting reconstructed domain of label switched path (LSP) and dual broadcasting data for host during switching, meanwhile an algorithm that we can fast find the nearest public point, through which group data is delivered, before and after host switched is employed to minimize delay of LSP reconstructing and redundancy of dual broadcasting. For the existing scheme of combining multi protocol label switching (MPLS) and IPv6, there have been some problems like how restoration of label switching path without mistake is established, so Predictive fast switching IPv6 base on MPLS arises [6], the scheme may minimize switching delay by defining forecast information table and algorithm of predictive MPLS transmitting and switching, furthermore by improving LSP topology construction and adopting dual broadcasting mechanism, data packets loss rate reduce effectively and ability of error recovery improves.

W-MPLS structure shows in Figure 7, RG(Root Gate) / FA is the root gateway of MPLS domain (i.e. FA); LER(Label Edge Router) is the labeling boundary router, connecting MN as first hop access router and responsible for label insertion between the data packet link road layer and network layer; LSR(Label Switch Router) is the label switching router, forwarding data depending on MPLS labels, most routers in network support MPLS protocol, LSP (label Switched Path) among FA, MN, CN and HA all transmit messages by looking for tag, instead of IP, using MPLS routers and switches that are both established ahead according to label distributing protocol, by this means higher speed can be obtained than that in which every router searches IP address in router transmitting table. The basic design concept of MPLS is that, when MN accesses router in turn, as shown in the graph, when A-LEAR1 is switched to A-LEAR2, binding registration messages are sent to new A-LER2 from MN, and then A-LER2 finally transmits the messages to FA by established LSP ahead, FA updates binding information from MN. A-LER2 at the same time notifies the front access router A-LER1 new care-of address of the MN, thereby A-LER1 transmits the data packets sent to MN packets to the new care-of address. In order to reduce the load on the RG and switching latency, optimization solution has been raised in this article that enhanced mobile LSR (i.e. ME-LSR) is introduced between MN and RG. When MN switches from the pre-network to the new network, from A-LER1 to A-LER2 as shown in the graph, ME-LSR1 receives the registration messages from MN, updates them to MN binding information table, establishes LSP about MN with new access router A-LER2 and then deserts the registration messages, not to forward them to FA/RG. ME-LSR1 intercepts the data packets and the new care-of address sent to MN.

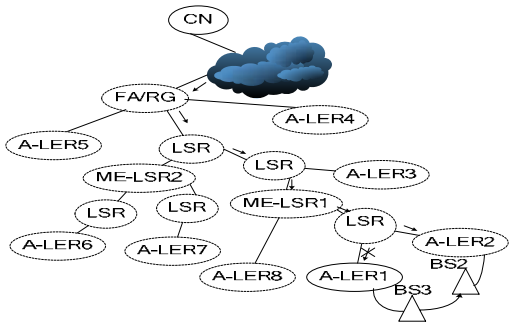


Figure 7. MPLS-optimized architecture

W-MPLS method makes use of MPLS label to forward data, instead of finding IP addresses, and it no longer needs to forward data by tunnel like Mobile IP, thereby speeds up the rate and effectively reduces forwarding delay. In the signal load aspects, MN does not need to deliver binding update request frequently to the distant FA and HA, which dramatically reduces signal load compared with MIPv6. Simultaneously due that the header of MPLS is smaller than that of IP and that W-MPLS does not require tunnels to forward data, thereby such characters reduce load of router and network; in the implementation complexity aspects, the scheme requests router to support MPLS, which increases entities of ME-LSR, thereby more complex than MIPv6 in designing.

VII. Comparision

Through above analysis of various schemes, it comes to conclusion clearly about the advantages and disadvantages of various options, the following comparisons are carried out according to the criteria of Mobile IPv6 from switching delay, data packets loss, signal load and design complexity.

Depending on actual demand and actual network characteristics, we can select the appropriate switching scheme to meet the needs of users as possible.

TABLE I. Comparison of Mobile IPv6 switch schemes

Classification	Scheme Name	Switch Delay	Packet Loss Rate	Signal Load	Complexity
Standard Mobile IPv6	MIPv6	Poor	Poor	Poor	Well
Fast Switching Scheme	FMIPv6	Well	Well	Very Poor	General
Hierarchical Management Scheme	HMIPv6	General	General	Well	General
Fast Switching Type Of Scheme-Level	FHMIPv6	Well	Well	Well	More Complex
The Mechnism Based On MPLS And Flow Label Class	W-MPLS	Better	Better	Well	General

VIII. CONCLUSION

As wireless technology develops, traditional Internet ways gradually do not meet the demands for the modern, so Mobile IPv6 technology becomes great concernment. Because of various problems on basic Mobile IPv6, it is not able to meet the needs of mobile users either, and then various modified handover methods are proposed.

As network technology evolves, the following areas will become the next hot research topics:

(1)Wireless access fast Switching based MPLS, switching strategy based on flow label [25], switching strategy base on multicast [23]

(2)Fast seamless switching, utilizing MN to cache TCP ACK to improve switching performance [24]

(3)Pursuing switching technology of low latency, simple to implement and easy to promote switching.

(4)With the diversification of wireless technology, such 802.11 series, GPRS, WCDMA have actualized seamless roaming of MN in different access network.

REFERENCES

[1] Johnson D, Perkins C, and Arkko J, “Mobility support in IPv6,”.RFC 3775, Internet Engineering Task Force, June 2004.

[2] Rajeev Koodli, “Fast handovers for mobile IPv6,” Internet D raft, draft-ietf-mip shop-fast-mipv6-02 txt, Internet Engineering Task Force, July 2004.

[3] Deng Ya-ping, Wu Ying-qiu, “Research on HMIPv6 handover latency of improved DAD policy,”Computer Engineering and Applications,2010,46 (3):94-97.

[4] Hong Yong-geun, M yung-Ki Shin, and Hyoung-Jun Kim, “Access router based fast handover for mobile IPv6,”Advanced Communication Technology, 2004. The 6th International Conference, 2004, 1: 129-132.

[5] Hesham Soliman, Claude Catelluccia, Karim ElMali,et al, “Hierarchical mobile IPv6 mobility management (HMIPv6),” Internet Draft, draft-ietf-mip shop-hmipv6-02.txt, Internet Engineering Task Force, June 2004.

[6] WANG hong-mei , Jia zong-pu etc, “Predictive fast handoff cheme for MPLS based mobile IPv6,” Computer Engineering and Applications.2007,43(31):141-144.

[7] PENG Jun, ZhANG Wei etc, “Information exchange-based fast handover cheme for hierarchical mobile IPv6,” Journal of Centeral South University (Science and Technology), 2009,40(3):749-755.

[8] Omae K, Okajima I, and Umeda N, “Mobility anchor point discovery protocol for hierarchical mobile IPv6,” Wireless Communications and Networking Conference,2004. WCNC.2004 IEEE, Volume: 4, 21-25 March 2004, 4: 2365-2370.

[9] Sethom K, Afifi H, and Pujolle G, “Wireless MPLS: a. new layer 2. 5 micro-mobility scheme,” ACM MobiWac 2004, Philadelphia, PA, USA.

[10] Gwon Y, Yegin A , Youngjune Gwon, et al, “Enhanced forwarding from the previous care-of address (EFWD) for fast handovers in mobile IPv6,” Wireless Communications and Networking Conference, 2004. WCNC. 2004 IEEE, 21-25 March 2004, 2: 861-866.

[11] REN San-yang, CHAI Rong etc, “Proxy mobile IPv6 based inter-domain mobility management approach and performance analysis,”Application Research of Computers, 2010, 27(3):1118-1121.

[12] Cai Kai , Yang Zhi-min, “The research and realize of HMIPv6 based on the level of care-of address pool,” Shandong:Shandong University, 2009.

[13] Grimminger J, Huth H P, “Mobile MPLS-a MPLS based micro-mobility concept,” Wireless World Research Forum, Stockholm, Sep 2001.

[14] [14]Ren Z, Tham C, Foo C, et al, “Integration of mobile IP and multi-protocol label switching,” IEEE ICC 2001.

[15] Chen Y, Yan Z, “Effect of the label management in mobile IP over MPLS networks,”A NA 2003.

[16] Xie K, Wong V ,and Leung V, “Support of micro-mobility in MPLS-based wireless access networks,” IEEE Proc. CNC’03, 2003, 1242-1247.

[17] SuKyoung LEE, “An efficient handover control scheme supporting micro-mobility in MPLS-based wireless Internet,” IE ICE T rans Commun 2005 E88-B: 151121516.

[18] Yang T, Dong Y, Zhang Y, et al, “Practical. App roaches for supporting micro-mobility with MPLS,” ICT 2002, Beijing, China, June 23, 2002.

[19] Vassiliou V, Owen H, Barlow D, et al, “M-MPLS: micro-mobility-enabled multi-protocol label switching,”ICC’03, 2003, 250-255.

[20] Kim H, Wong K S D, W ai W, et al, “Mobility-aware MPLS in IP-based wireless access networks,” In: Proc. IEEE Globecom 2001, 6: 3444-3448, San Antonio, TX, USA,November 2001.

[21] Kaddour M, Pautet L, “Towards an adaptable message oriented middle ware for mobile environment,” A SWN 2003.

[22] Wang Shengling et al, “Seamless Handoff Scheme in MPLS-Based Mobile IPv6 Network,” JOURNAL OF XI’AN JIAOTONG UNIVERSITY,2004, 38(10):1043-1047.

[23] Emst T, Castelluccia C, and L ach H-Y, “Extending mobile-IPv6 with multicast to support mobile networks in IPv6,” 1st European Conference on Universal Multi service Networks 2000(ECUMN 2000), 2-4 Oct. 2000: 114-121.

[24] Charles E Perkins, “Mobile IP at IETF,” ACM SIGMOBILE mobile computing and communications review (October 2003),7(4).

[25] Castelluccia C, “A hierarchical mobility management scheme for IPv6,” Roceedings of the Third IEEE Symposium on Computers and Communications. 1998 (ISCC ’98). 30 June-2 July 1998, 305-309.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 091-093

A Research and Implementation of Model Execution Method Based on MOF

Shuqiu Li, Shufen Liu, Xiaoyan Wang, Zhongcheng Geng
College of Computer Science and Technology Jilin University
Changchun, China
Email:lishuqiu@sina.com

***Abstract*—A modeling language xKL is designed based on a thorough study of Model Driven Architecture (MDA), Meta Object Facility (MOF) and Model Execution method, combined with an analysis of the effect of MOF in modeling language field. It is then implemented in two stage model building. The conclusions are: 1) xKL accords with MOF criterion in modeling structure. 2) xKL breaks the un-executable limitation in action and 3) xKL can incarnate the descriptive ability of Object Constraint Language (OCL) in model distraction. This paper interprets and explains the design and implement of executable method based on MOF model. This tools can be used in CSCW field widely.**

***Index Terms*—MDA, MOF, xKL, Model Executable**

I. INTRODUCTION

Nowadays, the complexity, diversity and volatility of software can be developers' nightmare. UML(the Unified Modeling Language)is the most popular modeling language at the moment, which can bring huge amount of financial benefits. But it has the limitation of inexecution. MDA[1] (Model Driven Architecture) is regarded as program language rather than design language. The advantages of development software by modeling language are efficiency, quality improvement and prolonging the software life. MDA defines a series of standards, among them the core is MOF[2](Meta Object Facility). MOF proposes a hiberarchy concept in modeling field: meta-metamodel layer, metamodel layer, model layer and instance layer.

OMG (Object Management Group) constitutes a series of standards such as MOF, XMI(XML-based metadata Interchange), CWM(Common Warehouse Metamodel), QVT(Query Views Transformations) and so on. However, OMG has a big limitation, that is, it proposes a whole structure but not makes it into practice. Therefore, it is necessary that the new modeling technology, theory and practice of model execution should be further studied.

II. THE RESEARCH OF MODEL EXECUTION METHOD BASED ON MOF

A. Definition of Model

The definition of model is: the formalized criterion of system function, structure and action [3]. First, model is a kind of system criterion to regulate system structure, function or action. Second, this criterion must be formalized, a strict definition without ambiguity. Therefore, a model must be bind with a modeling language which has been strictly defined with syntax and semantics. Model mainly includes two aspects: semantic information (semantic) and visible expression method (representation) [4].

B. Two Phases Modeling and Modeling Language

There are two kinds of traditional modeling method: common modeling and field modeling. Common modeling is widely applied, however, its descriptive ability is deficient in special fields. Field modeling can describe field problem accurately, but its application is limited. All those problems promote research on two phases modeling method.

For a big complex problem, modeling should be constituted step by step in different layers, thus to result in common and accurate modeling. Under such a circumstance, MOF 4-layer structure needs modeling for three times. However, given the metamodeling module provided by MOF, that is, the modeling between layer M3 and M2 has been completed by MOF, there are just two times modeling in need(two phase modeling).

The module of M2, M1 and M0 layer are meta-metamodel, metamodel and model. The modeling between M2 and M1 is metamodeling and the modeling between M1 and M0 is modeling. The hiberarchy of two phases modeling is shown as Figure 1 as follows. Two phases modeling can be described as:

(1)Metamodeling: According to the field knowledge, metamodel can be created by metamodeling module (meta-metamodel) provided by MOF, in other word, field model/ language [5]. This can make two phases modeling flexible and applicable in different fields, thus to create various field models.

(2)Modeling: Similar to yet more detailed than metamodeling, modeling is an object class modeling. At the same time, it is restricted by metamodel. The final model is executable under the circumstance of two phases modeling.

However, the ability of defining modeling language by MOF just stagnates on definition to structure. In other words, MOF only defines what to execute, but not how to execute. In one way this is an advantage: specific modeling languages can be implemented in a wider space.

xKL proposed in this paper is a modeling language whose abstract syntax model generated by MOF. And it focuses on the modeling field. xKL's abstract syntax structure model (structure metamodel) can be abstracted

first, and then xKL's abstract syntax action model (action metamodel) can be abstracted according to their action characteristics. OCL (Object Constraint Language) expression abstract model is the basis of designing xKL action metamodel. xKL is not an ordinary program language, but a text mode language. It can describe not only various modeling elements such as packet, class, attribute, method etc. but also restriction.

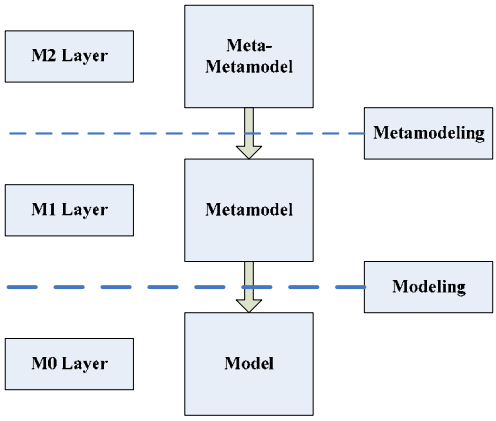


Figure 1 The Hiberarchy of Two Phases Modeling

III. DESIGN AND IMPLEMENT OF MODEL EXECUTION METHOD BASED ON MOF

A. System Structure

The system structure (5 parts) of model tool is shown as Figure 2:

(1)Model: The generation of model needs object, socket, quotative value, enumerate value and various simple type value. These conceptions are named modeling elements which are executable.

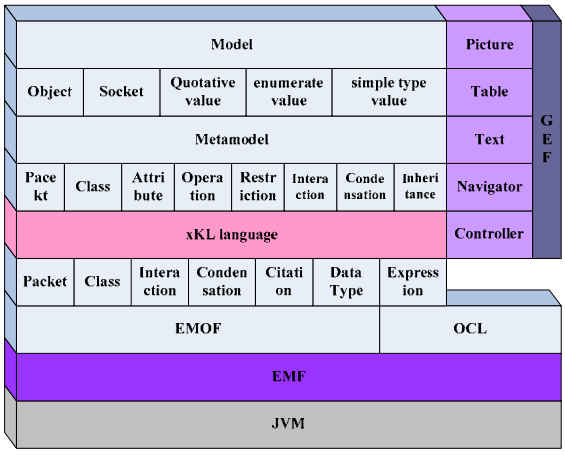


Figure 2 The System Structure of Model Tool

(2) Metamodel: The generation of metamodel needs packet, class, attribute, operation, restriction, interaction, condensation and inheritance. These concepts are called elements of metamedel.

(3) xKL: It is the basis of metamodel generation, and it provides suitable text description support for metamodeling elements. That is to say, metamodel elements can be mapped to xKL language one to one and vice versa. xKL also imports OCL expression criterion, one of the expression concepts, as basis of xKL.

(4) EMF: EMF is a modeling tool on Eclipse platform as a kind of implement of EMOF. Its Ecore model is based on EMOF, therefore, it can be used to generate EMOF model and expression model of OCL under EMF. These two models ensure the abstract syntax basis of xKL, in other word, xKL meatmodel.

(5) EMOF: The criterion of EMOF2.O proposes two new concepts, EMOF and CMOF (Essential MOF and Complete MOF). When EMOF is used for simple metamodel, it has a simple structure mapping MOF model to material implement.

B. xKL Metamodel Structure

The ultimate aim to design xKL is for providing MOF model an action semantic. Thus the generated model is not only based on MOF but also executable. Therefore, the idea of designing xKL abstract syntax is: first to generate the structure model of xKL abstract syntax based on MOF model (the structure metamodel of xKL), and then an executable action metamodel (the action metamodel of xKL) is supplemented.

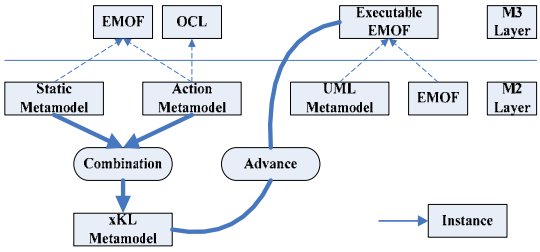


Figure 3 the Metamodel Structure of xKL

According to the idea of designing xKL abstract syntax shown as the left part of Figure 3, the xKL metamodel can be totally compatible to EMOF model. As a result xKL must accord with MOF standard. Therefore, the abstract syntax model of xKL can be divided into two parts:

- Structure Metamodel: Instance the metamodel elements of EMOF, generate and extend EMOF structure model, accord with EMOF standard.
- Action Metamodel: Instance the metamodel elements of EMOF, generate and extend OCL expression model, inherit the characteristics of OCL.

C. Structure Metamodel

xKL keeps the primary structure of EMOF and further fractionizes the syntax of categories of EMOF such as Type, Class. The structure metamodel of xKL accords with EMOF model framework. With rebuilding EMOF, some accessorial and action concepts are added to improve xKL to an executable language.

D. Action Metamodel

It is necessary to consider restriction modeling for the action metamodel generation of xKL. Adopting OCL abstract syntax model, regarding all sentences as expressions, that is to say, all sentence serials can be seen

as expression serials. So any action produced by xKL can be expressed by xKL expression serial. With reference to OCL expression abstract syntax model, xKL expression is defined in a similar way like IfExp, LoopExp, etc[6].

E. The Implement of Model Execution Method Based on MOF

The implement of model execution method is adopted with the Eclipse modeling framework and ANTLR plug-in method.

ANTLR [7] is a powerful compiling creation tool providing some functions as syntactic predicates and semantic predicates. Since EMF was developed, it has been focusing on MOF implement of OMG. EMF provides a Java code reproduction tool for developers concentrate on the model itself rather than details of implementation. The key contents under this framework are metadata, code generation and default serialization.

The code procedure of a model generation with EMF is as follows:

- Generate a UML model. There are four methods: Java class with special remark, ecore model described by XMI, XML Schema method and Rose UML class diagram file.
- Generate the EMF genmodel according to model file above.
- EMF generates Java code of model according to genmodel,.

Ⅳ. CONCLUSION

The model execution theory is researched and a modeling language xKL is designed to implement model executable mechanism in this paper. The study is based on an in-depth research of MOF. It proposes two stages modeling, the implement method of executable UML and OCL's important status in modeling field. And all those provide the theoretical support on model execution research. Then they are combined to a unique modeling system according to the demanding of model execution. xKL language can be an usable tool in CSCW field.

REFERENCES

[1] Frankel, D. S, "Model Driven Architecture: Applying MDA to Enterprise Computing", John Wiley & Sons, Inc. 2003.

[2] OMG. MOF 2.0 Core Specification, OMG formal/03-10-04 [S]. October, 2003

[3] Architecture Board ORMSC.Model Driven Architecture(MDA). OMG, Document Number Ormsc [EB/OL]. http://www.omg.org/docs/ormsc/01-07-01.pdf.

[4] Rumbaugh,J, Jacobson, Ivar and Grady Booch., "The Unified Modeling Language Reference Manual." Second Edition.ISBN 7-111-16560-8.

[5] Wei Zhang and Hong Mei, "A Feature-Oriented Domain Model and Its Modeling Process", *Journal of Software*. 2003,14(8):1345-1356

[6] Tongcheng Geng, "Research and Implement of Model Executable Method Based on MOF", *Master Dissertation, College of Computer Science and Technology ,Jilin University*. 2008

[7] Dan Yu, Hongzhi Yan, Jina Wang, "Design and Implementation of NC code Compiler Based on ANTLR", *Journal of Computer Applications. 2008*, 28(2):522-524, 527.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 094-097

The Application of SOFM Fuzzy Neural Network in Project Cost Estimate

Wen-Feng Feng[1], Wen-Juan Zhu[1], Yu-Guang Zhou[2]
[1]School of computer science and technology, Henan Polytechnic University, Jiaozuo, China
cbfwq3006@163.com
[2]China Overseas Holdings Limited, Hongkong, China

***Abstract*—Applications of neural network were widely used in construct project cost estimate. Aim at handling weakness of poor convergence and insufficient forecast, an improved method based on SOFM (self-organizing feature map) was proposed to replace the fashionable T-S fuzzy neural network. The method illustrated how to apply SOFM algorithm to improve the fault such as poor convergence and insufficient forecast, with the considering of analysis in basic principle of fuzzy neural network. After optimizing of T-S fuzzy neural network model, construct project cost estimate model had been built up. Finally, the model was set up with the purpose of comparing generalization ability by 18 examples and 2 testing samples. Comparing the simulation, a positive result was found that SOFM fuzzy neural network had a better performance in reducing the forecast error and iterating times. Therefore, this model is fit for handling construct project cost estimate.**

***Index Terms*—fuzzy neural network, self-organizing feature map, cost estimate**

I. INTRODUCTION

With the implement of evaluation of bid method which is proposed in "The construction contract and contract valuation management methods" and "tendering method", it is necessary to analysis project cost in correctly way during the process of project cost estimate. Construction cost is the essence properties of cost estimate in construct. It is the summation of general expenses. Traditional construct cost estimate method was applying mechanically estimate index and enterprise quota. The method of applying mechanically enterprise quota has great workload, and slow estimated speed. Meanwhile, quota has strong comprehensiveness. So, it can not reflect timeliness of concrete project feature. The method of applying estimate index is calculated according to completed similar project. Due to the uncalculated similarity between project, and the limitation of construct time and market condition, the modified method by price index and regression analysis can not content the requirement of estimate accuracy. How to acclimatize project cost estimate quickly and accurately to reality project cause more and more attention. For the past few years, scholars at home and abroad were proposing numerous methods. BP speediness estimate method is an effective way; however, it has many problems to solve the contradiction between reality scale and network scale because of the low speed in learning, and the over fitting performance. Reference [1] put forward T-S fuzzy neural network to settle this problem. Whereas owing to random parameter and samples acquired by equal interval, it made an ideal environment for actually samples. The network designed by this way had an inaccuracy output value and poor generalize ability. Some references raised a method that combine BP, K-means, and T-S fuzzy neural network together in order to improve the performance. However k-means need the parameter which has been set by expert. This article presents a T-S fuzzy neural network method based on KOHONEN. In another word, self-organizing feature map confirm the center and width of membership function to avoid the interferences which are made by artificial factors. And then the weight of every rule can be calculated by traditional product operator. At the end of the network, reverse calculation should be chosen to get the output value. Hereinto, the adjusting process can be realized by BP.

For the past few years, artificial neural network provide an efficient way to solve this question. Especially, the application of BP (back propagation) has a popular use. However, BP is easily to sink into birth defects take examples as topo-minimum, slow convergence speed, instability system and so on. This article present a method that combine fuzzy mathematic into neural network. Nevertheless, it is different from the model which proposed by Shi Feng in literature [1]. T-S fuzzy neural network can have a lower error, and faster convergence, merely poor generalization ability. The reason why is that center value and width are randomly set. In order to figure out this problem, this article put forward SOFM clustering method to get the center value and width, and then the center value and width should be put into T-S fuzzy neural network to compute and analysis with the purpose of getting the feasibility of this scheme.

II. SOFM FUZZY NEURAL NETWORK MODEL PRINCIPLES

A. SOFM Summary

Self-organizing feature map can be called as KOHONEN model or topology model. At the earliest, it was proposed by Malsburg. The current pattern is developed by Kohonen. The function of SOFM is adjust the weight by a great number of samples through

This study was supported by the National Science & Technology Pillar Program (No. 2007BAF23B0505).

self-organizing method, so as to make the output data can reflect disposition of the samples. The training algorithm has two parts, similarity match and update. The concrete steps can be stated as follows.

1) Initialization. A random number can be given to weight vector in output layer and then it can get a normalization dispose to acquire W_j, j=1, 2, …, m. An initial winning neighbourhood $N_j^*(0)$ can be set up, and the learning rate η was given to a initial value.

2) accept input data. Chose an input pattern in a random way from the training data and then normalize it to get X_n, $n \in c$.

3) In search of wining node. Compute the dot metrix of W_j and X_n, j=1,2,…,m. And find the maximum wining node j* of the dot metrix. Whereas if the input patterns do not have normalization, the Euclidean distance should be computed according to the following equation.

$$\left|x-\hat{w}_i^k\right| = \min_{1\le j\le m}\left\{\left|x-\hat{w}_j^k\right|\right\}$$

In order to find the wining node which has the nearest distance.

4) Define wining neighbourhood $N_j^*(t)$. Initially, neighbourhood $N_j^*(0)$ is comparatively large. During the training process, $N_j^*(t)$ gradually shrink in a unit radius.

5) Adjust weight. Node weight in the wining neighbourhood $N_j^*(t)$ should be adjusted according to the following equation.

$$\hat{w}_i^{k+1} = \hat{w}_i^k + \eta^k\left(x-\hat{w}_i^k\right)$$

$\eta(t,N)$ is a function of training time t ,and it is a function of topology distance between wining neuron j* and the jth neuron in the neighbourhood.

After finishing, centers and the distance between centers extracted from SOFM can be considered as center position and width of membership function [2].

B. Fuzzy System and Neural Network Summary

Fuzzy system is totally different from neural network among their basic feature and application. Both of them have great fault-tolerant capability during information processing. Differently, fuzzy system can simulate people's method of fuzzy logic thinking. Fuzzy mathematics is used for description, research, and handling fuzzy feature which object have. The essential concept of fuzzy mathematics is membership and fuzzy membership function. Among them, membership indicate u belong to degree of membership of fuzzy subset f. It can be shown as $\mu_f(u)$, and it is the number among 0 to 1. If the $\mu_f(u)$ close to 0, it indicates that the degree of fuzzy subset f is small. Otherwise, the degree of fuzzy subset is big. Neural network's fault-tolerant capability which shows up during information processing is from structural features of network. While our brain's fault-tolerant capability is stem from both of them------fuzziness of thought method and structural features of brain. This feature can give direction to the combination of fuzzy system and neural network.

C. T-S Model

T-S model designs controller by the method of PDC. Definitely, it means that each topo-subsystem separately design locality controller. Locality controller multiply by various locality weights, and then the value added together is the whole controller output data. Topo-fuzzy controller can be designed in a linear method, and it also can be plan by other mature theory. A simple linear method is proposed in this article. The feature of T-S model is that it not only can auto update the weight, but also can modify membership function of fuzzy subset. T-S fuzzy system is defined by the rule such as if-then. The rule can be stated as follows.

If variable is congregation Then action.

For instance, a very simple temperature adjuster made use of fan.

If temperature is very cold Then stop the fan.

If temperature is cool Then slow down the fan.

If temperature is normal Then keep the speed.

If temperature is hot Then speed up the fan.[3]

Under the circumstance when the rule is expressed as R^i, fuzzy reasoning can be shown as follows.

R^i: If x_1 is A_1^i ,x_2 is A_2^i ,…, x_k is A_k^i then $y_i = p_0^i + p_1^i x_1 + \ldots + p_k^i x_k$

A_j^i is fuzzy subset of fuzzy system. p_j^i (j=1,2,…,k) is parameter of fuzzy system. y_i is the output data which acquire according to fuzzy rules. The section of if is expressed in fuzzy logic and the section of then is a concrete value. The fuzzy inference process expresses that the output data is a linear combination of input data.

Suppose there is a output vector like x= $[x_1,x_2,\ldots,x_k]$.Based on T-S model, the membership of input variant x_j have been computed.

$$\mu A_j^i = \exp\left(-\left(x_j - c_j^i\right)^2 / b_j^i\right) \quad (1)$$

Among the equation, j=1,2,…,k;i=1,2,…,n

c_j^i is the center of membership and b_j^i is the width of membership. We should use SOFM to get the parameter of center and width. K indicates input vector and n indicates fuzzy subset.

Membership functions are computed by fuzzy, and fuzzy operator are adopted by multiply operator.

$$w^i = uA_j^1(x_1) * uA_j^2(x_2) * \ldots * uA_j^k(x_k) \quad (2)$$

Here, u indicates input data, and n indicates rule number.

In terms of fuzzy computation, output y_i can be work out.

$$y_i = \sum_{i=1}^{n} w^j\left(p_0^i + p_1^i x_1 + \ldots + p_k^i x_k\right) / \sum_{i=1}^{n} w^j \quad (3)$$

D. T-S Fuzzy Neural Network Algorithm

T-S fuzzy neural network can be split into four layers such as input layer, fuzzy layer, rules computation layer and output layer. Input layer connect with input vector x_i, and node has same dimension with input vector. The fuzzy layer adopts membership function (1) to make the input value into fuzzy membership value μ. Fuzzy rules computation layer adopts fuzzy multiply equation (2) to get w. Output layer adopts equation (3) to compute the output value of fuzzy neural network. The learning

algorithm of fuzzy neural network can be stated as follows.

1) Error computation

$$e = \frac{1}{2}\left(y_d - y_c\right)^2 \tag{4}$$

In the equation, y_d is expected output data, y_c is the real output data, and e is the error between expected output data and real output data.

2) Factor rectification

$$p_j^i(k) = p_j^i(k-1) - a\frac{\partial e}{\partial p_j^i} \tag{5}$$

$$\frac{\partial e}{\partial p_j^i} = \left(y_d - y_c\right) w^i / \sum_{i=1}^{m} w^i \bullet x_j \tag{6}$$

In the equation, p_j^i is the factor of neural network, a is the learning rate of the network, x_j is multiply product of the membership of input data.

3）Parameter rectification

$$c_j^i(k) = c_j^i(k-1) - \beta\frac{\partial e}{\partial c_j^i} \tag{7}$$

$$b_j^i(k) = b_j^i(k-1) - \beta\frac{\partial e}{\partial b_j^i} \tag{8}$$

In the equation, $c_j^i(k)$ is the center of membership function, and $b_j^i(k)$ is the width of membership function. Due to the rectification of parameters, we can get the new center and width, with the purpose of getting a more excellent output data. The parameter rectification process should be excute in the same time as factor rectification.

III. EXAMPLE APPLICATION AND ANALYSIS

A. Quantify Description of Construct Project Cost Samples

This study investigates the use of genetic algorithm in the design and implementation of neural network controller. Features of construction project have been chosen in using MATLAB to realize cost estimate. According to building operations technology which was written by Xi Zhang in Mach, 2008, 7 features were chosen as classified standards in construction such as base type, architecture form, number of plies, door and window, siding ornamental, wall, and plane assemble. The quantify description can be stated as follows.

m kinds and building project samples have been got. Each sample has 7 features. So a network input model can be fixed as follows.

$P_k=(P1_k,P2_k,\ldots,Pn_k)$ k=1,2,…,m, n=7

It can be seen that m kinds of vectors of building project samples could be built up by this method. There have 7 features in every sample. And then quantify description of any constructional engineering can be given. It can be shown as $T_i = (t_{i1} , t_{i2} , \cdots, t_{ij})$. Ti can be stated as the serial-number of the ith project. t_{ij} (j = 1,2, ⋯,7) indicates quantify values of jth feature in project i. Taking a project for example, if the project has 7 features such as brick foundation, 5 floors, timber door and aluminum alloy window, siding rock dash, standard brick, three chambers and one hall. So the quantify description can be expressed in T_i = (1,1,2,3,2,2,3).

B. Application of SOFM Fuzzy Neural Network in MATLAB

- normalization

The benefit of normalization not only can remove bizarre matrix, but also make the data of output layer more accurate.

We can choose 20 premises from a constructional operations company in China as these descriptions above. 7 features can be considered as samples. On the basis of the above method, training samples can be stated as table 1.

TABLE I. TRAINING SAMPLES

N	input data							output
	x1	x2	x3	x4	x5	x6	x7	y1
1	1	1	2	1	1	2	2	498
2	3	1	2	3	3	2	4	525
3	2	1	1	1	2	2	2	493
4	1	1	1	1	1	1	2	487
5	1	1	1	3	2	2	3	506
6	2	1	2	3	3	2	4	538
7	3	1	1	1	2	2	4	542
8	4	1	2	3	3	2	5	562
9	2	2	4	3	3	3	4	897
10	3	2	5	3	3	3	3	989
11	4	2	6	3	3	3	4	1045
12	5	2	4	2	4	3	4	876
13	5	4	6	3	4	2	4	857
14	5	2	4	3	3	3	4	923
15	6	2	3	3	3	3	4	948
16	6	3	4	3	3	3	3	747
17	6	2	4	3	4	3	4	689
18	6	4	6	3	4	2	3	936

The first step we should do is normalizing these data in order to acquire more accurate output values. Normalization should obey the following equation in MATLAB.

P(i,:)=(P(i,:)-min(P(i,:)))/(max(P(i,:))-min(P(i,:)));

- result analysis

We put the normalized samples into fuzzy neural network and SOFM fuzzy neural network, and set the iterate steps are 100. The performance figure can be stated such as figure 1 and figure 2. After comparing, it can be seen that the error of figure 1 is obviously than figure 2. From this, we can conclude that center and widths of membership function have a great meaning to the result of output data.

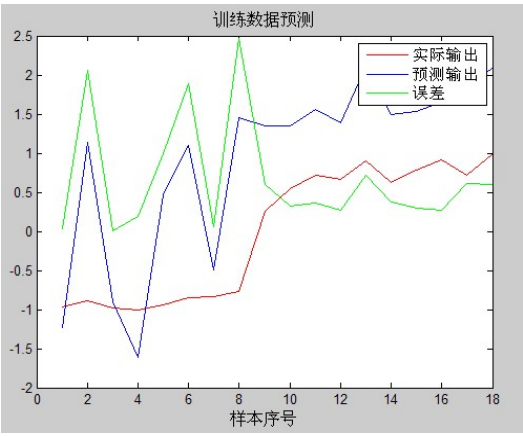


Figure1. General fuzzy neural network error

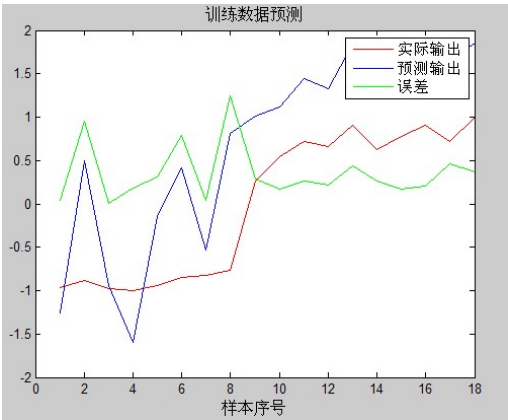


Figure2. SOFM fuzzy neural network error

Center and widths in a random way may have the error which is rather far from center and widths are set by SOFM.

Put the testing data into these two models to compare the differences. The forecast costs can be described as figure 3 and figure 4. Even though two models both illustrate that three kinds of data costs arrayed from huge to small can be stated as sample sector three, sample sector one, and sample sector two. It is said from accuracy that forecast costs from SOFM fuzzy neural network were more closely to real value 1138.28, 498.76, 1185.84.

Figure 3 shows the forecast result of general T-S fuzzy neural network without the process of SOFM system. In another word, the input data without the transformation of fuzzy set can get the similar result as figure 3.

From figure 3 and figure 4, we can get two conclusions.

(1) SOFM fuzzy neural network forecast costs are more closely to real values. It can be seen that generalization capability of SOFM fuzzy neural network is better than general fuzzy neural network.

(2) Because of the insufficient of sample numbers, input data can not accord with normal distribution. So it leads to the not ideal forecast result.

IV. CONCLUSION

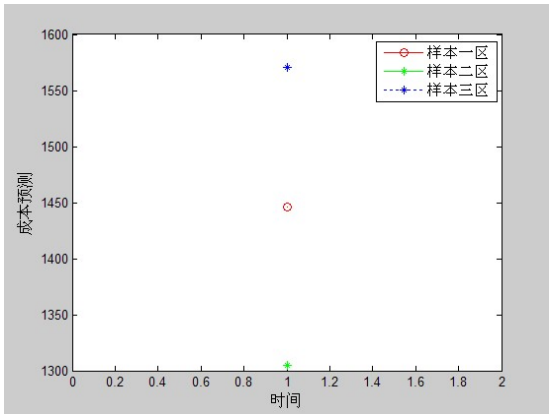


Figure3. General fuzzy neural network error

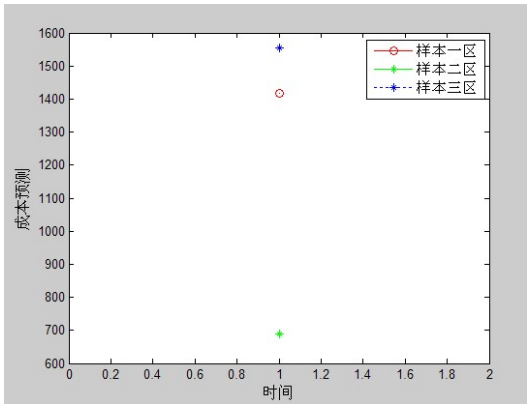


Figure4. SOFM fuzzy neural network error

This article proposes a blended learning algorithm which aims at handling the weakness of general T-S fuzzy neural network. In other words, SOFM can be used in this model to set the parameters of membership function, and the result in the output layer should be back calculated, and then the real output data make a comparison with expected data. Experiments indicate that this method not only improves convergence and efficiency of self-learning in network, but also guarantee conclusion section have higher information collection rate and accuracy rate. Meanwhile, simulation results demonstrate that the algorithm which proposed in this article has a faster speed in best approximation of target function. It is better than the general fuzzy neural network in performance.

ACKNOWLEDGEMENTS

This study was supported by the National Science & Technology Pillar Program (No. 2007BAF23B0505).

REFERENCE

[1] MATLAB Chinese forum. thirty case analysis of matlab neural network[M].Beijing University of Aeronautics and Astronautics press. April 21, 2010.

[2] Wen-Feng Feng, Wen-Juan Zhu, Yu-Guang Zhou. Application of SOFM in building classification[J].Computer Application. June, 2010.

[3] Lan Yao. Software quality forecast research based on fuzzy neural network. April, 2007.

[4] Xian-Ye Yang. Application of fuzzy neural network in land and water resource science. June, 2008.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 098-100

A kind of Design Schema of Intelligent Water Meter based on Radio Frequency Technology

Baoding Zhang[1], Yan Zhang[2]
[1] College of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, China
Email: hnzbding@126.com
[2] Mechanical and Electrical Engineering College, Jiaozuo University, Jiaozuo, China
Email: timdate@126.com

***Abstract*—Manual meter reading was main way in traditional water management. It was not only waste of human and material resources, but also very inconvenient. Especially in recent years, with the emergence of a number of high residential, this way of water management was obviously inefficient. In this paper, based on the study of existed water meters, a kind of design schema of intelligent water meter was introduced, which was based on PIC16F946 microcontroller. By this way, the efficiency of water management can be improved.**

***Index Terms*—meter reading, water management, intelligent water meter, PIC16F946**

I. INTRODUCTION

Currently, manual meter reading is still widely used in the three meters system including of water meters, gas meters and electricity meters, and lots of trouble has emerged. Especially with the development of technology, there are a lot of residential buildings in communities, and evidently the traditional meter-reading method has been unable to meet the requirements of the current situation.

There are three solutions to solve the problems of traditional manual meter reading. They are cable automatic meter reading system, wireless automatic meter reading system and intelligent water meter system.

Cable automatic meter reading system is very vulnera0ble and it needs a heavy workload of construction wiring. And in this kind of system, it is hard to find trouble spots once it is damaged, vulnerable to the impact of thunderstruck and over-voltage, too. Therefore, it is not easy to maintain for this kind of system [1].

Foundation Project: Youth Foundation Project of Henan Polytechnic University (Q2010-61)

Wireless automatic meter reading system needs not the complexity of indoor wiring, so it can avoid the shortcomings of cable meter reading system. But there is a fatal problem, power problem. In this kind of system, control module and the wireless transmission module require continuous power supply [2]-[5]. And in the system, there are generally a lot of wireless transmission nodes to transmit information for each other, which constitute to be a network. These nodes commonly used dry batteries as power supply, so the batteries must be replaced if they are run out of. For current technology and prices, the cost is larger. Therefore, the power issue has been an important factor of confining wireless automatic meter reading system to be used widely.

In contrast, IC-based smart card water meter system has great flexibility. It can not only avoid many drawbacks of manual meter-reading systems and wired meter-reading systems, but also reduce much of power consumption. In this paper a kind design schema of RF IC Card Intelligent Water Meter based PIC16F946 was introduced.

II. DESIGN SCHEMA

A. Components of system

The whole system is mainly composed of smart meter, RF cards, and water-using management software system. As shown in Figure 1, radio frequency smart card is information exchange medium between water meter and water management system, in this paper the design of intelligent water meter was mainly discussed [6].

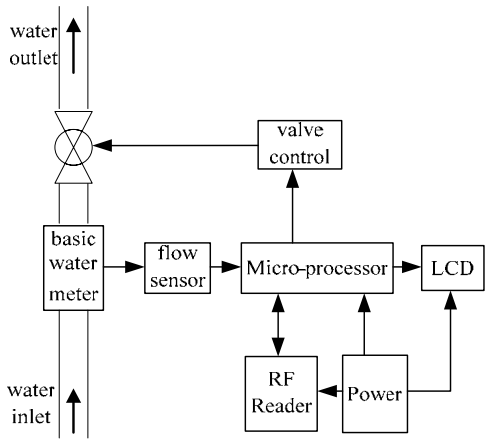


Figure 1. Components of System.

B. Design of intelligent water meter

The RF card intelligent water meter is mainly composed of basic meter, flow detection sensors, microprocessor, radio frequency reader module, valve control module, LCD display and power supply component[7].

Basic meter is an important component of measuring water amount accurately. Its main structure is the same as

ordinary mechanical water meter, which is the basis of intelligent water meter, and it is related to the accuracy of measuring.

Flow sensor is transmission part of the water meter. It is a critical component to transform mechanical signals to be electrical signals, by which the measurement of water flow can be calculated at last.

IC card is a medium for data transmission, which can be used for management and control. According to different functions and usage rights, cards can be divided into various types, in the water management system [8].

Microprocessor is like the "brain" in the whole intelligent water meter system. It can control all the components by anticipated set, such as valve control, data storage, LCD display, alarm control, power management and handling all kinds of interrupts including various card interrupts.

The system given in this paper used the chip PIC16F946 as controller. It is made by NanoWatt Technology, it is a microcontroller including of 64-pin and 8-bit CMOS Flash. And internal resources are rich. For example, there is a LCD driver, which can better control LCD flexibly. And the power consumption is very low, too. When operating voltage is 2.0V, the standby current is less than 100nA, and operating current is relatively low [10].

There is a watchdog in MCU. In order to reduce the power consumption, generally MCU can be set in the standby mode, when a radio frequency card is close to the card reader. And it will be waken up into normal mode by external interrupts.

C. Design of software

MPLAB IDE v8.10 is adopted as the development tool, which is based on the standard Win32 32-bit Windows operating system. This software is entirely self-developed by Microchip Company.

In the whole system, radio frequency cards are used to exchange information between water meters and PC. By function cards can be divided to be a lot of kinds of ones. For example, in addition to client card which is used for water purchase, there are many kinds of ones for system maintenance, such as the reset card, password setting card, time setting card, checking card, testing card, and so on.

The reset card is used to initialize the water meter. Checking card is for the collection of error messages occurred [11]. For example, magnetic failure, valve failure, battery failure and so on. All the occurring time and times of these failures were recorded in the MCU. Testing card is used for testing of water meters before the meter is out from factory. In the first sector of the card memory a series of numbers can be stored which will be used to distinguish the card belongs to which kind of one.

The flow chart of main function is as shown in Figure 2. Firstly, initializing is necessary. The first time running-flag is a byte in the EEPROM, if it is 0x5A,it is can be concluded that the water meter is not run for the first time, or it is run first. By that, some relevant variables will be initialized. Secondly, the global interrupts are turned on. In the system, three kinds of interrupts were designed [12]. They're card interrupt, flow sensor interrupt and timer interrupt. The card interrupt will caused the moment that a card is close to the water meter, and then the MCU begins to execute the card interrupt service program, it can identify card types by a byte in the card memory and will run corresponding program. A series of interrupts of flow sensor formed a string of pulses, and the water flow amount can be calculated by the pulses. The interrupts of timer are mainly used to compute the time. Lastly, after the interrupt service programs were run or there was no any interrupt at all. The MCU will be set in sleep mode. In this mode, the power dissipation is the least. If there is any interrupt, the MCU will be waken up and execute the related interrupt service code [13].

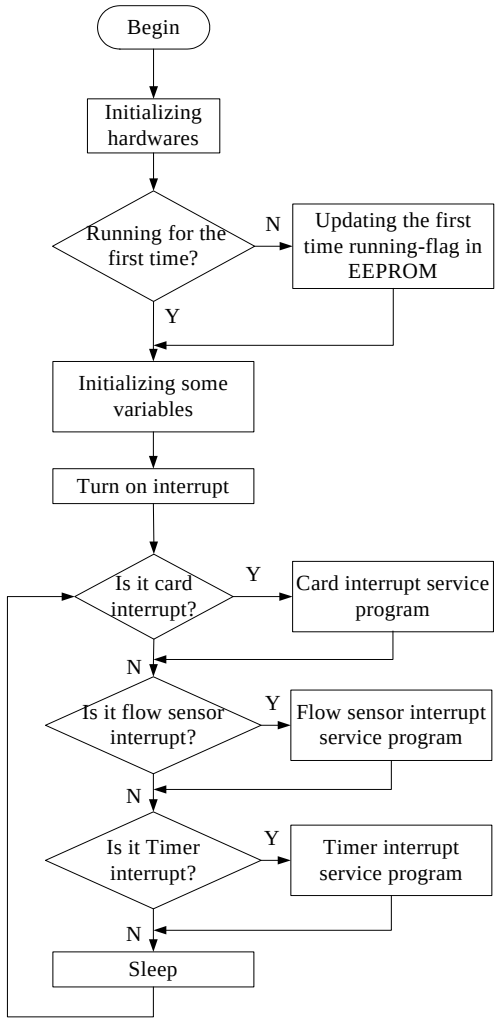


Figure 2. Flow chart of main function

III. CONCLUSION

In summary, a solution of a PIC16F946-based smart water meter was introduced. The power consumption of the intelligent water meter was very little with water amount display, alarm, error information recording and other functions. To test the feasibility of the given design, some experiments were carried out. Some parameters about the power were acquired, mainly including of the

total current of the different states. There were four kinds of currents, such as I1 (the total current while woken up by the clock interrupt), I2 (the total current while woken up by card interrupt, I3 (the total current while woken up by flow sensor interrupt) and I4 (the total current in sleep state). The experiments were done in the general environment. Five intelligent water meters were selected at random. The total currents under various states were recorded, which were shown in TABLE I. In the table, the five water meters were marked by M1, M2, M3, M4 and M5.The unit of currents was micro-amp. From the data in the table, it can be conclude that the design of the intelligent water meter is feasible.

TABLE I. CURRENT TEST

Current	M1	M2	M3	M4	M5	Mean
I1	80.0	79.9	80.3	80.0	79.7	79.98
I2	672	670	670	669	667	669.6
I3	82.1	82.3	82.5	82.0	82.3	82.24
I4	16.2	16.5	16.8	16.2	15.8	16.3

ACKNOWLEDGMENT

This paper was supported by Youth Foundation Project of Henan Polytechnic University under Grant Q2010-61

REFERENCES

[1] Islam, Nusrat Sharmin, Wasi-ur-Rahman, Md, "An intelligent SMS-based remote water metering system," ICCIT 2009 - Proceedings of 2009 12th International Conference on Computer and Information Technology, p 443-447, 2009.

[2] Jin Feng, Wang Jin-wen, Guo Fei-fei,"Application of SHA-1 algorithm in the design of prepaid intelligent water meter," Beijing Ligong Daxue Xuebao/Transaction of Beijing Institute of Technology, v 29, n 1, p 32-34+58, January 2009.

[3] Sun, Punan,"Development and application of integrated intelligent meter for measuring water ratio in petroleum," He Jishu/Nuclear Techniques, v 26, n 11, p 879, November 2003.

[4] Xu, Zhi-Qiang, Yan, You-Yun, Ran, Zheng-Yun, Jiang, Hai-Feng, "Optimizing measurement and applied system design of intelligent water meter," Huagong Zidonghua Ji Yibiao/Control and Instruments in Chemical Industry, v 30, n 3, p 63, June 2003.

[5] Murthy, C.N, Nagaraju. J.,"Intelligent thermal energy meter cum controller for solar water heating systems," Renewable energy, v 17, n 1, p 123-127, May 1999.

[6] Kim, Hiesik, Ayurzana, Odgerel, "Improvement of data receive ratio in remote water meter system by upgrading sensor," International Journal of Control, Automation and Systems, v 7, n 1, p 145-150, February 2009.

[7] Richards, Johnson, Michael C.; Barfuss, Steven L, "Metering residential irrigation water: Technological approaches and cost estimations," Journal / American Water Works Association, v 101, n 6, p 52-63+12, June 2009.

[8] Kang Yewei, Huang Yalou, Sun Fengchi. The Development of an Intelligent IC Card Cold Water Meter with Low Energy Consumption, Acta Scientiarum Naturalium Universitatis Nankaiensis Vol 39, No 5,2006.

[9] Meng Xiangyong; Shen Changjun; Sun Gang; Zheng Wengang. Transactions of the Chinese Society of Agricultural Engineering, Vol.24, Supp 2, 2008.

[10] ZHU Rui, HAN Qirui. Design of Supervisory Control System for Intelligent Housing Estate Based on PIC SCP, Computer Engineering, Vol 31, No 9, 2005.

[11] TIAN Peng, YIN Guang. The Study of System - Level Low Power Consumption Design in CMOS Circuit, JOURNAL OF LIAONING UNIVERSITY, Natural Sciences Edition, Vol 35, No 2, 2008.

[12] LIU Bing, QIAO Pei-li, ZHAO Yan. A Communication Solution of Automatic Meter Reading System, JOURNAL HARBIN UNIV. SCI .& TECH, Vol 10, No 2, 2005.

[13] Liu Ming, Zhou Feng-yu, Li Yi-bin, THE DESIGN OF AN APPL IED HANDY METER READING AND MANAGEMENT SYSTEM, Proceedings of the EPSA, Vol15, No 2 ,2003.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 101-105

The Distributed Task Scheduling Based on Real-coded Immune Algorithm

Lu Guiming[1] , Zhang Yunzhe[2]
[1] North China University of water resources and electric powe/Department of information engineering,Zhengzhou ,China
lgm@ncwu.edu.cn
[2]North China University of water resources and electric power/ Department of electric power,Zhengzhou ,China
Zhangyunzhe6113@126.com

***Abstract*—Task scheduling is a NP puzzle. Its algorithm is an important research direction. This paper proposes a task scheduling algorithm based on real-coded immune algorithm by studying and analyzing task scheduling models and immune algorithms existed. This paper discusses the coding method, the generation and update of population, the update of memory cells and the values of partial parameters. This paper explores the affinity function and the concentration function. In the end, this new algorithm is implemented in the software VC++, and is proven validity and feasibility by comparing and analyzing examples**

***Index Terms*—Immune algorithm, Real-coded, Task scheduling, Information entropy**

I. INTRODUCTION

Load distribution is a resource management module of a distributed system. It mainly redistributes system load reasonably and explicitly among processors to make the comprehensive performance of the system the best. Static load allocation algorithm makes a decision based on priori knowledge of the system to schedule a set of tasks so that each task has the smallest execution time in every goal PE. So the static load allocation is also called task scheduling problem.

In a task scheduling problem based on graph model, the Directed Acyclic Graph (DAG) can be used to make a model for a set of tasks. So the task scheduling problem boils down to scheduling optimally the set of tasks to the target PE of the system to make the scheduling length (i.e., execution time) the smallest under the premise of maintaining the precedence constraint between tasks. Except some instances with special constraint models, scheduling problems are usually still a NP puzzle, even though calculation costs and communication costs are made certain simple assumptions [1] .Many ways try to use mathematical tools (Figure, heuristic rules, etc.) to get a second-best solution, but heuristic algorithms existed have obvious limitations, such as either too complex and difficult to get a solution, or too time-consuming to be actually applied [2].

Immune algorithm[3] as a new global optimization searching algorithm[4], its self-regulation, antibody diversity, associative memory provide some theory evidences for resolving combinatorial optimization problems including task scheduling. It has been widely studied and used [5] - [11]. This paper proposes one task scheduling algorithm based on real-coded immune through task scheduling model.

II. THE MODEL OF THE PROBLEM AND RELATED DEFINITIONS

A. The Graph Model of the Task Scheduling Problem

Suppose a process set $P=\left(P_1,P_2,...,P_n\right)$ is executed in a series of the same processors, and the partial order relations ($<$) on the set P are given to compose the set $(P,<)$ of relations expressed with a quaternion $G=\left(V,E,C,W\right)$ called task precedence graph. The task precedence graph is a DAG, where $V=\left(v_j,1\le j\le n\right)$ is the set of task-nodes, and $n=\left|V\right|$ is the number of task- nodes; E is the set of directed edges representing the communication relation and the precedence relation between two nodes, and $e=\left|E\right|$ is the number of directed edges; W is the set of communication costs between any two directed edges; C is the set of computing costs between any two nodes; $w(v_i,v_j)\in W$ is the communication costs between two nodes of a directed edge. If these two nodes are dispatched to the same processor, the value is 0. $e_{ij}=\left(v_i,v_j\right)\in E$ is the directed edge. The $c\left(n_i\right)\in C$ is the execution time of the node $v_i\in V$.

In the task scheduling problem, the network connectivity between processors is given some typical assumptions [12]: （3）Storage capacity is unlimited; (2) Each PE has the same processing power; (3) Ignore the network congestion.

B. Related Definitions of the Problem

(1) The set $T=\left\{T_1,T_2,...,T_n\right\}$ with N sub-tasks, where T_i is the i-subtask.

(2) The set $P=\left\{P_1,P_2,...,P_n\right\}$ with M processors, where P_i is the i-processor.

(3) C is a matrix with $m \times n$ elements. Its element c_{ik} shows the time (assuming c_{ik} is known) that the task T_i is executed in the host P_k.

(4) W is a matrix with $m \times m$ elements. Its element w_{ij} shows the communication costs between the task T_i and the task T_j.

(5) D is a task distribution matrix with $m \times n$ elements, where

$$D_{ik} = \begin{cases} 1 & if \ \ T_i \ \ is \ \ implemented \ \ in \ \ P_k \\ 0 & else \end{cases},$$

(6)Information entropy: Let allele $\Gamma = \{k_1, k_2, ..., k_s\}$. $X = \{X_1, X_2, ..., X_n\} \in H_L^N$ is a population with N individuals,where each individual $X_i = a_{i1}a_{i2} \cdots a_{iL}$. Definition: $H(X) = \sum_{j=1}^{L} H_j(X) \Big/ L$,where $H_j(X) = -\sum_{i=1}^{s} P_{ij} \log P_{ij}$. P_{ij} is the frequency that the j-bit of N individuals gets the value k_i in the population X. $H_j(X)$ is the inforation entropy of the j-bit of X.

C. The Solution of the Problem

Suppose A is the scheduling strategy of some DAG. $W_A(P_i)$ shows the total time that the processor P_i completes the last subtask assigned under the scheduling strategy A. There is:

$$W_A(P_i) = t_k + w_k = \sum_{i=1}^{n} \left(c_{ik}d_{ik} + a\sum_{j=1}^{n}\sum_{\substack{r=1 \\ r \neq k}}^{m} w_{ij}d_{ik}d_{jr} \Big/ 2 \right),$$

where m hosts work in parallel. $W(A)$ shows the time that the system completes all of the tasks, and there is $W(A) = \max(W_A(P_i), \forall i(1 \le i \le m))$; Furthermore, min ($W(A)$) is the target required.

In which, t_k is the time that the k-host executes all of its tasks, w_k is the time that the k-host communicates with other hosts, and the constant a is used to adjust the overlap between communication time and execution time .In certain actual design process, the right value of a can be selected according to different degrees of overlap.

III. The Task Scheduling Based on Immune Algorithm

A. The Description of the Algorithm

Flowchart (Fig.1) shows that antigen corresponds to the problem to be solved and antibodies correspond to a solution of the problem.

- Input the objective function and constraint conditions as the antigen of immune algorithm;

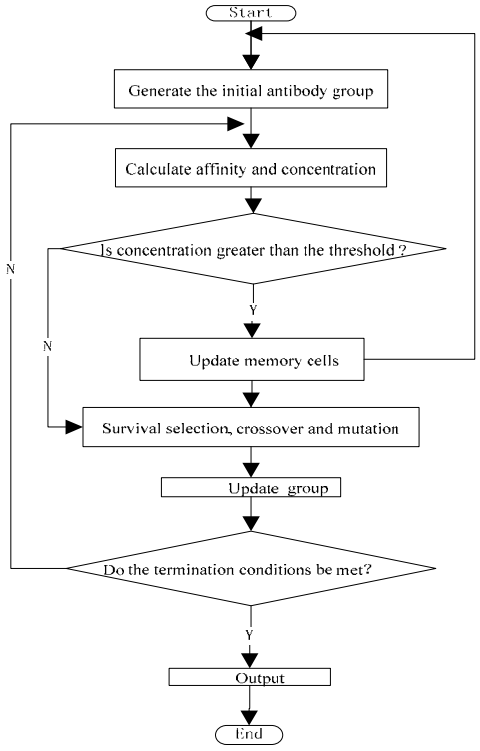


Figure 1. The algorithm flowchart

- Generate the initial antibodies based on real-coded;
- Compute affinities and concentrations;
- Update memory cells: when the concentration of antibody v exceeds the threshold value, memory cells are generated to record the antibody v representing a local optimal solution; if similar solving problems are met again in the future, memory antibodies can be directly searched for from the memory in order to improve the solving efficiency.
- Update population: get new antibodies though survival selection, crossover operations and mutation operations, and replace the antibodies in the last generation.

B. Coding Method

Immune algorithm commonly uses binary coding. But for optimization with multiparameter, the use of binary coding would lead to encoding too long and the efficiency of search reduced. This paper applies real-coded to task scheduling problems. In this method, a chromosome consists of a distribution sub-string and a scheduling sub-string, and these two sub-strings are encoded in different ways. The encoding form is expressed with a unidimensional character string, where any bit $LString_i = a$ of the distribution sub-string shows that T_i is assigned to the P_a ,and any bit $RString_i = b$ of the scheduling sub-string shows that the predecessor task of T_b will be placed before the i-bit.

C. Antigen Recognition (Analyzing Problems)

In this paper, the problem and the characteristics of its solution are analyzed, and the right expression form of the solution is designed. The solution of this algorithm is based on real-coded. The approach is to input the objective function and constraint conditions of the problem as the antigen of immune algorithm.

D. Initialize Population

The method that the distribution sub-string of chromosome is generated is that: Generate randomly one random number in internal [1: m] for each pick from 1 to n, where n is the number of tasks, and m is the number of processors. But the scheduling sub-string requests a special way to generate for the need of meeting the constraint relation between predecessor and successor of task.

E. The Affinity Computing

Affinities include the affinities between the antigen and antibodies, as well as the ones between antibodies. The affinity A_v between the antigen and antibody v can be get from the transformation of objective function $f(x, y, z)$ (i.e. $\min(W(A))$):

$$A_v = 1/(1 + f(x, y, z)). \quad (1)$$

The affinity between two antibodies reflects the similarity of these two antibodies. The more similar they are, the greater the affinity is. On the contrary, the less similar they are, the smaller it is. For the affinity $B_{v,w}$ between antibody v and antibody w ,it is:

$$B_{v,w} = 1/(1 + H(2)). \quad (2)$$

F. The Concentration C_v is the Proportion that Similar Antibodies Account for in Groups

The formula is as follows:

$$C_v = \sum_{w=1}^{N} S_{v,w} / N, \quad (3)$$

where N is the total number of antibodies;

$$S_{v,w} = \begin{cases} 1 & B_{v,w} \geq T_{ac} \\ 0 & B_{v,w} < T_{ac} \end{cases}, \quad (4)$$

where T_{ac} is the predetermined similarity threshold value.

G. Memory Cells (the memory of antibodies) Update

When the concentration C_v of antibody *v* is over than the threshold value T_c, memory cells are generated to record the antibody *v* that represents a local optimal solution; when memory cells have not yet reach the upper limitation, they begin to join the antibody; if these memory cells reach the upper limitation, the newly added antibodies replace the original antibody with the largest affinity with it.

H. The Survival Rate Computing of Antibodies and Survival Selection

The survival rate formula of each antibody is as follows:

$$e = A_v / C_v . \quad (5)$$

Equation (5) shows that antibodies with larger affinity with antigen and antibodies with low concentration have the stronger ability to survive to the next generation. The method is that: Do a getting scores test to antibodies, then select the individuals with high scores into the next generation according to a predetermined elimination rate, and wash out the antibodies with low survival rats. They will be treated as immunization supplement after being washed out.

I. Crossover and Mutation

To ensure that the newly generated individuals after crossing are still feasible solutions, the distribution sub-string and scheduling sub-string are crossed in different ways. The distribution sub-string is crossed in the classical single-point-hybrid method; while the scheduling sub-string is based on a special hybridization method. The method is that: Select randomly a cross-point and remain the gene before the cross-point unchanged, meanwhile, rearrange the gene after the cross-point according to the sequence of their crossing each other.

The mutation operations of the distribution sub-string and the ones of scheduling sub-string are also different. The mutation operations of the distribution sub-string are that: Select randomly one mutation bit in distribution sub-string, and then replace the processor randomly in the bit with another one. While the mutation operations of the scheduling sub-string are that: Let the task that the mutation bit selected randomly represents migrate between predecessor nodes and successor nodes to ensure that the solution obtained after migrating is still a feasible solution.

J. The Parameter Selection of Immune Algorithm

Generally, mutation rate P_m should get its value in interval (0.005, 0.03), crossover rate P_c in (0.5, 0.9), and elimination rate in (0.1, 0.25).

IV. Coding Implementation of Algorithm and Experimental Analysis

The algorithm in this paper is encoded with the software VC++ to verify its effectiveness and feasibility .And it is contrasted with the ones in [13] and [14].Its parameter settings are as follows:

Crossover rate: 0.9,
Threshold value: 0.8 and 0.4,
Mutation rate: 0.007,

Elimination rate : 0.15,
a =0.3

A. *Partial Code of the Program*

```
#define popsize 50 //The size of initial population
#define cellnum  20  //The number of memory cells
#define maxgen 30 //The maximal number of evolutionary iteration
#define pcross  0.9  //Crossover rate
#define pmutation 0.007 //Mutation rate
#define conthres1 0.8 //Threshold value while calculating antibody concentration
#define conthres2 0.4 The threshold value of concentration
#define peliminate 0.2 // Elimination rat
class point:public CObject
{
      … …
      … …
           point()//Constructor function
           {
            … …
           }
           point()//Destructor function
           {
            … …
           }
           unsigned int gen,score[popsize],minsurv;
      float tt[popsize];//Store the final optimization result
void main();
     void crossover(int,int);//Crossover operations
   float objfunc (pp);  //Affinity calculation between antibody and antigen
   void  mutation(int);//Mutation operations
float objfunc2();  //Affinity calculation between antibodies
      float consistence(pp);//Concentration calculation of antibody
      void  initpop();  //Generate initial population/
   void  sscore(int); // Calculate survival rate
      void  output();//Output the result

 };

… …
void point::crossover(int i,int j) //Crossover operations

{
 int k,m,p;
 srand((unsigned)time(NULL));
 k=(rand()%11)/10;
 p=rand()%10;
 if(p<pcross*10)
 {
 //Carry out crossover operation
   … …
   … …
 }
 return;
 }
void  point::mutation(int i) //Mutation operation

{
 int k,j1,m,p;
 srand((unsigned)time(NULL));
 k=(rand()%11)/100;
 p=rand()%100;
   … …
      … …
 if(p>pmutation*100)
 {
  … …
  … …  //Carry out mutation operation

 }
 return;
}
```

B. *Experimental Analysis*

The table 1 is used to record the finish time required and evolution iteration times of algorithm when the optimal solution is obtained by using the immune algorithm above in the case that the number of processors and the number of subtasks are respectively 2, 3, 5 and 20, 30, 50. In the Fig.2 are the performance curves when the algorithm in this paper and the algorithms in [13] and [14] evolve respectively under the circumstance that there are 3 processors and 50 subtasks.

TABLE 1: EXPERIMENTAL RESULTS OF THIS ALGORITHM

The number of processors	*the Number of subtasks*	*The finish time*	*evolution iteration times*
2	20	362	49
	30	425	255
	50	638	184
3	20	216	82
	30	340	293
	50	489	249
5	20	174	63
	30	293	319
	50	387	462

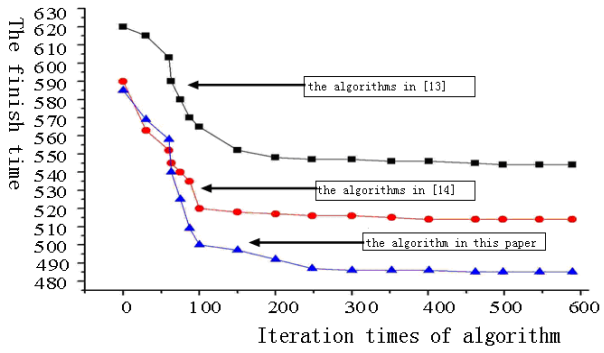


Figure 2. the performance curves of three algorithms (m=3,n=50)

According to the calculation and comparison for the data in table 1 and the performance curves of three algorithms in Fig.2, this algorithm has the following characteristics:

(1) The optimal solution that this algorithm can find has some improvements than the ones in [13] and [14], especially when the number of subtasks and the number of processors are larger.

(2) When the number of subtasks and the number of processors are larger, the convergence rate of the algorithm is slower than the one in [14] and almost the same with the one in [13].But the algorithms in [13] and [14] require more iteration operations to obtain an optimal similar solution with this algorithm.

(3) Fig.2 shows that the algorithm converges to the optimal solution with evolution iteration times creasing. Therefore, the algorithm has convergence and can obtain the optimal solution.

V. CONCLUSIONS

Task distribution and scheduling is a NP puzzle. In view of the shortcomings that traditional optimization algorithms are difficult to obtain a global optimal solution, this paper proposes a task scheduling algorithm based on real-coded by studying and comparing the models of task scheduling problems and immune algorithms in existed references. This algorithm uses the unidimensional real-coded to explore the affinity function and the concentration function, and it is implemented in the software VC++. It is proved to be efficient and feasible by analyzing a series of data obtained by computing, as would provide an effective reference method for solving task scheduling problems.

REFERENCES

[1] E.G , Coffman . Computer and Job-Shop Scheduling Theory.Now York:John Wiley & Song, Incorporated, 1976

[2] Meng Jianliang, Yi Liu, Niu Weihua. Algorithm and Realization of Task Scheduling of Distributed System and Development[j].Information Technology.2003;27(1):16-19

[3] Mo Hongwei.Theory and Application of Artificial Immune System [M].Harbin Institute of Technology Press, 2003:131-231

[4] Luo Zongjun . A class of combinatorial optimization probems and its algorithm [J].Applied Mathematic, 1996; 9(3):399-402

[5] Zeng Yi.Improvement of an immune algorithm [J].East China Jiaotong University, 2007; 24(1):123-128

[6] Ge Hong, Mao Zongyuan.The study of several parameters of immune algorithm [J].South China University of Technology (Natural Science Edition), 2002; 30(12):15-18

[7] Dipankar Dasgupta Ed. Artificial Immune Systems and Their Applications [M].Berlin , Heidelberg:Springer-Verlag, 1998

[8] Zheng Rirong, Mao Zongyuang, Luo Xinxian. Analysis and research of improved artificial immune algorithm [J]. Computer Engineering and Applications 2003;39(34): 35-37

[9] Li Jincheng, Zhang Guozhong,Teng Honghi,Zhou Sheng,Wu Hongxia . Research on immune algorithm [J].Journal of Shenyang Institute of Aeronautical Engineering.2005; 22(5):82-85

[10] de Castro.L N.Timmy J.An artificial immune network for multimodal function optimization[C]. Proc. of IEEE Congress on Evolutionary Computation, Honolulu, USA. 2002:699-704

[11] Wang Lei,Pan Jin,Jiao Licheng.The Immune Algorithm [J]. Acta Electronica Sinica, 2000; 28(7):74-77

[12] Wu Jie [theU.S.],Gao Chuanshan translated.Distributed System Design [M].Mechanical Industry Press, 2001:192-216

[13] Sun Jun,Xu Wenbo. A GAs-based Algorithm for Task Scheduling on Distributed Systems [J].Computer Engineering and Applications,2003

[14] Zhong Qiuxi,Xie Tao,Chen Huowang. Task allocation and scheduling based on genetic algorithm [J].Computer Research and Development, 2000; 37(10):1197-1203

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 106-108

Value Rational Consideration of E-learning

Huang Feng
College of Mathematics and Information Science; Henan Polytechnic University; Jiaozuo; China
hfg@hpu.edu.cn

***Abstract*—As an idea of learning and a mode of learning, e-learning dramatically highlights the subjectivity of the learners and make value rationality known. Meanwhile, value rationality appears damage and breaking in this process. Therefore, it is necessary to make learners fully exert subjective activity and intensify effective supervision on e-learning to construct e-learning and value rationality harmony. In addition to this, it must give full attention to combine the real world with the virtual world.**

***Index Terms*—e-learning; value rationality; subjectivity**

I INTRODUCTION

With the rapid progress of information technology and the great development of internet, as an idea of learning and a mode of learning, e-learning has been favored by more and more people. This new learning approach greatly highlighted the subjectivity and initiative of learners. However, value reason of the subject hasn't played adequate role in the e-learning, and even has the development of deformity. We must urgently put forward to improve and perfect it.

II VALUE RATIONALITY IS BOOSTED IN THE E-LEARNING

Web-based learning known as e-learning and digital learning. It refers to learning which is based on the development of computer and network technology. Its main feature is high subjectivity, interactivity and virtuality, and it assures the resources available with abundance and expansibility. And in this process, the subjectivity has been greatly highlighted. Subjectivity refers to an initiative, independent and creative nature of the internal regulations that people form, establish and demonstrate in practical activities. Value rationality refers to human spiritual strength that could regulate and control human desire and behavior to achieve the aims and aspirations. As human peculiar rational existence, it intrinsically lead the subject's value judgment, value orientation and value choice, etc.

Value rationality is the mapping of subjectivity on the level of value. In the web-based learning, the prominence of subjectivity makes value rationality widely known.

A. E-learning greatly advances the autonomy of learners.

In the e-learning, the learners' autonomy have been fully exhibited. First, e-learners have full autonomy to choose study time which is superior to the traditional learning. In traditional classroom, the learners are in a passive position due to many factors. In web-based learning，the learners could choose the time to study and adjust independently based on value rational consideration. Second, e-learners have full autonomy to choose learning content according to their subjective desire and value demand. Third, e-learners could independently evaluate their learning effectiveness in accordance with their value rationality.

B. E-learning greatly contributes to the learners' creativity.

In the learning activities, creativity is one of the main and important content of the subjectivity. First, learners 'choice of network resources could activate the creativity of the subject. Network resources are very rich, covering a range of knowledge and content. Learners creatively choose learning content to start learning on the basis of subject quality and value demand. Second, the process that the learners analysize and understand the knowledge could fully reflect the creativity of the subject. Especially, abstract thinking ability is applicated in processing and handling the knowledge. This makes original state of knowledge changed. The learners could get update and deeper recognition. Third, the learners put creative knowledge on the network for interact learning which greatly enrichs the creativity of the subject. E-learning is a strong interactive approach to learn, not a one-way learning. Learners upload their new knowledge to the network which could not only enrich web resources, but also further check, repair and develop the original knowledge.

C. E-learning greatly promotes the reflection of the subject.

On the one hand, e-learners could consciously think how to adjust the direction and progress of their own learning, develop their cognitive level and optimize their thinking ability and structure, based on their own choice, cognitive level, structure and degree. On the other hand, the process that e-learners gradually update knowledge and get more information would be a perverse incentives for reflection.

III VALUE REASON IS BROKEN IN THE E-LEARNING

On the one hand, value rationality is boosted in the e-learning. On the other hand, value rationality also appears breaking owning to the deviation or deformity of the subject's thought and behavior.

A.Weakening of self-control

First, reduction of self-control. Network filled with a lot of unhealthy information although its resources are rich and various. Learners often lose their control ability. Their original value reason is weaken or broken without a firm self-control. In reality, many e-learners finally slide into decadence and degeneration, and even crime

because they are addicting to internet and losing themselves,

Second, alienation from the real world. With the rapid development of network technology, virtual world becomes more and more realistic. If you indulge in the virtual world, you may be loss of the real emotions and value judgments, and then deep mud in the virtual world, or even indulge yourselves. This will lead to learners' alienation from the real world.

B. Weakening of rational thinking

While the network resources are rich, most of all is visualize, emotional, dynamic, and much more to those unhealthy information and resources. They mainly take the external sensory stimulation and experience as the important driving force. Learners who are lack of rational criticism can easily be undermined and even deprived of rational thinking. A large number of non-rational information will eat up and misappropriate learners' brain space and value rationality.indulging in online games and unable to extricate themselves are good examples.

C. Loss of self-criteria of value

Network is a "double-edged sword." How to use it correctly lies primarily in learners' value reason. Correct value reason will guide and regulate learning goals, process and effect to facilitate the development of the subject. On the contrary, they will be lost and go wrong. Some students cannot bear too much tempatation from network and then lose their value reason, with the result that studies is neglected.

In the e-learning, unhealthy and broken value retionality of the subject is the main reason that results in the deviation or abnormal development of their ideas and behavior.

IV CONSTRUCT HARMONY BETWEEN VALUE REASON AND E-LEARNING

It is one of major issues to build up harmony between value reason and e-learning, which is worthy of in-depth reflection on. It requires value reason should be applied in the e-learning. Based on current situation, the following three countermeasures are put forward.

A. Bring the subjectivity into full play correctly

First, we should adhere to and implement the ideas of man's subjectivity by Marxism. It is an important core of Marxism theory and a fundamental theoretical basis for the response to this problem.

Marxism holds that " The person‘s essence is not single and personal proper and abstract thing, on its actuality, it is all the totals of the social relation". The internal regulations of human have been the most essential grasp. Marxism fully affirmed that people have played a central and dominant role in the social and historical development. Marx pointed out that "the so-called entire history of the world is nothing but a history about humans' labor, which is a process of the evolution of man by nature". That is why Marxism historical materialism is essentially a conception of history and development with a core of human and its subjectivity.

The View of Marxism Practice develops around human and its subjectivity. The view of practice is the chief basic view of Marxist philosophy. Marxism holds that practice has a strong objectivity and human is the subject. The autonomy of practice is embodied that human is not only able to understand objective laws but also use it in accordance with his wishes and needs in practice. Meanwhile, practice also has obvious creativity which represents that he can create various things even beyond the laws of nature. Human creativity highly displayed subjective initiative which reflects human have played a dominant role in practice. Autonomy and creativity together constitute the human subjectivity. Human subjectivity can constantly develop and enhance with the development of the practice.

Secondly, we should norm learners' value choice. It is the premise to bring learners' subjectivity into full play correctly. Learners must consciously make the correct value choice and study in accord with their own demand. And they must reflect and regulate actively in case of deviation in this process.

Thirdly, we should train learners' critical consciousness and creative ability. Marxism holds that critical consciousness and creative ability are the core to bring the subjectivity into full play correctly. Learners should enhance their conscious analysis and judgement abilities to develop their self-awareness, initiative and independent. Faced the complicated network resources, rational criticism on the level of value, distinguish the true from the false, and then choose proper information and knowledge for self-development. In particular, they could criticize unhealthy information to strengthen correct value judgments and choices. Innovation is an important manifestation of subjectivity. In network learning, learners should sort and process the information they received to create new knowledge and explore new ideas, and then continuously self-develop on the basis of the right specifications.

B. Strengthen effective monitoring for e-learning

The important feature of e-learning is the learners' subjectivity. It means that we must pay attention to learners' effective monitoring for the organic unity and internal harmony between value reason and e-learning.

First, develop correct and reasonable learning goals and plans based on objective analysis of their study habits and value demand. Correct and reasonable goal will determine what kind of information learners choose and enhance direction and effectiveness. The correct and reasonable plan is to regulate the learning activities in the process.

Second, pay attention to reflection and regulation of learning process. In fact, the effective control of e-learning focused on process monitoring. Because it is to implement learning goals and plan and finally to determine learning effectiveness. In the learning process, learners should take the initiative to self-reflect and self-test to see if their behaviors deviate from their

self-worth in the following aspects, such as, content selection, processing, exploration and innovation. If there is a deviation, they should regulate and amend in time. Many learners with network addiction are the bad case.

Third, evaluate learning effectiveness actively. Learners should actively evaluate learning progress. On the one hand, learners should evaluate if the information and knowledge they get is consistent with their needs to ensure the coherence between planning, process and results. On the other hand, learners take the initiative to communicate with others in order to test self-learning validation. Network communication is one way between teachers and classmates in real life. This part will affect learners' next step. If the outcomes don't meet their needs and expectation, they will have to re-learn and improve learning outcomes.

Forth, strengthen ideological and moral cultivation initiatively. In network learning, learners must consciously strengthen ideological and moral cultivation, and continually enhance the value pursuit and evaluation standards. This is a key factor to norm learners' behavior. First of all, e-learners consciously study the scientific theories of Marxism, especially Marxist ethics to equip their minds. Secondly, learners should consciously use moral rules to restrain own online learning. Thirdly, learners should have the courage to self-analysize and self-criticize and be good at self-improvement. More important is to self-develop, and then to continuously improve and enhance the ideological and moral levels and value judgments.

C. Pay attention to combine virtual space with real world experience

As a special way of learning, network learning is carried out in the virtual space. However, the virtual nature is essentially derived from the practice of social life, which is a virtual extension of the real world. Therefore, we could not stress the virtualization of learning space and platform in a single. Learners should pay attention to combine the virtual space with the real interactive experience. Learners should constantly exchange, verify and amend information by communicating with teachers and classmates initiatively. In addition, communication can also promote learners to ground in reality. Learners must not depreciate the real world experience, otherwise easily mire into the virtual network world.

CONCLUSION

In the e-learning, we must scientifically use value rationality to regulate the whole process, and then to pursue and carry out the organic unity and internal harmony between value reason and e-learning. In this way, we correctly use network resources and carry out scientific learning for the purpose of healthy development of the subject.

REFERENCES

[1] Marx and Engels. *Marx and Engels, Selected Works* (Vol.I) . Beijing: People Press, May 1972 edition, p18.

[2] Marx and Engels. *The Complete Works of Marx and Engels* (Vol.42). Beijing: People Press, September 1979 edition, p131.

[3] Zhang Jinshun, Lu Huiju. The main characteristics of e-learning. *Journal of Guangxi College of Education*,2006(5)

[4] Ye Min. The factors of web-based learning. *China Science and Technology Information*,2008

[5] Liu Xiling. The influence of web-based learning on undergraduates. *Private Science and Technology*, 2008 (10)

[6] Zhu Qingshan.The breaking and reconstruction on technological progress and human reason. *Guangxi Social Science*,2005 (11)

[7] Liu Xuelan, Liu Ming. E-learning and the development of human subjectivity. *Journal of South China Normal University (Social Science Edition)*, 2004 (1)

[8] Zheng Yuanjing. The value considerations of scientific and technological reason. *Journal of Hunan University of Arts and Science(Social Science Edition)*, 2006 (2)

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 109-111

Promoting E-learners' Self-monitoring with Mind Map

Pan Ziyan
College of Computer Science and Technology, Henan Polytechnic University, JiaoZuo, China,
zypan@hpu.edu.cn

***Abstract*—In the e-learning,we pay more attention to promote learners' self-monitoring ability and improve the quality and effectiveness of learning. However, learners' ability of self-monitoring is not satisfactory. In this regard, introducing mind map to enhance e-learners' self-monitoring is a very effective method. This paper interprets the application of mind map and the problems which should pay attention to.**

***Index Terms*—e-learning, self-monitoring, mind map**

INTRODUCTION

With the development of modern information technology and internet, network is increasingly becoming an important means for learning. However, the quality of e-learning is not satisfactory. There is a great difference between expectations and actual results for e-learning. Among the many reasons, learners' self-monitoring ability is an important factor which affects the e-learning process directly as well as to achieve good results. Therefore, improving the learners' self-monitoring capability is very critical for the quality of e-learning. Introducing mind map into the e-learning can promote learners' self-monitoring effectively.

PART I E-LEARNING AND SELF-MONITORING

E-learning also known as "Digital Learning" and "online learning", which not only breaks through traditional limits of time and space but also brings extensive learning resources for us. However, it makes learning fast and convenient, and at the same time has some negative effect. First, the separation of teachers and students leads to weak monitoring of teachers and learners in the network learning. Second, learners are prone to psychological problems such as loneliness and the lack of a sense of belonging because of the separation between learners. Third, massive information and its randomness and irregularity lead to the "Information-mazing" and "Information Overloading" phenomenon. Forth，internet, with its special virtual and open features, brings much convenience as well as anomie moral phenomenon. Based on these negative effects, it becomes extremely important to improve the capacity of learners' self-monitoring.

The capacity of learners' self-monitoring refers to learners' self-monitoring, self-regulation and self-awareness that actively regulates learning strategies according to their own characteristics and learning tasks. Learners with weak self-monitoring feel easily discouraged, frustrated, even disoriented. Learners with strong self-monitoring will consciously develop learning objectives and study plans, choose learning methods, actively regulate learning behavior, discipline themselves to finish their learning task on time，subject to social development and self-development.

PART Ⅱ THE CURRENT SITUATION OF E- LEARNERS' SELF-MONITORING

Due to the traditional education, many learners lack a certain autonomy, independence and self-control. Consequently, they are not suited to e-learning, particularly with regards to the capacity of self-control. Incorporate is as follows:

First, some learners are lacking in the ability, wishes and motivation in autonomous study, which results in the uneffectiveness. All of these make the students study passively, and then make many students feel that "learning is important but hard to study". Self-learning ability is a higher level of integrated learning, which includes good self-awareness and self-evaluation, cognitive processes of their own cognitive abilities, identificating all kinds of feedback information and regulation of the learning methods. There are also non-cognitive factors.

Second, learners' self-monitoring ability is prevalently weak. Learners are vulnerable to external interference and temptations because web resources are very mixed in terms of quality. The "Information-mazing" and "Information Overloading" phenomenon are low self-monitors' typical model manifestation. Those learners are unable to extricate themselves, lost in the virtual space, thinking down and seriously affected the learning owning to the network indecent information. In contrast, high self-monitors are people who often use certain criteria to assess their learning process, in order to adjust, guide, supervise, evaluate and feedback learning to achieve the desired objectives.

Third, self-evaluation needs to be strengthened. Self-evaluation is to evaluate and feedback own learning, which is the key of ongoing independent study. Network learning requires learners are fully responsible for their own learning, self–assess their learning process and outcomes. Due to the influence of the traditional education, many learners can not give full play to their own subjective dynamic role, and are not also good at self-evaluation in every aspect of learning which greatly affects the learning.

PART Ⅲ PROMOTING E-LEARNERS' SELF-MONITORING WITH MIND MAP

With reference to training the capability of e-learners' self-monitoring, many scholars have pointed out that the learning process requires learners to shift from external control to internal control, use search engines to retrieve information, integrate web resources, and establish a complete support system. How can we do that? It is an important way to train learners a conscious strategy of cultivating self-control capacity in the learning process. Introducing mind map into the e-learning is very effective for this, which aims to enhance the learners' self-monitoring ability.

1.Mind map and its features

Studies have shown that it is not much different in the capacity from person to person. The main differences are thinking mode, ways of thinking and their actions. Early 1960s, British psychologist Tony put forward the concept of mind map. It aims to express ideas orderly that people learn knowledge, solve problems and innovate by drawing pictures.

Tony pointed out that the biggest enemy of thinking is complex, the biggest obstacle is chaos. Mind map can make thinking clearly, visible and be tracked. It can help students concentrate, capture information more efficiently, see the "Panorama", and think orderly and efficient. At present, mind map has been widely used in various fields, and is also the focus that our education research pays close attention to.

2.Promoting learners ' self-monitoring with mind map

Self-monitoring including targeting, planning, selecting methods, time management, effort regulation, implementation, feedback and analysis, and take remedial measures. Mind map can help learners promote their self-monitoring from several following respects.

(1)Stimulate interests and carry out inquiry learning

It is a core task for traditional learning and e-learning to inspire learners' learning interests and motivation. Mind map can stimulate learners' interests. Firstly, the learners will constantly have new discoveries to improve the will and effort level in their effort to build mind map. It is a happy inquiry process that not only to find and solve the problem, but also to take the initiative to explore. In this way, Learners will mobilize their enthusiasm, initiative, curiosity to create a good learning state. Secondly, learners can clearly understand the basic concepts that constitutes knowledge and conceptual relationship by drawing. Third, mind mapping can help learners construct knowledge framework, identify their definite learning objectives, select one paragraph at a time. As a result, learners could easily focus attention and study effectively. Meanwhile, the perfection and refinement mind map process can reflect learners' level of understanding and mastery the knowledge. As the evaluating standard on the effect of learning, it can help learners provide timely feedback, improve learning efficacy and strengthen motivation.

(2) Take overall situation and grasp the details

To master a course, learners must firstly construct a complete framework of knowledge and form a mind panorama in order to enhance the overall grasp and make a reasonable schedule and progress based on the actual situation in the learning process. Using mind map in the e-learning is a process to construct knowledge structure and curriculum panorama. In this process, learners can constantly enrich and refine panorama, edit scattered points into the panorama. This will not only help learners corer the overall situation but also grasp the details and exclude irrelevant information.

(3) Full-process navigation and assisted learning

Effective learning strategies are usually "K-W-L" strategy. K stands for what learners had known before learning new knowledge, that "know"; W stands for what learners want to learn, that "what"; L stands for what the students learned in the end, that "learned" . In web-based learning, mind map can be used for K, W, L phases. In the K phase, mind map can help us collect all the relevant ideas, knowledge and experience by divergent thinking to facilitate the learning of the subject. In W phase, it is used to construct a system of knowledge to order and deepen the learning content while learners brows the web. In the L phase, Writing mind map from memory over and over again could make up the deficiencies, hold the essentials, consolidate the knowledge. Meanwhile, mind map has a variety of design templates. Learning can select the most appropriate characterization to organize information to carry out the best learning result according to their type of thinking and previous experience. For example, concept and character description could use style 1; the grasp of the causes and results could use styles 2.

(4) Collate information and aggregate points

Web resources are very rich. Massive information easily lead to "Information-mazing" and "Information Overloading". It is a challenge for many e-learners to efficiently organize and grasp the information. Our ultimate goal is to understand the information and get the desired result, not to access and share information. Therefore, learners must have the information literacy of analysis, processing and innovativation. Mind map is a very useful tool for this. It is not only beneficial to learners order all kinds of knowledge, information and ideas, clear its internal hierarchy, and hence make use of these resources, but also to understand the learners what to think, how to think, and what factors affect the entire process for the purpose of promoting them to reflect their knowledge acquisition process to improve analysis and problem-solving abilities.

(5) Provide feedback and evaluate effectiveness

E-learners must consciously look for feedback and evaluate effectiveness because it has less external monitoring and feedback. The process that learners draw up mind maps again and again effectively reflects on the variation diagram that learners understand and master new knowledge. It could help them accurately evaluate their own learning effect, make up the deficiencies and grasp the key points, and then take remedial measures to adjust the learning plans and learning strategies based on

the results of feedback.

3.Some attentive questions

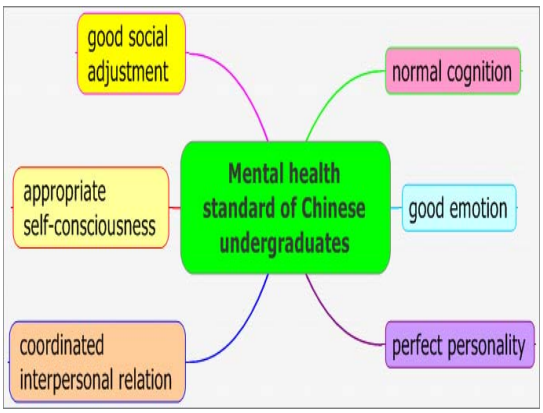
(1)To implant mind map into learners' brain. Give a man a fish and he ends for a day. Learning of learn is the most important learning. It is the most important to foster the ability to study on their own. Such an open autonomous learning environment demands higher self-monitoring for learners. It needs to complete from the dependence on external control to internal control for the learners who are accustomed to the traditional learning model. This requires teachers to guide and help. To this end, we first introduced mind map into the minds of learners to demonstrate how to use mind maps to motivate, organize knowledge, monitoring learning, feedback in the whole process of learning. Thereafter, guide the students consciously use mind maps to lead all aspects of learning to achieve good learning results.

(2)To prefect the learning support system. Learning support services plays an important role in the e-learning. For most learners, they may not achieve the expected learning outcomes without enough support and help. General assistance is not enough. The majority of learners need more individualized help. They need to get niche targeting feedback on their concerns. Thus, to provide full, comprehensive, timely, and convenient learning support services is very important to train e-learners of good study habits. In addition to helping the learners to navigation, inquiry, advice, guidance and coaching, online examination and management, it is particularly important to nurture learners' autonomy motivation and promote their conscious self-monitoring in order to gradually adapt to e-learning.

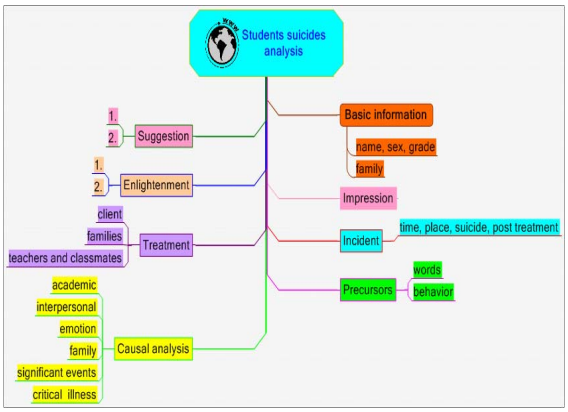
(3)To strengthen the sense of self-study. Even if we have embedded mind map into learners' mind, introducted it into the learning process, and constantly perfect the learning support system, we are still unable to achieve good learning results if the learner is not a good sense of self-learning. The internal cause is the thing change basis, the external factor is the thing change condition, the external factor has an effect through the internal cause. Therefore, e-learning must primarily emphasize learners' autonomy and creativity so that learners become the subject of information processing and the active constructers of knowledge.

CONCLUSION

In short, it is a very effective method that promote e-learners ' self-monitoring with mind map, thereby improve the quality and effectiveness of e-learning.



styles1



styles 2

REFERENCES

[1] Du Xiangli, Cui Jia. The Training of Students' Self-monitoring in Network Circumstance. *Journal of Guangzhou Radio & TV University*,2007(1): 13-17

[2] Ren Ruixian, Feng Xiuqi. Analysis of the factors influencing self-controlled e-learning. *Open education research*,2004(1):33-37

[3] Chen Xiaozong. Study of network-based autonomous learning system. *Education and Vocation*,2009(11)

[4] Liu Xiaoning. A research review on mind map in China. *Journal of SiChuan College of Education*,2009(5):

[5] Xu Yeping. The application of mind mapping in education. *Science & Technology Information*,2008(27)

[6] Gao Li, Meng Suhong. The Application of Mind Mapping in education teaching. *China Modern Education Equipment*,2007(6)

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 112-115

An Improved Clustering Algorithm

Tianwu Zhang[1], Hongshan Qu[2]
[1] Computer Science & Engineering Department, Henan Institute of Engineering, Xinzheng, China
Email:xxzhtw@163.com
[2] Computer Science & Engineering Department, Henan Institute of Engineering, Xinzheng, China
Email:qhs@haue.edu.cn

***Abstract*—This paper introduces an improved clustering algorithm GCA(Gravitational Clustering Algorithm), it is extended in such a way that the Gravitational Law is not the only law that can be applied. This algorithm can decide automatically the number of clusters in the target data set, and find any clusters with arbitrary forms and filter the noisy data. The experimental results show that GCA algorithm creates high quality greatly.**

***Index Terms*—clustering, clustering algorithm, gravitation**

I. INTRODUCTION

Cluster analysis has recently become a highly active research branch in data mining and been widely used in many fields such as market research, pattern recognition, data analysis and image processing[1]. Clustering is an unsupervised learning technique that takes unlabeled data points and classifies them in different clusters. A cluster is a collection of data objects that are similar to one another within the same cluster and are dissimilar to the objects in other clusters[2]. The approaches for data clustering which have been worked out so far can be classified in five broad categories: partitioning method, hierarchical method, density-based method, grid-based method, model-based method[3]. These algorithms mostly classify the types with distance and density, which results in the flaws that they cannot express the directions between dots and embody spontaneously the collecting and loosing relations of the dots precisely, and cannot indicate the multiple relations between the dots. To solve the problems mentioned above, this paper introduces Gravitation and Newton's Second Law of Motion into the process of clustering, and proposes an improved algorithm GCA(Gravitational Clustering Algorithm), and verifies this hypothesis with experiment about the efficiency.

II. RELATED THEORY

Let x be an object in the n-dimensional euclidean space, that is moving in the direction given by the vector $\vec{d}$, and t be a real number representing an instant of time. Let x(t) be the object position at time t, v(t) be the object velocity at time t, ant $\overrightarrow{a(t)}$ be the object acceleration at time t.

corresponding author: Tianwu Zhang. Tel: 13526616958; Email: xxzhtw@163.com.

AP-PROC-CS-10CN007

A. Gravitational Law

The force exerted from one object x over another object y is expressed by the following equation:

$$F(t)=\frac{Gm_x m_y}{d(x(t),y(t))^2} \tag{1}$$

B. Newton's Second Motion Laws

If m_x is the mass of the object x, then the force exerted on the object is defined according to Newton's Second Motion Law as follows:

$$F(t)=m_x a(t) \tag{2}$$

According to Gravitational Law and Newton's Second Motion Laws, the acceleration vector y can be deduced as follows:

$$\overrightarrow{a(t)}=\overrightarrow{d(t)}\frac{Gm_x}{|\overrightarrow{d(t)}|^3} \tag{3}$$

The speed and position of the object at time t+Δ(t) are approximated as follows:

$$v(t+\Delta(t))=v(t)+\overrightarrow{a(t)}\Delta(t) \tag{4}$$

$$x(t+\Delta(t))=x(t)+v(t)\Delta(t)+\frac{\overrightarrow{a(t)}\Delta(t)^2}{2} \tag{5}$$

Because the acceleration vector function is a complex equation, to find the position function of a given object under the influence of one or more gravitational fields is difficult task. As a result, the movement of an object is approximated by using the acceleration vector given by (3) in the equations (4) and (5). Therefore, the movement equations of an object y under influence of the gravitational field of an object x are:

$$v(t+\Delta(t))=v(t)+\vec{d}\frac{Gm_x}{|\vec{d}|^3}\Delta(t) \tag{6}$$

$$y(t+\Delta(t))=y(t)+v(t)\Delta(t)+\overrightarrow{d(t)}\frac{Gm_x\Delta(t)^2}{2|\overrightarrow{d(t)}|^3} \tag{7}$$

C. Optimal Disjoint Set Union-Find Structure

A disjoint set Union-Find structure has many similarities with clustering. It is a structure that supports the following two operators[4]:

- Union(A,B,C): Replace the two sets A and B by their union set C.
- Find(x): Return the name of the set containing the element x.

D. The mass of data points

GCA is a clustering algorithm based on Gravitational Law and Newton's Second Motion Law. In this way, for an n-dimensional data set with n data points, each data pint is considered as an object in the n-dimensional space with mass equal to 1[5].

III. PROPOSED APPROACH

The basic ideas behind applying the GCA algorithm are:

- A data point in some cluster exerts a higher gravitational force on a data point in the same cluster than on a data point that is not in the cluster. Therefore, points in the same cluster move in the direction of the center so that GCA can determine the number of clusters in the data set.
- A noise point does not belong to any cluster because the gravity produced by other points is too weak to attract it. The point keeps almost intact and will not be assigned to any cluster.
- The terminal condition of algorithm is one agreed iterations. After several rounds of polymerization, the number of clusters in the target data sets has been fixed and would remain unchanged.

To improve the algorithm's execution efficiency, the equation (7) can be simplified as: Let $\Delta t=1$， $v(t)=0$，$\frac{G}{2}=1$. Given the possibility that excessive movement of data points will lead to less number of clusters, the decreasing function $f(x)=\frac{1}{x^3}$ was introduced.

And finally, the equation for the data point y affected by x is:

$$y(t+1)=y(t)+\vec{d}*f(|\vec{d}|) \quad (8)$$

The proposed clustering algorithm can be described as:

Input: N different data points and related parameters R, ε

Output: different clusters

GCA(*R*, *ε*)

1）for *i*=1 to *N* do

2） Initial(*i*); //*create a new set containing the single element i*

3）for *i*=1 to *R* do //*R is the number of iterations, as the algorithm termination conditions*

4） for *j*=1 to *N* do

5） begin

6） *k*=random point index and $k\neq j$;

7） MOVE(y_j,y_k)； //*see equation (8)*

8） If $dist^2(y_j,y_k)\leq\varepsilon$ then Union(*j*,*k*,*k*)； //*ε is the smallest distance value*

9） end

10）for *i*=1 to *N* do

11） Find(*i*);

12） return disjoint-sets

GCA algorithm assigns every point, normal or noisy, to different clusters. We use a function Get-clusters to disjoint sets generated by GCA and return the collection of clusters that have at leas the minimum number of points [6].

Get-clusters algorithm is:

Input: Intermediate clustering results

Output: Final clustering results

Get-clusters(*clusters*,*α*)

1）newclusters=ϕ;

2）Minpoints=α； //*α is the minimum number of points that a valid cluster should include*

3）for i=0 to number of clusters do

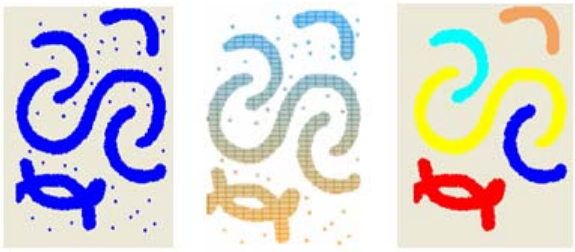
4） if size($cluster_i$)≥ Minpoints then

5） newclusters= newclusters ∪{ $cluster_i$}

6）return newclusters

IV. EXPERIMENTAL RESULTS AND ANALYSIS

Experimental environment: Pentium Ⅳ 2.93G CPU,



(a)Original data set (b)Result of K-means (c) Result of GCA

Figure 1 Comparison of algorithm K-means and GCA

(a)Original data set (b)Result of K-means (c) Result of GCA

Figure 2 Comparison of algorithm K-means and GCA

RAM 1G, the operating system is windows XP，the algorithm is written in C#.

GCA algorithm's time complexity is O (N), N is the number of data points. It is suitable for algorithm to be applied on large data sets because its time complexity is a linear function.

In order to evaluate the performance of GCA, experiments were performed on two synthetic data sets which are included in the literature [7].

The data set in Figure 1(a) contains 11,182 points, 5 natural clusters. Figure 1(b) shows the results obtained by K-means. Figure 1(c) shows the results obtained by GCA. The values for GCA parameters were: R=400, ε=1e-4，α=50. The comparison of two figures shows that GCA algorithm is better than K-means. GCA can decide automatically the number of clusters in the target data set, and find any cluster with arbitrary form and filter the noisy data. On the other hand, the clusters obtained by K-means include some noisy points and was very sensible to be disturbed.

The data set in Figure 2(a) contains 12,917 points, 3 natural clusters. Figure 2(b) shows the results obtained by K-means. Figure 2(c) shows the results obtained by GCA. The values for GCA parameters were: R=500, ε=1e-4，α=50.The comparison of two figures shows that GCA algorithm is better than K-means. GCA can decide automatically the number of clusters in the target data set, and find any cluster with arbitrary form and filter the noisy data. On the other hand, the clusters obtained by K-means include some noisy points and was very sensible to be disturbed.

KDD Cup 99 data set is the data set used for The Third International Knowledge Discovery and Data Mining Tools Competition, which was held in conjunction with KDD-99 The Fifth International Conference on Knowledge Discovery and Data Mining. KDD training data set consists of approximate 4,900,000 simulated attack records, and each of which contains 41 features. We select a subset T containing 20,000 normal and 80,000 attack records and four relevant attributes of each record. Table 1 shows the results obtained by GCA and K-means.

As is shown in table 1: GCA generated 18 clusters and K-means generated 25 clusters. It shows that GCA's clustering effect is more concentrated than the K-means's and the effect of gravity was expressed well. And also, GCA improved the clustering quality in the way that the proportion mixed by normal and attack records which obtained by GCA was smaller than K-means. Although GCA's execution time was affected by using gravity, the total execution time of GCA and the K-mean are almost equal because that GCA can get less number of clusters.

TABLE I.
COMPARISON OF ALGORITHM K-MEANS AND GCA

Cluster	GCA		K-means	
	normal	***attack***	***normal***	***attack***
1	9	42624	18	42631
2	4	13843	7	12709
3	14592	214	12526	824
4	52	2541	0	133
5	0	67	35	78
6	1	71	4	34
7	3	31	0	18
8	1	23	3	17
9	1	18	5	19
10	1653	21	1738	118
11	783	11	639	153
12	46	2	68	19
13	184	5	197	25
14	2359	13	2292	27
15	0	78	16	6
16	9	2368	62	2467
17	227	16886	1785	19619
18	76	1184	586	1032
19			2	6
20			1	7
21			6	10
22			5	9
23			0	13
24			3	11
25			2	15

V. CONCLUSIONS

A new clustering algorithm was presented in this paper. Several experiments with synthetic data sets and with a real data set were performed in order to show the performance of the proposed approach. The proposed approach will not be affected by the noisy data, and it can find clusters with any regular shape without allocating the number of the clusters which make up for the former drawback in partitioning method. This algorithm is superior than classical K-means in both effect and precision and improve the quality of clustering. Although the performance reached by GCA algorithm can be affected by the parameters *R* and ε.

REFERENCES

[1] Jiawei Han, Micheline Kamber, Data Mining：Concepts and Techniques, Data Mining：Concepts and Techniques, 2nd ed. Beijing: China Machine Press, 2007, pp. 383–384.

[2] Huizhe Zhang, Jian Wang, "Improved Fuzzy C Means Clustering Algorithm Based on Selecting Initial Clustering Centers," Computer Science, vol. 36, Jun. 2009, pp. 206–209.

[3] Xiaoqin Tang, Ruyuan Dai, “Technique of Cluster analysis in Data mining,” Computer Science, vol. 19, Jan. 2003, pp. 3–4.

[4] Alfred V. Aho, John E. Hopcroft, Jeffrey D. Ullman. The Design and Analysis of Computer Algorithms. Beijing: China Machine Press, 2007, pp. 124–125.

[5] Baozhi Qiu, Feng Yue, “Gravity-based Boundary Points Detecting Algorithm,” Journal of Chinese Computer System . vol. 29, Feb. 2008, pp. 279–282.

[6] Jonatan Gomez, Dipankar Dasgupta, and Olfa Nasraoui, “A new gravitational clustering algorithm,” in Proceedings of the Third SIAM International Conference on Data Mining 2003. SIAM Press, May. 2003, pp. 83–94.

[7] George Kapis, Eui-Hong(Sam) Han, Vipin Kunmar. “CHAMELEON: A Hierarchical Clustering Algorithm Using Dynamic Modeling,” IEEE Computer, vol. 32. Aug. 1999. pp. 68–75.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 116-119

Characteristic Investigation Impulse Radiation of Two UWB Antennas

Li Bao-ping [1], Wang Yan [2]
[1] College of Computer Science & Technology, Henan Polytechnic University, Jiaozuo Henan, China
Email : libaoping@hpu.edu.cn
[2] Wanfang College of Science & Technology HPU, Jiaozuo Henan, China
Email : wywf@hpu.edu.cn

***Abstract*—This paper compares the characteristics in the time-domain of Ultra-wide Bandwidth (UWB) monopole antenna and wide slot antenna. Whether the UWB antenna is suitable for the signal transmitting is not only measured by antenna's width but also by the characteristics of the time domain. These two antennas are proved to be fit for sending and receiving signals in frequency domain. In this paper, the relationship of the excitation and the radiation signals is studied in time domain. And the distortion and fidelity results of the two kinds of antenna are compared.**

***Index Terms*—UWB, monopole antenna, wide-slot antenna, time-domain characteristics, fidelity**

I. INTRODUCTION

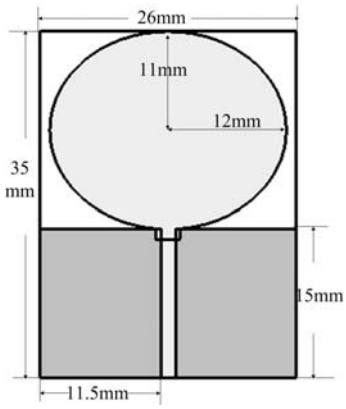
UWB systems usually use UWB short electromagnetic pulse, and the single pulse signal has two prominent characteristics. One is the excitation signal waveform with a steep leading edge, the other is the excitation signal with an ultra wide bandwidth, from DC to microwave band. So they must be UWB as transmit antennas or receive antennas [1]. The fidelity in time-domain as the most important characteristic of UWB antennas is defined as the maximum correlation of the normalized input voltage and the normalized electric field strength in far region[2]. In order to reduce the radiation ultra narrow pulse waveform distortion, minimize the frequency dispersion and spatial pulse dispersion, the mid-phase in whole work frequency band of UWB antennas must be kept unchanged. When an ultra short pulse (UWB signal) is used to stimulate antennas, there will appears ringing effect, so the response signal has dispersion in time domain. How to avoid UWB antennas ringing effect is one of the bottlenecks in need to resolve [3].

The following two characteristics of antennas in time domain were studied usually: 1. the pulse response, which is an important indication of the UWB antennas performance. 2. The radiation signals wave form with different angles. 3. The antennas gain with different frequency. How to insure the antennas gain in whole ultra wide bandwidth is one of the most important factors in UWB antennas design [4].

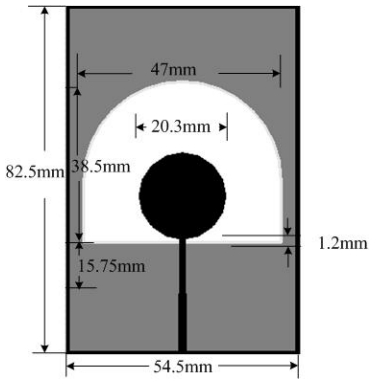
II. THE SIMULATION STRUCTURE AND TOOL

In order to design UWB antennas with excellent performance, we usually used numerical simulations add actual measurement based on antenna theory. Because of the antennas structure is complex and diverse, it is difficult to use analytical method to the theoretical prediction. At present, the numerical methods are finite element method, finite integral time-domain (FDID) method, finite difference time domain (FDTD) method, and moment method (MoM). In above methods, FDID and FDTD methods are better to analyze antennas characteristics of time-domain. The CST simulation software used in this paper is based on FDID method.

In this paper, we take the short-pulse voltage signal fed monopole antenna and microstrip feed line wide-gap antenna as the examples, the antennas structure shown in Figure 1.



1-1 the Monopole Antenna Structure.



1-2 the Wide-Gap Antenna Structure

Figure 1. the Antennas Structure

Li Bao-ping (1981-), Lecturer, Research area: Mobile communication and radio frequency technology.

This monopole antenna using a new type of feed matching technology, on the back of the ground plane opened a gap corresponding to the positive microstrip feeder line. Its role is to regulate coupling between positive radiation element and negative ground units to a certain extent. The antenna's bandwidth would be broadened when choose right gap size. The wide-gap antenna choose semi-circular and circular feed microstrip line for enhancing coupling between the wide-gap and the feed patch.

III. COMPARISON OF FREQUENCY-DOMAIN CHARACTERISTICS OF THE TWO ANTENNAS

As shown from Figure 2, the monopole antenna and the wide-gap antenna is similar to work in 3-10GHz, the following radiation field analysis about the two antennas is based on similar work band.

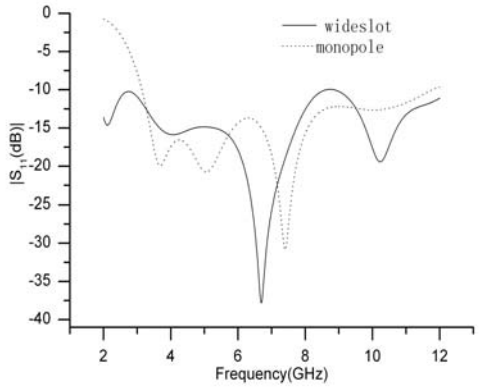


Figure 2. Comparison of work frequency of the two antennas

IV. TIME DOMAIN ANALYSIS OF THE RADIATION FIELD

In this simulation, the excitation signal is Gaussian pulse, the frequency range 3-10GHz, time-domain waveform shown in Figure 3.

The radiation waveforms with 30m distance in different angles can be obtained through CST software simulation, as shown in Figure 4.

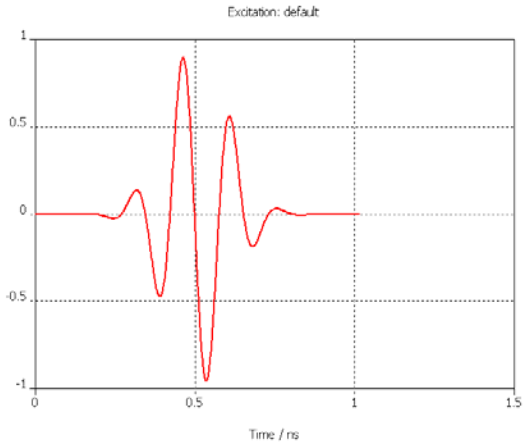


Figure 3. Excitation signal in time-domain

From the Figure 4 we can see, the monopole antenna and the wide-gap antenna "ringing" are not very serious, and suitable for UWB system. The following calculated the antennas waveform fidelity factor, the fidelity factor f is the maximum correlation value with the normalized radiation pulse p and the excitation p0. Mathematically expressed as

$$f = \max_{\tau}\left(\int_{-\infty}^{\infty} p(t)\,p0(t+\tau)dt\right) / \sqrt{\left(\int_{-\infty}^{\infty} p^2(t)dt\right)\left(\int_{-\infty}^{\infty} p0^2(t)dt\right)},$$

so the radiation field signal must be normalized when calculate the fidelity factor. The fidelity factors are shown in table 1 through MATLAB programming calculation.

TABLE I. THE MONOPOLE ANTENNA ANT THE WIDE-GAP ANTENNA FIDELITY

Angle/Fidelity factor	monopole	wide-gap
0，90	0.98501	0.95196
90，90	0.97095	0.88736
180，90	0.98291	0.94872
270，90	0.94547	0.90607
15，90	0.98619	0.95033
30，90	0.98095	0.99325
45，90	0.97425	0.98503
60，90	0.97099	0.9732
75，90	0.9397	0.97074
105，90	0.96708	0.9686
120，90	0.97146	0.95915
135，90	0.97446	0.97988
150，90	0.97934	0.98929
165，90	0.98447	0.9385
90，0	0.81703	0.75626
90，90	0.97095	0.88736
90,180	0.81703	0.75626
90,270	0.94547	0.90607
90，15	0.83155	0.89545
90，30	0.90844	0.91765
90，45	0.94712	0.90776
90，75	0.96487	0.91068
45，45	0.92584	0.77214

V. SUMMARY

This paper analyzes the two antennas distortion under the same excitation in the time-domain. As table 1 show, these two antennas have good fidelity characteristics for the 3-10GHz Gaussian pulse. When the frequency range is extend to 2-12GHz, the fidelity decline significantly. Thus, waveform fidelity factor with excitation pulse shape is sensitive. During the research, we found that the

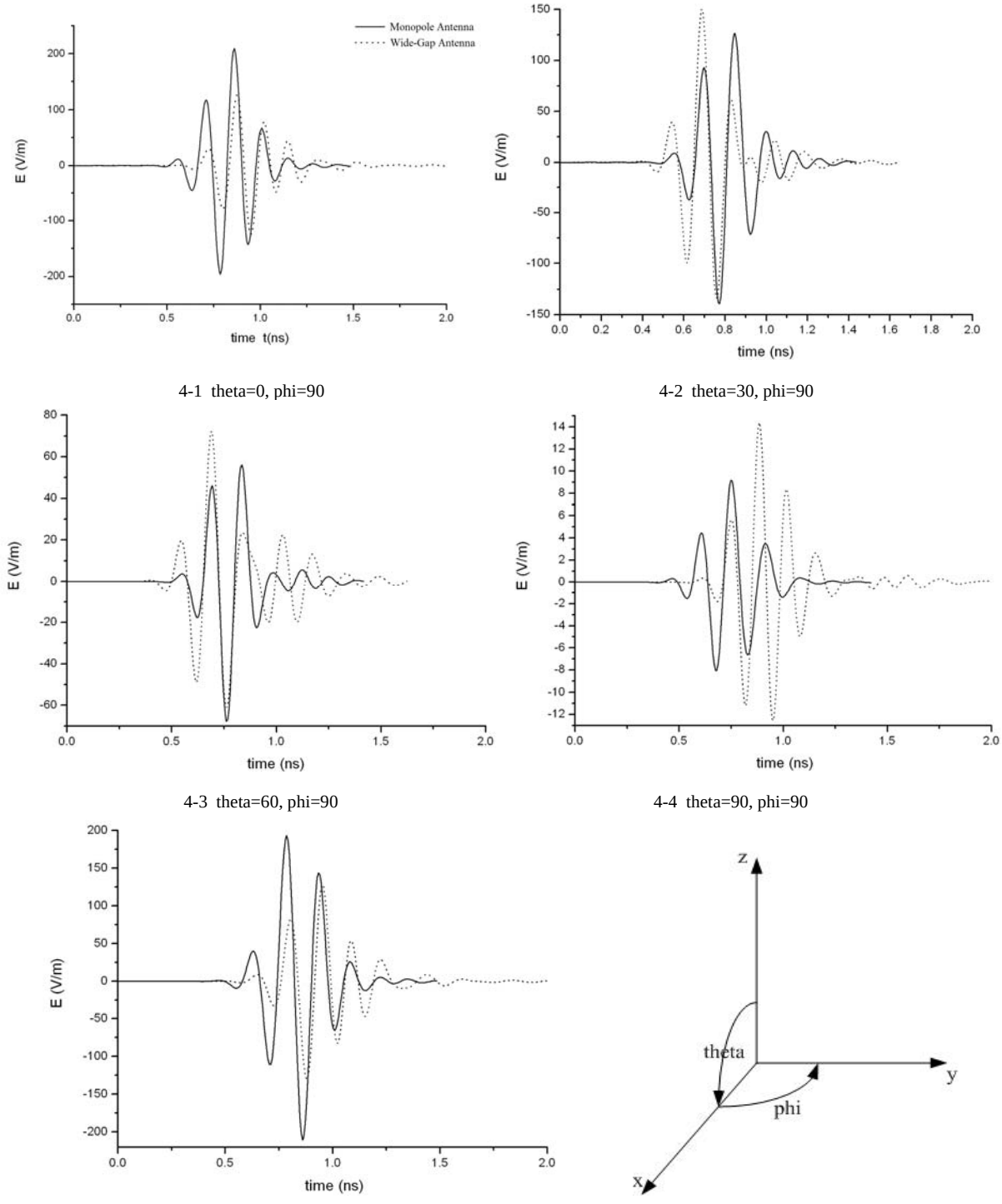


Figure 4. the Radiation Field of the Monopole Antenna and the Wide-Gap Antenna in Different Angles

distance change has small influence in fidelity factor with the same antenna and in the same direction, and the different antenna structure will affect the fidelity. In most directions of radiation, the monopole antenna fidelity is better than the wide-gap antenna. Calculated by the other antennas, we found that the symmetrical structure antenna fidelity is better than asymmetrical structure antenna. During the research in radiation field, we found that the main radiation fidelity superior to the other directions.

REFERENCES

[1] Z.Q. Peng, *Transient electromagnetic field*, Higher Education Press.

[2] D. Chen and C. H. Cheng, *A Novel Ultra-Wide Microstrip-Line Fed Wide-Slot Antenna*, Microwave and optical technology letters, vol. 48, No. 4,April 2006.
[3] C.Y. Huang and W.C. Hsia, *Planar elliptical antenna for ultra-wideband communications*, Electronics Letters 17th March 2005 Vol. 41 No. 6.
[4] Sergey N. Makarov, *Antenna and EM Modeling with MATLAB*, Beijing University of Posts and Telecommunications Press.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 120-123

Research of Campus Heterogeneous Database Middleware Based on SOA

Song Haige[1], Zhang Zhibin[2]
[1]College of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, China
Email: shg@hpu.edu.cn
[2]Department of Modern Education Technology, Henan Polytechnic University, Jiaozuo, China
Email: Zhangzhibain@hpu.edu.cn

***Abstract*—Along with the rapid development of campus information technology, many universities precipitates a great deal of information resources, how well make use of existing resources to build a centralized information resource management system is a serious problem. This paper designs a deep integration architecture of campus information based on SOA, adopts middleware technology to implement the integration scheme that highly requires a real-time data exchanging. The paper constructs a central database, which can synchronize its data to the corresponding application databases. Through the secure and reliable public data exchanging, all applications can be integrated on the basis of sharing public data and the data integrity, the data accuracy and the data consistency can be effectively ensured .**

***Index Terms*—SOA; central database; data synchronization; middleware; public data exchanging**

I. INTRODUCTION

With the rapid development of information technology in universities, many universities purchase and develop a number of applications after decades of information construction at the same time a number of information resources are precipitated. However, all these applications cannot be exchanged via the Internet and cannot share data eventually forming many islands of information, leading to repetitive construction and work. Therefore, the depth of information integration construction must integrate existing information resources and develop new resources, build a centralized information resource management mechanism to ensure that all applications can share data and achieve a real-time exchanging.

Data with different data sources and heterogeneous platform interfaces can be described in a unified and transparent data mode through the coupling way. System resources can be connected, integrated and collaborated though Web service in the process of data integration, as is a serious problem to be solved in the research of campus information integration platform. SOA (Service-Oriented Architecture) that has coarse-grained, loosely coupled, composite structure based on Web Services in particular provides a new solution for applications and data integration.

In this paper, the design of a deep degree of integration architecture for campus information based on SOA is proposed. New integration project can realize highly a real-time data exchanges through middleware technology. The paper constructs a central database , which can synchronize its data to the corresponding application databases. Through the public data exchanging system, all applications can be integrated on the basis of sharing public data.

II. SOA ARCHITECTURE

A. SOA Architecture

Service-oriented architecture is a component model [1], it can connect applications with different functional units (called services) through these well-defined interfaces and contracts. Interface that is defined in a neutral way can be independent of the handwork platforms on which they run, or the operating system and programming languages in which they are written, so that services based on such system can interact in a uniform and common way. SOA is regarded as an architectural style that emphasizes implementation of components as modular services that can be discovered and used by clients.

1) Services mainly have the following characteristics:

a) Services may be individually useful, or they can be integrated—composed—to provide higher-level services. Among other benefits, this promotes re-use of existing functionality.

b) Services communicate with their clients by exchanging messages: they are defined by the messages they can accept and the responses they can give.

c) Services can participate in a workflow, where the order in which messages are sent and received affects the outcome of the operations performed by a service. This notion is defined as "service choreography."

d) Services may be completely self-contained, or they may depend on the availability of other services, or on the existence of a resource such as a database. A service might perform a task without needing to refer to any external resource, or it may have pre-loaded all the data that it needs. Conversely, a service that performs currency conversion would need real-time access to exchange-rate information.

B. Three Important Roles of SOA

SOA architecture is composed of service provider, service requester and service registry [2]. Basic operations include service registration and publication, service discovery and binding, as shown in Fig. 1. Service provider publishes service information to service registry.

Henan Province Soft Science Research Project(No. 102400450064), Song Haige, 1979, female, han, Nanyang, Henan, lecturer, research area: computer application.

Service requester locates a service that meets its needs through searching for service. Once service requester search for the suitable service, it will directly activate the service according to the description of information in registry .

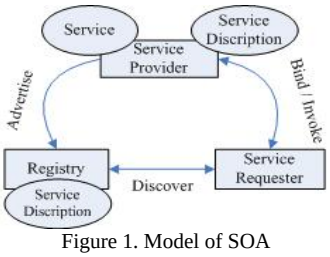


Figure 1. Model of SOA

1) Service requester: Service requester may be an application, a software module or another service in need of other services. It initiates a service inquiry from registry and binds a suitable service , then invokes the service according to interface contract.

2) Service provider: Service provider, which may be a network addressable entity, accepts and implements the user's request. It will publish own services and interface contracts to service registry so that service requester can discover and access the services.

3) Registry: Registry is a supporter of service discovery. It contains a repository of service and allows interested service requester to search for and access service provider's interface.

Ⅲ. WEB SERVICE

A. Web Service Concept

With the development of the Internet and relative technology, Web Service [3] is a product in a certain developing stage. Web Services have the interoperability on the complete different platforms, which is intended to achieve interoperability among all the applications through the Web standard.

Web services are modular components that may provide information to applications rather than to humans, through an application-oriented interface in a web environment. The information is described using standardized XML, so that it can be parsed and processed easily rather than being formatted for display.

Web services publish details of their functions and interfaces, but they keep their implementation details private; thus a client and a service that support common communication protocols can interact regardless of the platforms on which they run, or the programming languages in which they are written. This makes Web services particularly applicable to a distributed heterogeneous environment.

B. Web Service Key Technology

The key specifications used by Web services are:

1) XML(eXtensible Markup Language)—a markup language for formatting and exchanging structured data. XML language can transform data with different formats into the same structure and provide a unified data format for web service.

2) SOAP(originally Simple Object Access Protocol, but technically no longer an acronym)—an XML-based protocol for specifying envelope information, contents and processing information for a message.

3) WSDL(Web Services Description Language)—an XML-based language used to describe network service, or endpoint. A WSDL document can be used to dynamically publish Web service, to find a published Web Service and bind Web Service.

4) UDDI(Universal Description, Discoveryand Integration) —a soap-based client function for a framework for describing and finding a web service. UDDI can access the agreements of registered information through registry.

Many other protocols that focus on security, asynchronous communication and semantic expression are gradually being added to Web service.

Ⅳ. DATA INTEGRATION BASED ON SOA

A. Data Integration Technology

Typical data integration solutions can be divided into two categories: One is the materialized method, and the other is the global model method.

Data Warehouse belongs to the materialized method whose integration strategy is to pre-process and convert data copies coming from several heterogeneous data sources according to a centralized and unified view requirement in order to conform with the model of data warehouse. The data-sharing integration of heterogeneous database based on the data warehouse model has the advantage of on-line analysis and data mining. Disadvantage is the duplicate storage and updating difficulty of data. This method generally applies to large enterprises for analyzing its vast historical data .

Middleware system is a global view, which presents a global model in the middle layer to hide data details of the underlying layer so that Integration of data source is regarded as a unified whole by users. The actual data is not stored in the middle layer under this system. It is suitable for the integration environment that is relatively fast for the speed of updating data, is impossible or difficult to load all data from data sources. When user submits a query statement, middleware will separate it and send it to the underlying servers with different data source. Because the difference in the types of data source servers, it can complete the consistency of service interface for data sources of heterogeneous databases through wrapper function layer. User's query based on global model do not need to know the characteristics of each data source, middleware will divide query statement into sub-queries based on each local model of data source. Data integration project based on middleware, because of the advantages of real-time data exchanging and flexible scalability, is widely adopted. Considering middleware technology, the design of data integration layer based on SOA architecture is proposed.

B. System Architecture

System architecture [4] is divided into user layer, data

integration middleware layer and resource layer, which is shown in Fig. 2. Each data accessing operation that is initiated by transaction service can invoke this software layer. User layer may access the bottom of heterogeneous data source through an application interface that is provided by the middle layer. Data integration middleware layer, which is core layer of this architecture. This paper adopts the standard SQL language as query language to eliminate the difference between data sources of heterogeneous databases in the underlying layer. Resource layer, the bottom layer is mainly used to store and manage persistent data whose type may be text files, XML documents, relational database.

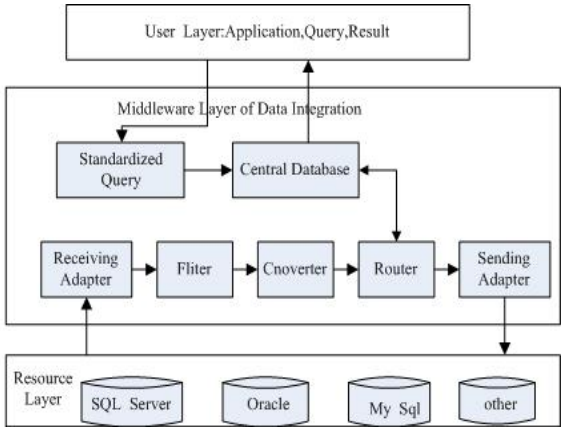


Figure 2. Data integration architecture based on SOA

Main flow of the system is as follows: user layer submits the standardized SQL query statements, directly queries in the central database, gets the suitable query results and returns them to user; In order to ensure the data consistency between the central database and the underlying databases, the middleware layer of data integration adopts "receiving adapter → filter → converter → router → sending adapter" to complete the data exchanging process. Firstly, receiving adapter receives the data that waits to be updated from the application databases in resource layer and then submits it to filter. Secondly, according to configurable business rules and information standards filter filters out these data inconsistent with rules and standards and then submits the matching data to converter. Thirdly, converter encrypts and decrypts these data, generates the corresponding data packet and submits it to router. Fourthly, according to the updated plans router updates the corresponding data of the central database. Finally, according to the synchronous scheme synchronous component submits the data packet to router, synchronizing to the corresponding data of the corresponding application databases.

C. The Introduction of Main Components

1) Standardized query: This module may parse the submitted SQL query statement by user and verify its correctness of SQL syntax.

2) Adapter: Adapter is actually a interface by which information with different formats can be converted.

Adapter is available to support the following protocols or services: HTTP(Hypertext Transfer Protocol) and HTTPS (secure hypertext transfer protocol), JDBC (JAVA Database Connectivity), TCP (Transmission Control Protocol), UDP (User Datagram Protocol), SOAP (Simple Object Access Protocol), WSDL (WEB Services Description Language) and so on.

3) Central database [5]: Through the secure and reliable public data exchanging, all applications can be integrated on the basis of sharing public data and the data integrity, the data accuracy and consistency can be effectively ensured in the process of information integration.

4) Fliter: According to configurable business rules and information standards, filter analyzes and deals with the data that waits to be exchanged, filters out these data inconsistent with rules and standards.

5) Converter: It converts the data to a suitable data format that can be received by receiver based on information standards.

6) Router: Based on configurable routing policy, the data can be securely exchanged and reliably transmitted between the application systems.

D. Public Data Exchange System

The data changing tracking components that are deployed in each application server are used to track the changing of data according to the data tracking schema, and then generate a relative data changing packet and submit it to the central database. Simultaneously, the data changing synchronous component is deployed in the central database server, which generates the sequence packets of updating data that are synchronized to each specific application database responding to the data Synchronization schema.

The logical structure of common data exchanging system is shown in Fig. 3.

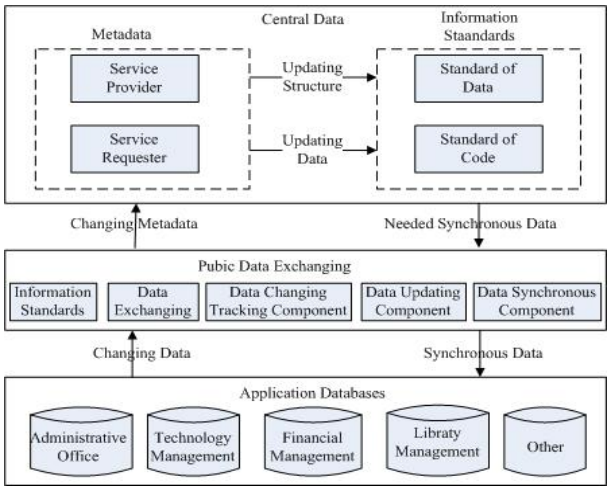


Figure 3. The logical structure of public data exchanging system

1) Public data exchanging system function

a) Information standards: A series of standards should be set, such as the data standards, code standards

and so on in order to provide a basis for the public information exchange. Metadata and the central database should be well maintained in order to provide a safe and reliable hub for the public data exchanging .

b) Data exchanging management: The automatic function of data exchanging and the supplementary function of data exchanging can be provided. Data changing tracking component, data updating component and data synchronous component should be effectively deployed and managed. If necessary, the abnormal data can be restored according to the log of data exchanging.

2) Public data exchanging process

Firstly, data changing tracking component tracks the changing of data from service provider and then generates the data waiting to be updated. Secondly, data updating component according to configurable data rules filters the data and converts it to the standard format data and then updates the corresponding data of the central database based on the updated plans. Finally, data synchronous component according to synchronous scheme synchronizes to the corresponding data of the corresponding application database.

The process of public exchanging is shown in Fig. 4.

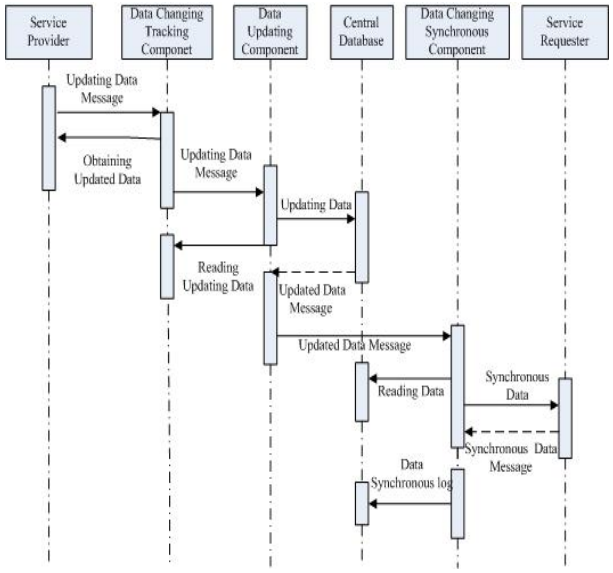


Figure 4. Flow chart for data exchanging

The introduction of public data exchanging process is as follows:

a) Service provider: Service provider is an application system that provides a business data (to the data item). Any of the data (to the data item) has a unique data provider.

b) Service requester: Service requester is also an application system that uses a business data (to the data item). Any of the data (to the data item) can have multiple data providers.

c) Data changing tracking component: It is used to intelligently track the changing data of the application databases, generate the data changing packet and then submit it to the central database.

d) Data updating component: This component can analyze, filter and convert the data that waits to be exchanged according to the regular data conversion mode.

e) Data synchironous component: This component can synchronize the data that waits to be exchanged to the corresponding application database.

Ⅴ. CONCLUSION AND FUTURE WORK

At present, campus information integration is lack of a unified architecture and standard. As the next-generation architecture, SOA is one of the best schema that is used to solve heterogeneous system integration. This paper proposes a new data integration middleware layer based on SOA and describes main components in details. It adopts schema based on XML to establish global model and local model and implements the deep integration of relational database. But some specific details need to be further improved, future work will mainly be improved in the following areas:

1) The type of data source needs to be increased. This paper discusses only integration query of relational database, and now, especially in Network, a lot of information is based on XML or other text format that will be the focus of future data integration research.

2) In practice, most of databases are distributed more scattered, network situation is more complicated, one of databases may be no longer respond, or result set is lost during transmission via network. Considering the complexity of network situation, query between heterogeneous databases should be tested and optimized in a distributed environment.

3) The accuracy and completeness of result set merging should be tested. The accuracy of a single data query is easy to implement, but the complete verification for result set merging of many databases has not yet to find a suitable way.

In short, data integration based on SOA is a very complex problem referring to a wide range of knowledge domains, many issues need to be overcome and be improved in the future of study and research.

REFERRENCS

[1] Thomas Erl," Services Oriented Architecture Concept, Technology and Design," China Machine Press,2007.

[2] ENDREIM, ANG J, ARSANJAN IA. Patterns : Service Oriented Architecture and Web Services [EB /OL]. (2004-04) . http: / /www. redbooks . ibm. com / redbooks/pdfs/ sg246303. pdf .

[3] Keith Ballinger,"Architecture and Implementation of .Net Web Services,"Beijing: China Electric Power Press,pp.5-83,2004.

[4] ZhenYu-Gang, Liu Luying, Kang Jian-chu,"Architecture and Implementation of an XML-based Heterogeneous Database Integration System," Computer Engineering, vol.32,no.2,pp.85- 87,2006.

[5] Jin Zhin-qiang,Teng Gui-fa,Sun Chen-xia,Zhu Ya-tao,"Design and Realization of a Heterogeneous Data Acess Interface Based On XML and a Dynamic Data Integration Model,"Journal of Agricultural University of Hebei,vol.32,no.2,pp.131-135,2009.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15, August 2010, pp. 124-128

Evaluation on E-government Websites Based on Rough Set and Genetic Neural Network Algorithm

Dang Luo, Yanan Shi
North China University of Water Resources and Electric Power, Zhengzhou, China
iamld99@163.com,shiyanan2003@163.com

***Abstract*—This paper researches on e-government website evaluation. After establishing the evaluation index system, this paper reduces the evaluation index system by rough set. Then, this paper introduces genetic algorithm which are optimized to BP neural network weights and thresholds, and establishes e-government website evaluation model based on genetic neural network algorithm. It is exemplified that the evaluation result is reasonable, and the evaluation model provides a new way of thinking for evaluation on e-government websites.**

***Index Terms*—index system, rough set, genetic neural network algorithm, e-government website evaluation**

I. INTRODUCTION

With the network popularization, the development of the e-government has made rapid progress. As an important component of the e-government construction, the e-government website is also a window which provides the society management and service by means of information technology. The construction and operation of an e-government website, is directly related to the government image, and also affects the level of management and service. Thus, it is a very important issue to emphasize websites building and improve the level of design, operation and management. How to strengthen the evaluation and establish a scientific index system to solve the problem in the development of e-government websites has become a problem which can not be ignored [1].

E-government website evaluation is complicated system engineering, with many subjective and objective factors affecting the evaluation, so it is very meaningful to adopt what kind of evaluation methods to make evaluation results objectively reflect the actual level of websites, so as to provide scientific basis for administrative decisions. This paper based on Rough set theory and genetic neural network evaluates the e-government websites comprehensively. Summarizing the existing evaluation index system, the paper establishes an e-government website evaluation index system, and makes use of rough set theory to simplify the established index system, then establishes e-government website evaluation model based on genetic neural network. Finally, this paper exemplifies the scientificity and validity of the model.

II. ESTABLISHMENT OF E-GOVERNMENT WEBSITE EVALUATION INDEX SYSTEM

A. Basic Theory of Rough Set

Definition1[2] Knowledge representation system $S=(U,R,V,f)$; where U is a non-empty finite set of objects ,also known as the universe of discourse; $R=C\bigcup D$ are a set of attributes, the subset C is the condition attribute set and the subset D is the result attribute set; $V=\underset{r\in R}{\vee}V_r$ is the set of attribute values , V_r is the range of values of the attribute $r\in R$,which is the range of the attribute r ; $f:U\times R\to V$ is a information function which assigns the at tribute value of every object x in U.

Definition 2[2] R is a equivalent relation, $r\in R$, if $ind(R)=ind(R-\{r\})$,then r is unnecessary in R; Otherwise, r is necessary in R.

Definition 3[2] If every $r\in R$ is necessary in R, then R is independent; Otherwise, R is dependent.

Let $S\subseteq P$, if S is independent and $ind(S)=ind(P)$, then S is one of reductions of P .All of reductions are $red(P)$, and the set which is consisted of all of necessary relations in P is the core of $P(core(P)=\bigcap red(P))$.

B. Establishment of Index System

The purpose of e-government website evaluation is to provide a reliable basis for e-government construction decisions. In order to evaluate the level of the e-government websites effectively, first we should establish a scientific, comprehensive evaluation index system. Currently, e-government construction has not yet formed a clear standard system, so there is no unified evaluation index system of e-government websites. Evaluation criteria in the international community mainly consists of e-government evaluation index system of Accenture Inc, e-government strategy evaluation system of Gartner Inc, e-government evaluation index system of UNDPEPA and ASPA and so on. Chinese scholars have also put forward to their own evaluation index system, for example, ref. [3] proposes e-government website evaluation method based on Web log analysis, establishing e-government website evaluation index system from the quality of construction, websites function services and the benefits and costs of websites; ref. [4] establishes the science and

technology system website evaluation index system from government affair, online service, online database of science and technology, public participation, website operation and maintenance ; ref. [5] proposes evaluation index system of government portal websites from four stages; ref. [6] establishes evaluation index system of the government portal websites from the website contents, website design, web technology, online work, public participation and economic services; ref. [7] establishes evaluation index system of the government portal websites from the site infrastructure, information disclosure, online services, public participation and interaction and website design.

On the basis of following the principle of operability, comprehensiveness and systematicness in the process of establishment, we have synthesized the study of the e-government evaluation index system home and abroad, and evaluated it from three aspects of website contents, website features and website construction referred to ref. [8-9]. Specific index system is as follows: first class evaluation index: website contents B_1, website function B_2, website construction B_3; second class evaluation index: comprehensiveness C_{11}, practicality C_{12}, accuracy C_{13}, timeliness C_{14}, authoritativeness C_{15}, personality C_{16}, novelty C_{17} and richness C_{18}; online capability C_{21}, information retrieval capability C_{22}, interactivity C_{23}, advocacy capacity C_{24} and category management capability C_{25}; security C_{31}, compatibility C_{32}, scalability C_{33}, stability C_{34}, operating speed C_{35}, profession C_{36}, beauty C_{37}, page hierarchy C_{38}, overall structure C_{39}.

According to e-government website evaluation index system, 10 experts in e-government website evaluation from different universities in Henan Province evaluates 18 government portal websites, and give every index value and the final evaluation results. This paper finally reduces the dates by rough set analysis software Rosetta.

Rosetta is a logic data analysis toolkit based on rough set theory. It was developed by computer and information science departments of Norwegian Science University and institute of mathematics of Warsaw University cooperatively. It is a good rough set theory software and experiment platform, and it provides a variety of date pretreatment function and common rules for reduction and algorithm in rough set, which supports the whole process from the data pre-processing rules to predict and analyze. Specific steps of attribute reduction are as follows:

a)Data pretreatment: Every index value is continuous, so the data must be discretized before attribute reduction. First run Rosetta and input data table, then discretize the data. 9 discretization methods are given by Rosetta; and this paper selects equal frequency discretization method [10];

b)Attribute reduction: As for attribute reduction for the pretreatment data, 8 attribute reduction methods are given; and this paper selects genetic algorithm.

According to attribute reduction method above, the e-government website evaluation index system is shown in Figure 1.

III. E-GOVERNMENT WEBSITE EVALUATION MODEL BASED ON GENETIC NEURAL NETWORK ALGORITHM

A. *The Principle of Genetic Neural Network Algorithm*

Neural network has good self-learning, self-adaptive, parallel processing and nonlinear computing capacity. Therefore, it has made a wide range of application in intelligent control, nonlinear optimization, signal processing and so on[11]. As BP neural network is a search algorithm along the gradient descent, once the quantity of the train sample is large, the input-output relationships become more complex and the network's convergence rate will become slow, easily falling into local minima. However, the genetic algorithm is a random search optimization method based on the principle of natural selection and genetics, which has a strong global optimization capability. Therefore, combining BP neural network with genetic algorithm, using genetic algorithm to optimize neural network weights and thresholds, we can make the network have a faster convergence speed. In this way, it not only plays a mapping and generalization ability of neural network, but also avoids the local minimum problem [12]. Genetic neural network algorithm steps are as follows:

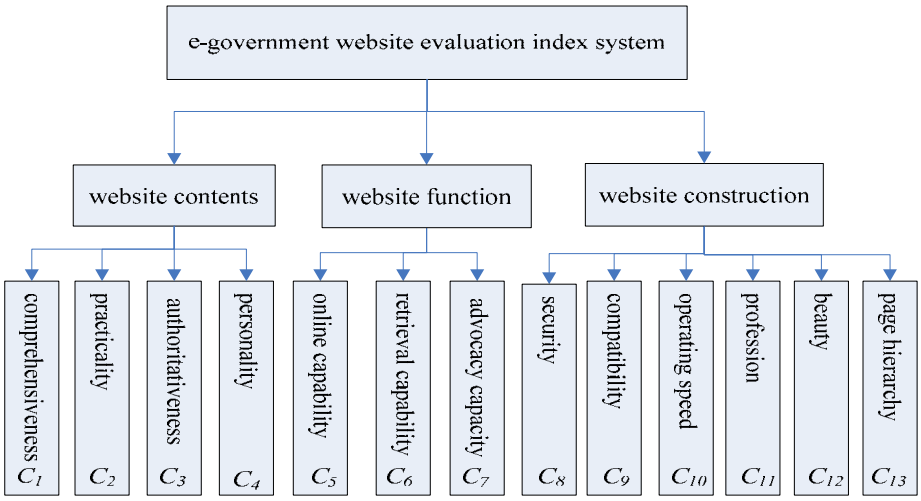


Figure 1 E-government website evaluation index system

a) Initialize the algorithm parameters: Set the population size, crossover probability, mutation probability, the number of network layers, the number of neurons in each layer and so on;

b) Set the fitness function: The fitness function is the derivative of network error which is the sum-of-squares of the error between the output value and expectation;

c) Encoding and population initialization: Using real number coding, link the weights and thresholds to form chromosomes, and then encodings respectively correspond to weights from the input layer to the hidden layer, weights from the hidden layer to the output layer, the thresholds of the hidden layer and the output layer;

d) Genetic operation: Use roulette selection method to choose the individuals, use consistent crossover to take cross operation on temporary population, and use site mutation operator to take mutation operation on temporary population;

e) Decode: Decode the selected optimal individual, and then obtain the initial weights and thresholds of BP neural network;

f) Training network: Train the network according to BP network neural parameters of the first step;

g) Training finished: Training does not finish until the global error is less than the pre-set target or the number of modifying is more than a pre-set value. Otherwise, turn on the 6th step;

h) Prediction or evaluation: Make use of the trained network to predict or *evaluate*.

B. Constuction of E-government Website Evaluation Model based on Genetic Neural Network Algorithm

1) Design of BP network structure

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin in this template measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

In 1989 Robea Hecht Nie1son has proven that a continuous function in any closed interval can be approximated by a hidden layer of BP neural network, so three layers BP neural network can make an arbitrary mapping from n-dimension to m-dimension[13]. Therefore, this paper adopts three-layer network structure:

For the input layer, taking the number of secondary indicators as the number of nodes of the input layer by the index system, there are 13 nodes.

For the hidden layer, in general, we can select the number according to the following empirical formula:

$$S = \sqrt{m+n} + L \qquad (1 \le L \le 10) \tag{1}$$

Where S is the number of hidden layer nodes, m is the number of input layer nodes and n is the number of output layer nodes. In this paper the number of hidden layer nodes is set to 6.

For the output layer, this paper would like make a proper evaluation for e-government website by the results of the output layer, thence in the paper the output layer node is set to 4. (1,0,0,0) indicates that the assessment result is excellent(“e”), (0,1,0,0) indicates that the assessment result is good(“g”), (0,0,1,0) indicates that the assessment result is medium(“m”), (0,0,0,1) indicates that the assessment result is poor(“p”).

2) The step of e-government website evaluation model based on genetic neural network algorithm

This evaluation model is achieved by means of goat toolbox and neural network toolbox of Matlab7.0. Specific steps are as follows:

a) Collect the corresponding input and output sample according to e-government website evaluation index system, and then standardize the sample;

b) Set the genetic algorithm parameters to take genetic operation, and then get the initial weights and thresholds of BP neural network;

c) Input the input and output sample to train, and training does not finish until the global error is less than the pre-set target or the number of modifying is more than a pre-set value;

d) According to the established network, entering

TABLE I THE BASIC DATA OF 18 GOVERNMENT PORTAL WEBSITES OF HENAN PROVINCE

No.	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	C_{11}	C_{12}	C_{13}	results
1	0.90	0.95	0.79	0.90	0.85	0.95	0.47	0.78	0.87	0.75	0.89	0.87	0.98	e
2	0.93	0.89	0.70	0.80	0.93	0.87	0.82	0.71	0.90	0.89	0.95	0.98	0.89	e
3	0.70	0.70	0.70	0.60	0.75	0.90	0.65	0.77	0.73	0.68	0.72	0.79	0.80	g
4	0.69	0.56	0.55	0.54	0.60	0.58	0.98	0.56	0.80	0.53	0.47	0.60	0.50	m
5	0.60	0.30	0.70	0.10	0.50	0.56	0.56	0.23	0.45	0.56	0.45	0.40	0.87	p
6	0.87	0.50	0.20	0.43	0.50	0.61	0.33	0.43	0.20	0.71	0.65	0.76	0.10	p
7	0.62	0.61	0.56	0.52	0.57	0.59	0.58	0.90	0.41	0.50	0.52	0.54	0.67	m
8	0.78	0.80	0.78	0.75	0.64	0.65	0.61	0.78	0.70	0.77	0.73	0.67	0.67	g
9	0.88	0.90	0.56	0.90	0.87	0.93	0.92	0.81	0.30	0.89	0.78	0.98	0.81	e
10	0.90	0.34	0.90	0.98	0.80	0.89	0.81	0.86	0.93	0.93	0.78	0.90	0.97	e
11	0.50	0.53	0.65	0.61	0.54	0.30	0.58	0.67	0.54	0.94	0.50	0.43	0.56	m
12	0.77	0.76	0.77	0.72	0.76	0.75	0.73	0.70	0.68	0.76	0.71	0.65	0.80	g
13	0.78	0.90	0.81	0.99	0.85	0.78	0.87	0.81	0.78	0.80	0.92	0.90	0.71	e
14	0.70	0.75	0.75	0.72	0.76	0.80	0.70	0.75	0.67	0.72	0.86	0.55	0.71	g
15	0.80	0.71	0.77	0.70	0.71	0.68	0.75	0.68	0.78	0.78	0.90	0.71	0.69	g
16	0.91	0.80	0.82	0.92	0.87	0.66	0.89	0.85	0.82	0.86	0.89	0.82	0.80	e
17	0.72	0.56	0.72	0.77	0.78	0.76	0.56	0.90	0.65	0.72	0.70	0.79	0.70	g
18	0.60	0.54	0.51	0.40	0.55	0.50	0.52	0.65	0.60	0.76	0.43	0.65	0.58	m

TABLE II INITIALIZED WEIGHTS AND THRESHOLDS OF BP NEURAL NETWORK

weights from input layer to output layer						thresholds	weights from hidden layer to output layer				thresholds
0.917	0.719	0.859	0.665	0.081	0.793	0.817	0.043	0.070	0.324	0.339	0.311
0.095	0.443	0.449	0.823	0.016	0.927	0.396	0.632	0.577	0.235	0.935	0.451
0.415	0.936	0.194	0.102	0.254	0.713	0.299	0.295	0.445	0.380	0.602	0.159
0.496	0.365	0.523	0.503	0.542	0.285	0.714	0.299	0.666	0.320	0.129	0.409
0.639	0.381	0.160	0.737	0.753	0.485	0.606	0.003	0.713	0.087	0.746	—
0.744	0.563	0.166	0.022	0.060	0.604	0.813	0.187	0.237	0.231	0.207	—
0.452	0.505	0.241	0.370	0.548	0.357	—	—	—	—	—	—
0.345	0.410	0.413	0.219	0.107	0.289	—	—	—	—	—	—
0.286	0.686	0.170	0.266	0.666	0.970	—	—	—	—	—	—
0.073	0.286	0.819	0.934	0.795	0.390	—	—	—	—	—	—
0.003	0.828	0.382	0.571	0.961	0.115	—	—	—	—	—	—
0.931	0.265	0.106	0.544	0.073	0.232	—	—	—	—	—	—
0.179	0.579	0.256	0.980	0.012	0.506	—	—	—	—	—	—

index value of the evaluation object to obtain the corresponding output, we can make a reasonable evaluation for e-government website by means of the output.

IV. EXAMPLE ANALYSIS

This paper selects 18 government portal websites of Henan Province as the research objects. We regard those index value that are reduced by rough set as the basic data of training and testing(TABLE I).The neural network trains with the index value of numbered 1-15 government portal websites. We test the evaluation results of neural network with the index value of numbered 16-18 government portal websites.

a) The basic data sample: The input and output sample is shown in TABLE I.

b) Genetic operation: Setting the population size is 100,the number of genetic generation is 200,the range of initialized weights and thresholds is [0,1] ,Parameter norGeomSelect is set to 0.09, Parameter arithXover is set to[2,0], Parameter nonUnifMutation is set to [2 200 3]; the genetic algorithm optimize to obtain the initial weights and thresholds of BP neural network which is shown in TABLE II.

c) The training of BP neural network: The transfer function of the hidden layer is set to logsig, the transfer function of the output layer is set to Purelin, and the training of BP neural network adopts LM (Levenberg-Marquardt) methods; the learning accuracy is 0.0001, the largest training times is 1000, the learning rate is 0.05, and the momentum factor is 0.60. After a few minutes, the training does not halt until training error reaches the demanded precision. The final network error performance curve is shown in Figure 2.

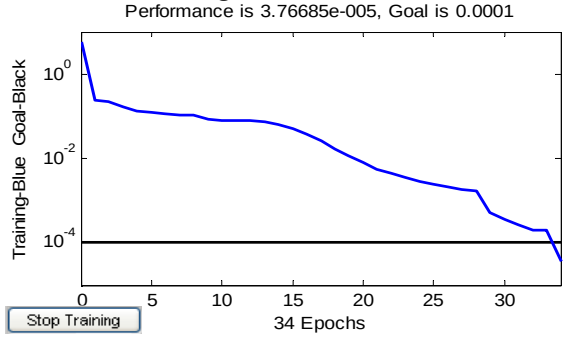


Figure 2 BP network error performance curve

d) Evaluation results of this model: When the network training is completed, we input the test sample to verify the adaptability of the network, and the test sample is numbered 16-18 sample. The comparison between network validation results and the actual evaluation results is shown in TABLE III.

TABLE III THE COMPARISON BETWEEN NETWORK VALIDATION RESULTS AND THE ACTUAL EVALUATION RESULTS

objects	16	17	18
output	0.8701	-0.1697	-0.0628
	0.1022	1.1334	-0.0085
	0.0353	-0.0356	0.9376
	-0.0063	-0.0018	0.1316
validation	e	g	m
actual result	e	g	m

From the validation and the results of comparison, we can see that the evaluation results through this network are basically same as the actual results. However, the training samples are insufficient relatively, so each output has a slight change, but this slight change does not affect the final evaluation results. The comparison indicates that the evaluation model has good simulation capabilities, and it is feasible to have a reasonable evaluation of e-government websites by genetic neural network algorithm.

V. CONCLUSION

On the basis of the existing literature, the paper establishes an E-government website evaluation index system, and simplifies the evaluation index system by rough set theory. On this basis, the paper introduces genetic algorithm which are optimized to BP neural network weights and thresholds, to solve the problem that the randomness of the initial weights and thresholds of the BP neural network causes the low network calculation accuracy and a fall into the local solution of the network easily. Then the paper establishes e-government website evaluation model based on genetic neural network algorithm. Through the case analysis, this model has fast speed of convergence and the result of the evaluation model is reasonable. So this model is a kind of effective method of e-government website evaluation, and also provides a new way of thinking for evaluation on e-government websites.

VI. CONCLUSION

On the basis of the existing literature, the paper establishes an E-government website evaluation index

system, and simplifies the evaluation index system by rough set theory. On this basis, the paper introduces genetic algorithm which are optimized to BP neural network weights and thresholds, to solve the problem that the randomness of the initial weights and thresholds of the BP neural network causes the low network calculation accuracy and a fall into the local solution of the network easily. Then the paper establishes e-government website evaluation model based on genetic neural network algorithm. Through the case analysis, this model has fast speed of convergence and the result of the evaluation model is reasonable. So this model is a kind of effective method of e-government website evaluation, and also provides a new way of thinking for evaluation on e-government websites.

ACKNOWLEDGMENT

This work was Supported by the Science and Technology Attack Projects of Henan Province (072102340009); Supported by the Natural Science Foundation of Department of Henan Province (2009A110011); Supported by the Philosophy and Social Sciences Plan Project of Henan Province (2007BJJ014); Supported by Soft Science Foundation of Science and Technology Department of Henan Province (082400440100).

REFERENCES

[1] Yue Ying. A Study on the Evaluation of Government Websites of China[D]. Soochow University master's thesis,2008:1-3.

[2] Zhang Wenxiu, Wu Weizhi, Liang Jiye,Li Deyu. Rough set theory and method[M]. Science press, 2001: 12-25.

[3] Wang Yi, Wang Suozhu. A Synthetic Approach to E-government Website Evaluation Based on Web Log[J]. Information Science, 2007,25(10): 1495-1498.

[4] Xu Xiaolin,Li Weidong. Evaluating S&T Government Web Site and Countermeasure of Improving S&T E-Government[J]. China Soft Science, 2005, (6):13-18;

[5] Wang Xuehua,Ge Dongxue. Research on the evaluation system of government portal website [J].Journal of Dalian University of Technology(Social Sciences), 2006, 27(1):59-61.

[6] Cheng Xuan. Research on the Evaluation Index system in Government Portal Websites [D].HuaZhong Normal University master's thesis, 2007:18-35.

[7] Yang Xingkai. Research on the Evaluation Indexes and Methods in Government Portal Website[J].Soft Science, 2007,21(4):34-37.

[8] Li Xinshi,Wu Xiaoyun. Based on websites Guangxi e-government evaluation [J].China Management Informationization, 2008, 11(16):102-104.

[9] Lu Fangmei,Wang Lubin.Study on Evaluation Indicator System of E-government Based on Information Architecture[J].Computer Engineering and Applications(supplement), 2006:93-112.

[10] Yu Kun,Liu Zhigui,Huang Zhengliang. Overview of the Discretization Methods in the Application of Rough Set Theory[J]. Journal of southwest university of science and technology, 2005, 20(4):32-36.

[11] Wang Lijun,Liu Xiaoyan. Fault diagnosis of large mechanism based on genetic-neural network[J]. Machinery Design & Manufacture, 2009, (6): 155-157.

[12] Gao Yanna,Zhu Daoling,Cheng Yuqi,Zhang Xiaofeng.Rural land price for compulsory acquisition appraisal model based on genetic neural network algorithm[J]. Systems Engineering-Theory & Practice, 2009, 29(4):103-110.

[13] Mao Zhiyong. Evaluation Model of Customer Satisfaction in B2C Electronic Commerce based on BP Neural Network [J]. Science Technology and Industry, 2008, 8(5):49-51.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 129-132

Research on Coordinate Transformation Method in Three-dimensional Reconstruction of Architecture

Hai Lin-peng[1], Chen Chong[2], Wang Yu-kun[3]
[1] Institute of Computer Science & Technology, Henan Polytechnic University, Jiaozuo, China
Email: hailp@hpu.edu.cn
[2] Institute of Computer Science & Technology, Henan Polytechnic University, Jiaozuo, China
Email: chennchong@163.com
[3] Institute of Computer Science & Technology, Henan Polytechnic University, Jiaozuo, China
Email: wyk@hpu.edu.cn

***Abstract*—Based on the analysis of the basic principle of three-dimensional reconstruction, this paper discusses the key problem in the process of three-dimensional reconstruction, and gives prominence to the most important process, which is making coordinate transformation and space projection to construct the model. The paper mainly adopts the method of the view separation by coordinate axes differing from the way of searching the silhouette of the views, achieving the coordinate conversion from view coordinate into space projection coordinate to get the three-dimensional coordinates which meets the needs of the conditions of three-dimensional reconstruction, and the 3D architectural model is rebuilt.**

***Index Terms*—three-dimensional reconstruction, orthographic views, view separation, coordinate transformation**

I. INTRODUCTION

For some decades now, with the development of the performance on computer graphic demonstration and the appearance of several kinds of strong functional developing software (such as OpenGL, DirectX and OpenGvs), the exhibition of realistic images on common microcomputer became reality.

The present research on the algorithm [1] about three-dimensional reconstruction centering on mechanical objects is very much, but the instantiation research oriented to architectural working drawing is rare, without the practical rebuilding tools for architectural objects. The objects of architecture reconstruction are more abstract and inconstant in form than mechanical objects, so some new thought and attempt is needed in the reconstruction, especially for more research penetrating to application layer. It is thus clear that architecture reconstruction has great theatrical value and practical sense, which is the extension of traditional three-dimensional reconstruction in the field of architecture.

II. THE BASIC PROCESS OF THREE-DIMENSIONAL RECONSTRUCTION

Three-dimensional reconstruction is the process from two-dimension to three-dimension in which we get the data information of three-dimensional objects through the semantic analysis of the samples based on the understanding of two-dimensional engineering drawings, according to the geometric and topological information in orthographic views or multi-view graphics. The process framework [2] of three-dimensional reconstruction is illustrated in Fig.1.

After importing the two-dimensional drawings into the computer, the file in DXF form is chose from AutoCAD. The drawing is in the global coordinate system. In DXF file, two-dimensional drawing elements are not recorded by sort according to the relations of feature entity, but they are recorded by drawing order, which don't reflect

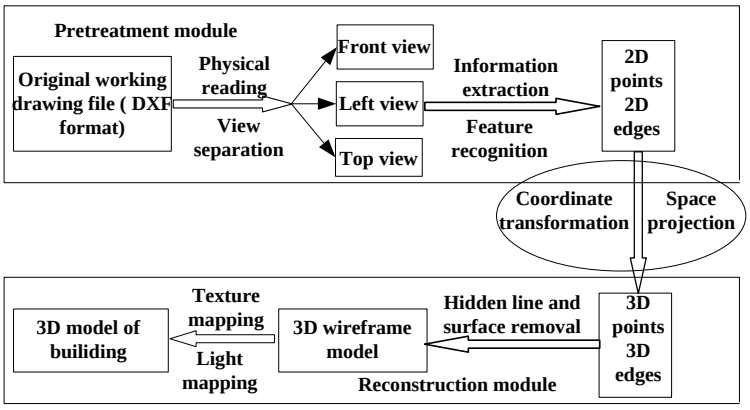


Figure 1. The process framework of 3D reconstruction

relationship. Therefore, view separation is made before transformation, and the origin point of view is determined. The view is separated into front view, left view and top view under orthographic projection, from which the useful information is extracted by feature recognition and information extraction. Not all the information in orthographic views is useful for us, and we only extract the useful information for our later rebuilding. The effective way of getting 3D points and 3D edges from 2D points and 2D edges is space projection and coordinate transformation.

The useful 2D points and 2D edges in need can be transformed into 3D points and 3D edges by space projection and coordinate transformation, which are stored by appropriate data structure. According to the principle of hiding the invisible line and surface, the hidden lines and surfaces are removed. With OpenGL and these data structure, the 3D wire frame model is built, which is rendered by texture mapping and light mapping to reconstruct 3D architectural model with more realistic effect. How to make coordinate transformation will be emphatically introduced as follows.

III. COORDINATE TRANSFORMATION IN THREE-DIMENSIONAL RECONSTRUCTION

In the process of three-dimensional reconstruction, how to get the three-dimensional coordinates and build the model is the key to three-dimensional reconstruction with known geometrical and topological information. The important way is coordinate transformation to get three-dimensional information from two-dimensional ones.

A. View separation

Before coordinate transformation from two-dimension to three-dimension the view separation must be made. It is that all the contents in the drawing are divided into three domains according to the ownership of orthographic views, in which each domain corresponds to one view. For each primitive in the drawing, judge separately which view these coordinates belong to.

Traditional view separation approach, not only angle discriminance [3], the seed-pot view separation algorithm [4] but also the maximum enclosing rectangle algorithm all need to search outer outlines of each view and compare them. In order to reduce the comparison with each two-dimensional entity, a new view separation method - the view separation [5] by coordinate axes is used to determine the separation line of orthographic views. As shown in Fig.2, each entity is projected onto x axis and y axis, and their union of projection onto x axis- line segments AB and CD and their union of projection onto y axis- line segments EF and GH are separately obtained. Draw a horizontal line lx at the point of y= (y_G + y_F)/2, and draw vertical line ly at the point of x= (x_B + x_C)/2, which divides the plane into P1, P2, P3, P4 four parts. According to the property of orthographic views, front view is in region P1, left view is in region P2, and top view is in region P3. Which region each entity locates in is just determined to separate the views.

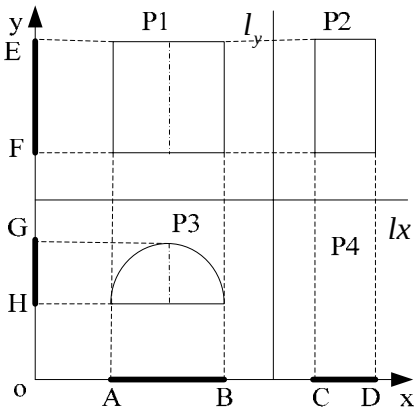


Figure 2. View separation by coordinate axes

B. Coordinate transformation

In order to convert the two-dimensional coordinates of every view in unified coordinate system into space coordinates, three different coordinate systems are set as follows: drawing coordinate system (x-y-z), view coordinate system (u-v-w), space projection coordinate system (X-Y-Z).

The relationship of the three coordinate systems [6] is shown in Fig.3. Drawing coordinate system is for drawing the initial input graphic, space projection coordinate system is used for building and describing the object of reconstruction, and view coordinate system is the intermediary between two-dimensional coordinates and space coordinates. Set the original coordinate as (x0 , y0) satisfying condition L=M.

1) Transformation from drawing coordinate to view coordinate

In Fig.3, the original point of each view drawing coordinate system coincide with space projection original (x0, y0), the coordinate axes u, v separately in parallel with x, y axis, in which the coordinate axis w always points at the paper outside according to right-handed rule. View coordinate (u, v, w) can be calculated according to drawing coordinates (x, y, z), and their relation can be expressed by matrix transformation as follows:

$$\begin{bmatrix} u & v & w & 1 \end{bmatrix} = \begin{bmatrix} x & y & z & 1 \end{bmatrix} A \qquad (1)$$

Transformation matrixes of every view are respectively as follows:

$$A_v = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ a & -b & 0 & 1 \end{bmatrix} \quad A_h = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ a & b & 0 & 1 \end{bmatrix}$$

$$A_w = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -a & -b & 0 & 1 \end{bmatrix}$$

Where a, b are coordinate parameters of projection original (x0, y0) in the same coordinate system, and subscript v, h, w denote respectively front view, top view and left view.

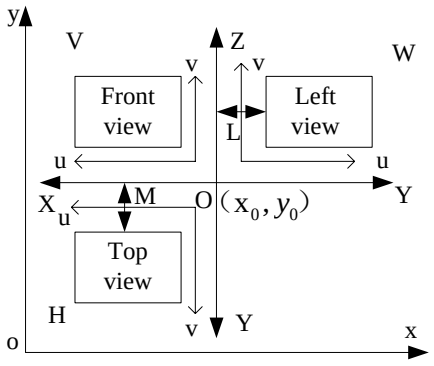


Figure 3. Corresponding relation among three coordinate systems

2) Transformation from view coordinate to space projection coordinate

The origin of view coordinate coincides with space projection coordinate, and then the conversion relation from view space to space projection coordinate can be expressed by transformation matrix B:

$$\begin{bmatrix} X & Y & Z & 1 \end{bmatrix} = \begin{bmatrix} u & v & w & 1 \end{bmatrix} B \qquad (2)$$

Transformation matrixes of every view are respectively as follows:

$$B_v = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad B_h = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$B_w = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

3) Transformation from drawing coordinate to space projection coordinate

After the above transformation, finally the matrix transformation from drawing coordinate to space projection coordinate can be expressed as follows:

$$\begin{bmatrix} X & Y & Z & 1 \end{bmatrix} = \begin{bmatrix} x & y & z & 1 \end{bmatrix} AB \qquad (3)$$

Let mapping operator T=A • B. Transformation matrixes of every view are respectively as follows:

$$T_v = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ a & 0 & -b & 1 \end{bmatrix} \qquad T_h = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ a & b & 0 & 1 \end{bmatrix}$$

$$T_w = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & -a & -b & 1 \end{bmatrix}$$

Using these three matrixes, the projection relation among orthographic views is built at the same time. Therefore, space projection coordinates from orthographic views are figured out by matrix transformation.

4) Extraction of basic points

According to the basic theory of descriptive geometry [7], the space points need to satisfy the principle of "length of the positive, height of the flush, width of the equal", and their projection onto two different directions can completely determine their position in the space, namely the points and edges should have entire equity relationship in different views. As for orthographic views, if V, H and W denote respectively the point set of front view, top view and left view. Then the point v in front view has x, z coordinate represented by x(v), z(v). In the same way, the points in top view and the points in left view are respectively expressed by x(h), y(h) and y(w), z(w).

The coordinate of the point in different view should satisfy the relation as follows:

x(v)= x(h),y(h)= y(w), z(v)= z(w) (v∈V, h∈H, w∈W)

The space point set F from orthographic views can be expressed as follows:

F={(v, h, w)(V, H, W/ x(v)= x(h)， y(h)= y(w)， z(v)= z(w))}

The corresponding relationship of these points is just the basic starting point of three-dimensional reconstruction.

The points which satisfy the requirements above are just the basic points needed in three-dimensional reconstruction, which is the way of getting three-dimensional coordinates from two-dimensional coordinates. The point coordinates are saved in 3D point table, while the edge coordinates are saved in 3D edge table.

IV. RECONSTRUCTION OF THE MODEL

Through coordinate transformation of three-dimensional reconstruction, the three-dimensional point coordinates and edge coordinates are got, from which how to rebuild the simple 3D model of the architecture by OpenGL programming is realized. The program framework [8] of three-dimensional reconstruction is illustrated in Fig.4.

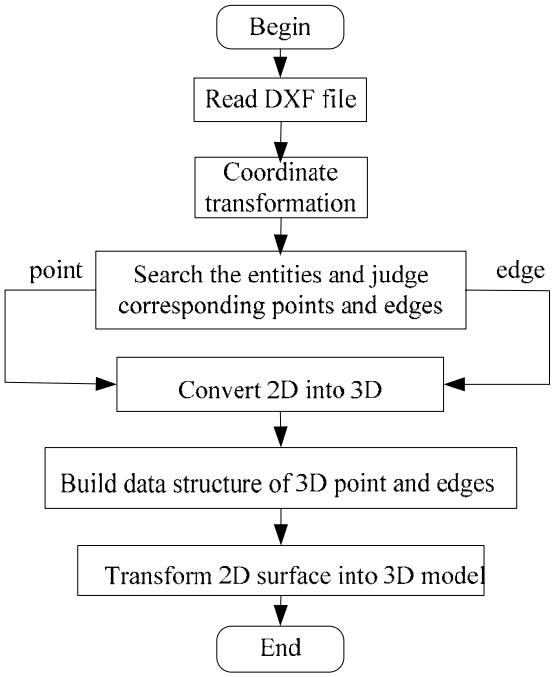
With the data structure of 3D points and edges, the useful coordinates are extracted and the position relation of each entity is calculated. Then the 3D model by VC++ programming is rebuilt. According to orthographic views of the architecture (Fig.5 is font view, Fig.6 is left view, Fig.7 is top view) the simple 3D model rebuilt is illustrated in Fig.8. For the building with multilayer, its complete structure can be finished by splicing according to instance technology.

After modeling, scene optimization technology, such as hidden processing, illumination model and texture mapping is used to achieve the drawing of realistic image.

V. CONCLUSION

In three-dimensional reconstruction, the two important modules pretreatment and reconstruction both center on coordinate transformation from 2D to 3D. Only when

finding the way from low dimension to high dimension, the 3D model can be rebuilt successfully. Make the DXF



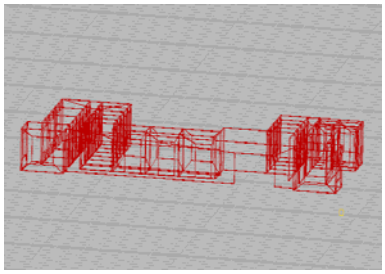


Figure 5. Front view

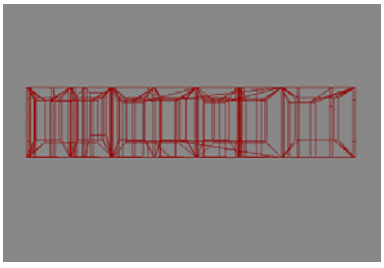


Figure 6. Left view

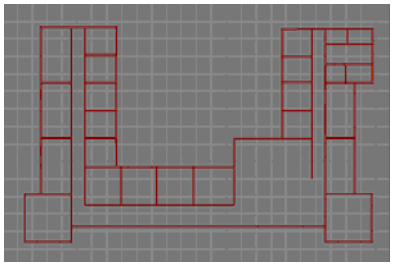


Figure 7. Top view

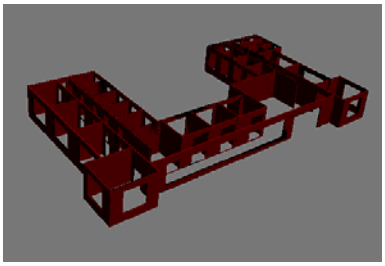


Figure 8. 3D model

file as the input, and organize the modeling data to complete the reconstruction combining with VC++ and OpenGL. At last, the virtual realistic image is rebuilt.

The implementation of three-dimensional reconstruction is useful for the reconstruction of according to common two-dimensional orthographic views, not only in architecture but in other fields. It all depends us for further study.

ACKNOWLEDGMENT

The authors would like to thank Institute of Computer Science and Technology in Henan Polytechnic University for their sponsoring to the subject and all the numbers helpful for my paper.

REFERENCE

[1] Li Xiao, Zhu Pengfei, “About 3D Reconstruction Algorithm Layout,” Computer Knowledge and Technology, Vol 5, No. 16, June 2009, PP. 4299-4300.

[2] Laifeng Shi, Beiji Zou, “Research on Pre-processing and Information Extracting for 3D Reconstruction from Engineering Drawings,” Application Research of Computers, vol.24, pp.161-165, Apr. 2007.

[3] Hiroshi Sakurai, David C.Gossard. “Solid Model Input Through Orthographic Views,” Computer Graphics, Vol 17, No.3, May 1983, pp. 243-252.

[4] Gao Wei, Wu Zhongqi, Tong Hongwei, “An Automatic Method of Recongnizing the Outline of Engineering Drawing,” Computer Applications and Software, Vol 11, No.1 pp. 243-252, Apr 1996.

[5] Liu Shixia, Hu Shimin, Wang Guoping, Sun Jiaguang, “Rec0nstructing of 3D Objects from Orthographic Views,” Chinese Journals Computers,vol.23, No.2. pp.141-146, Feb. 2000.

[6] Zhang Ai-jun, Zhu Chang-qian, Wang ji, “Coordinates Transformation of Engineering Drawing for 3D Reconstruction from Orthographic Views,” Journal of Southwest Jiaotong University, Vol 36, No.1, pp.57-61.Feb 2001.

[7] Min Jiang, Qiangde Li, “3D Reconstruction from Geometry Elements Based on AutoCAD2000,” Computer Applications and Software. vol.23, No 6, pp.87-88, Jun. 2006.

[8] Chunli Wang, Lipo Wang, “Development and aApplication of 3D Reconstruction,” Journal of Shenyang University, vol.15, pp. 25-26, Jun. 2003.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 133-136

Research of supervision system of coal mine safety based on VC

Li Changqing, Zhao Min
School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo City, China
Email: zhangwenjia6921@163.com, 181073417@qq.com

***Abstract*— This paper introduces the composition and functional structure of mine safety monitoring system, in the VC++6.0 environment, using SQL Sever database and multi-threading technology, unify the technology of the computer and communication, a VC-based mine safety monitoring system has been proposed and describes the system architecture design and implementation of key technologies. The application results show that the system is high efficiency、good real-time、reliable operation,the coal mine warning of a major disater prevention capabilities have been greatly increased,ensuring mine production safety is of great significance.**

***Index Terms* —monitoring system; VC ; SQL Sever; multi-threading technology**

I. INTRODUCTION

In recent years,great achievements have been made in China's coal industry,however,the number of coal mine safety accidents is increasing and there are many other issues,for example,the monitoring equipment is not complete and management of coal falls behind,coal mine production safety situation is still not optimistic.Therefore,one of the necessary means to resolve the current issue of coal mine production safety is the use of computer technology and communication technology on coal mine production safety monitoring and the establishment of a comprehensive mine safety monitoring system.

There are many funtions of mine safety monitoring system,such as:collection of analog、switching volume and total volume; transmission, storage, processing, display, print, sound and light alarm, and control.The system can be used to monitor the concentration of methane, carbon monoxide concentration, carbon dioxide concentration, oxygen concentration, wind speed, negative pressure, temperature, smoke, feed status, throttle state, windshield state, hairdryer state, local fan-off, the main fan-off, etc.The system can also be used to achieve sound and light methane overrun alarm, power and methane latch control, etc.From using mine safety monitoring system,accurate information can been provided for the monitor and it can have an effective guidance to the production.

II. OVERALL DESIGN

This section will introduce composition of mine safety monitoring system (-A).Next, composition of mine safety monitoring system software will be analyzed (-B).

A. .Composition of mine safety monitoring system

Mine Safety Monitoring System (Figure 1) contains the following components: monitoring host、UPS power supply, transmission interface, transmission cables, monitor sub-stations, printers, audio, various sensors and switches and so on..Sensors of different kinds are used to monitor the environmental parameters on the underground (monitoring the coal mine gas and face a variety of toxic and hazardous operating conditions) and also used to monitor the process of the manufacture (monitoring the production processes of various parameters and the operation of critical equipment state parameters). Sub-stations are located on the place of more sensors and are responsible for collection of underground electrical signal coming from the various sensors.After treatment and conversion electrical signal will be transported to the ground central station in the form of frequency.This system can receive the ground center station control command executing alarm and power-off function.The diagram can be shown in Figure.1 below.

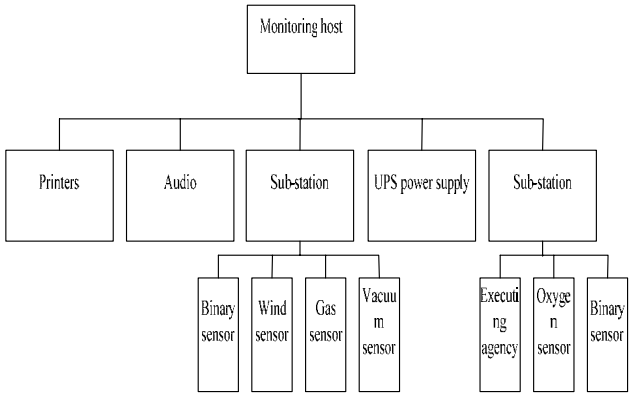


Fig.1. Supervision system of coal mine safety

B. .Composition of mine safety monitoring system software

The Mine Safety Monitoring System is developed by Visual c + +6.0,the application can be designed by users according to their needs.Various forms of data or information can be displayed in the form of graphic, curve, statements or documents.The user interface is intuitive、friendly、strong visualization and both has image and text.The application can be connected to SQL Sever database by using ADO and the powerful data processing capabilities.The data can be inqueryed、counted and analyzed through the operation of various functions of database. The computer storage capacity and high speed、high precision and wide range of features

and artificial intelligence are play so as to make information of monitoring system security、integrity、accuracy and timeliness of much better protection

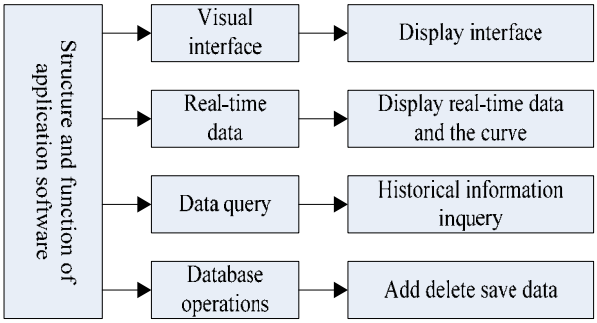


Fig.2. Overall function of monitoring system

According to the national safety standards for coal mine safety monitoring system and actual production of coal requirements,a monitoring software with the following features has been designed.Data acquisition、data control、data adjustment、data storage and query、printing、human-computer dialogue、analog alarms、network communication、real-time multi-task(real-time transmission, processing, storage and display information, and request real-time control, can cycle to run without interruption). The diagram can be shown in Figure.2 below.

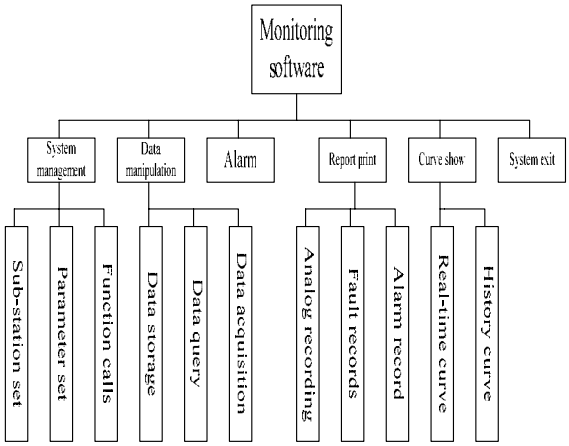


Fig.3. Overall function of monitoring system

III. DETAILED DESIGN

The goal of monitoring software is to estabilsh a software system which is consistent with "Coal Mine Safety Regulations" and "Standard of Mine Safety Monitoring System Software Design", adapting mine monitoring and control of environmental parameters.Device drivers are designed using visualize VC++ which can complete the data exchange between host and information transmission interface;use visualize VC++6.0 as development tools and build SQL Sever database,using VC technology for data access and data storage and processing.So,this section first presents the design of database(III-A) and data acquisition will be presented(III-B),Next,data processing will be discussed(III-C) as well as data display(III-D).

A. Database design and connection

SQL Sever 2000 database is used in this system and there are many advantages of using SQL Sever 2000 database, such as large memory capacity、high speed、high precision 、wide range and artificial intelligence.So the systematic、integrity、accuracy and timeliness of mine safety monitoring system can be better protected.The following table is part of the data table design:

Table name	Meaning	Function
KJ_XX	Table of mine information	Mainly used to set mine numben,name and mine leader
FZ_XX	Table of sub-station	Mainly used to set sub-station number,name and activate logo
TD_SENSOR	Table of channel sensor	Mainly used to set sub-station number, channel number, sensor type, installation location, power range, alarm , limit thresholds, alarm threshold solution, power threshold
PZ_ SENSOR	Table of sensor configuration	Mainly used to set sensor type number, name, type, name, unit, upper and lower range, frequency 1,2,3,4, coefficient of 1,2

Tab.1 Part design of database table

ADO (ActiveX Data Object) is designed for the latest and most powerful data access paradigm by Microsoft.And it is a user-friendly interface to the application layer. The program uses ADO to access the database, and there are many merits, such as easy to use, fast and spending less memory and disk remains small.

ADO library consists of three basic interfaces: _ConnectionPtr interfaces, _CommandPtr interface and _RecordsetPtr interfaces. ADO data source connection is created by connecting smart pointers.A pointer _ConnectionPtr m_pConnection to object pointer need to be created. Here is some code to connect to the database:

```
    BOOL CMyApp::InitInstance()
{
        AfxEnableControlContainer();
    AfxOleInit();
        m_pConnection.CreateInstance(__uuidof(Conne
                        ction));
}
 try
        {
                m_pConnection->Open("driver={SQL
Server};Server=127.0.0.1;DATABASE=CoalMine;UID=
sa;PWD="","","",adModeUnknown);
        }
        catch(_com_error e)
        {
        AfxMessageBox("Database connection failed ");
                return FALSE;
}
```

B. Date acquisition design

Data can be sent to the host from the sub-station through the interface.Firstly the host sends control signals 8FH, communication interface card receives the signal and then transmit data up to the host; Secondly the host determine the status of sub-station through the byte is 0 or not.If the byte is 0, the sub-station end the data transfer.Data collection asks for a timer from the system, WM_TIMER message is sent to the system at intervals of some time, users receive the message and start the process of collecting data once.

Macroeconomic effects of multi-task can run on the introduction of multiple threads in data collection.Using MFC in Visual C++ programming, the thread can be divided into worker threads and user interface thread, a data collection can be created in the view class's initialization function.Making the program start to read the collection of data from the main station.

C. Data processing design

The data collected from the sub-stations by the host through monitoring interface card are binary.Firstly you need to install for the frequency value, and then according to the sensor range the corresponding value is converted to the specific.The frequency of the sensor using in the system is 200-1000HZ and the s[i][j] (j value 1-4) is used to store the first I sub-station 4 analog frequency value and using the r[i][j] to store the parameters of specific value.The formula is as follows:

Gas (low density) linear counterparts 0-4.00% r [i] [j] = (s [i] [j] -200) * 0.005

Gas (high concentration) linear counterparts 0-40.0% r [i] [j] = (s [i] [j] -200) * 0.05

Linear speed corresponding to 0-15 m / s r [i] [j] = (s [i] [j] -200) * 0.019

Temperature 0-40 degrees r [i] [j] = (s [i] [j] -200) * 0.05

Negative linear mapping 0-5000Pa r [i] [j] = (s [i] [j] -200) * 6.25

Carbon monoxide linear counterpart 0-500ppm r [i] [j] = (s [i] [j] -200) * 0.625

D. Data display design

Data display can be divided into parameter display and sub-station display.Parameter display shows the information of gas、wind speed、negative pressure、CO、temperture、carbon monoxide,etc and it also include sensor installation location、name、status、measurements、sub-station installation sites and sub-station status.Sub-station display can show the analog、switch、value of control of current active sub-station and a test page is designed for use by debugging,showing the frequency of each analog.

Microsoft FlexGrid Control and SSTabCtrl controls of Visual c++ are used to display the data of the software in real-time monitoring. Microsoft FlexGrid Control controls display the data as the shape of the grid and SSTabCtrl controls display the data as paging.By timer trigger,the cycle intervals using the latest data collected and refresh list.In the data table,when the analog is over normal value(need alarm or power failure),the corresponding values change into red and it is blue when normal working.

IV. CONCLUSION AND FUTURE WORK

This paper mainly described some difficulities which could arise while the design of supervision system of coal mine safety.It first metioned composition of mine safety monitoring system and composition of mine safety monitoring system software.Next it gave the detailed design.This software is a real-time multi-task image-user software,it has many advantages,such as practical、simple、friendly interface、strong stability、fault tolerance、better maintainability、scalability and versatility.

Most of the mentioned methods have been implemented and tested on coal mine and some questions need to be improved.For example data dynamic display is not very good and fault diagnosis needs to be strengthened.What's more,directed the rescue after the accident need to be improved. Design a more intelligent mine safety monitoring system software will help to improve mine safety standards.

REFERENCES

[1] "Coal Mine Safety Regulations"[S]. National Coal Mine Safety Supervision Bureau, 2005.

[2] Sun Jiping. Coal Mine Safety Monitoring System [M]. Beijing: China Coal Industry Institute of Science and Technology of Labor Protection Safety Monitoring Professional Committee, 2005.

[3] Zhang Dinghua etc. KJ2005 colliery control system software design and implementation [J]. Computer Applications 2006.

[4] Wei Liu, Hong-mei Wang, Qing Xiao, Jian Yang. Internet of Things Concept [J]. Telecommunication Technology, 2010,vol.430, pp. 5-7.

[5] Bao-yun Wang. Internet of Things Technology Research [J]. Journal Of Electronic Measurement And Instrument, 2009,12(23).

[6] Zhi-feng Liu,Hong-hai Zhang,Jian-hua Wang. The EPCglobal network construction based on RFID technology[J].Computer Applications,2005,vol.25.

[7] Zhi-yu Ren, Pei-ran Ren. Internet of Things and EPC / RFID technology [J]. Forest Engineering, 2006,22 (1).

[8] Daniel W . Engels . A Comparison of the Electronic Product Code Identification Scheme& the Intemet Protocol Address Identification Scheme.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 137-139

The Strategies of Matrix Allocation and Efficient Analysis on Parallel Algorithm of Matrix Multiplication in multiple processors system

Jun Liu[1], Li Chen[2]
[1] Network Center, Henan University of Finance and Economics, ZhengZhou, China
Email: lj@hnufe.edu.cn
[2] Computer Center, Henan University of Finance and Economics, ZhengZhou, China
Email: cl@hnufe.edu.cn

***Abstract*—Parallel matrix multiplication has been investigated extensively in the last two decades. There are different approaches for matrix-matrix multiplication. We analyses the factors of affecting the efficiency for matrix multiplication parallel algorithm in the multiple processors system at first. Then a mathematical model which is about how to allocate matrix data to the processors is presented. The strategies of allocating matrix were discussed in the end.**

***Index Terms*—Matrix Multiplication, Allocation of Matrix, Parallel Algorithm, Multiple Processors system**

I. INTRODUCTION

Matrix multiplication is used much in many problems' solving process, and it has good parallelism in itself and is suitable for parallel processing. There are different approaches for matrix-matrix multiplication. For example, Cannon's algorithm[1], Fox's algorithm[2], Bentsen's algorithm, DNS algorithm[3] and other algorithms[4][5]. According to the degree of parallelism, matrix multiplication algorithm can be classified into the following several kinds: inner-product method, middle-product method, outer-product method and using n^3-parallelism, the degrees of parallelism of these algorithms respectively are: 1, n, n^2, n^3 (n is the dimensions of matrix).

In order to obtain the most desirable performance, algorithm must be fit for the system architecture of computer. The inner-product method is a typical series algorithm and is fit for being executed in single processor SLSD structure while outer-product method and n^3-parallelism are fit for being executed in processor array SLMD which matches the dimensions of matrix well.

According to the categorical method, Flynn, multiprocessor system is always attributed to MIMD system implementing the total parallelism of tasks and processes. The number of units in multiprocessor system should not be too great and the middle-product method, of which the degree of parallelism is lower, is suitable when it does matrix multiplication.

Assume that A and B are both n dimensioned matrixes and C is the product of their multiplication.

The middle-product method been described in C programming language likes the follows:

```
For(j=1; j<=n; j++)
    For (k=1; k<=n; k++)
        C[][j] = C[][j] + A[][k] * B[k][j]
```

Where [][j]=C[][j]+A[][k]*B[k][j] is a vector expression.

C[][j] and A[][k]are both vectors which consist of the *j*th and the *k*th column of A respectively. "+" is parallel summation of n elements in one time while "*" represents the multiplications of scalar B[][j] and vector A[][k].

The middle-product method's degree of Parallelism is n, the exertion of Parallelism is the best when the number of processors is equal or greater than n, however it will see problem when execute big matrix algorithm in multiple processors system with a much lower number of processors, that is, how to allocate matrix A by row (or allocate matrix B by column) reasonably in order to balance data in each process unit and make them execute parallel as much as possible.

In practical application, the condition is seen much that the dimensions of matrix is great while the processor number is lower, so under this circumstance it possesses both theoretical and practical significances to analyses the parallelism of matrix multiplication theoretically.

II. THE BLOCK DIAGRAM OF PARALLELISM ALGORITHM OF MATRIX IN MULTIPLE PROCESSORS SYSTEM

The topology structures of multiple processors system which the algorithm we will describe here is fit for are star LAN or Bus PON and their varieties types. A node I/O is responsible for reading in the data of matrix A and B, allocating them to each slave node, gathering and exporting data of matrix C, and it also takes part in the matrix multiplication meanwhile. Each slave node accepts both matrix A and B that have been split from the I/O node and returns the results to it after finishing computing. Figure 2.1 and 2.2 show the system topology structures.

This Project Supported by the Natural Science Research of Henan Provincial Education Department of (No. 2010A520004).
This Project supported by the Key Technologies R&D Program of Henan Province of China (No.0624260017, 072102210029)

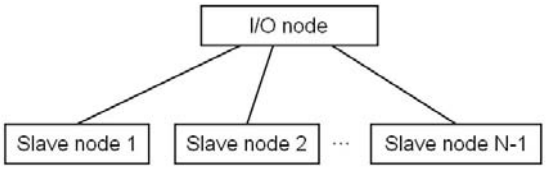


Figure 2.1

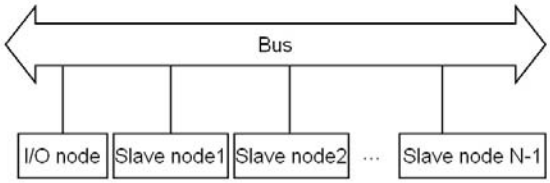


Figure 2.2

Assume the matrixes are A_{mxy} and B_{yxn} that do multiplication, C_{mxw} is the product. Figure 2.3 shows the block diagram of program, where 2.3, the left part represents the executing process of I/O node while the right part represents the executing processes of each slave node.

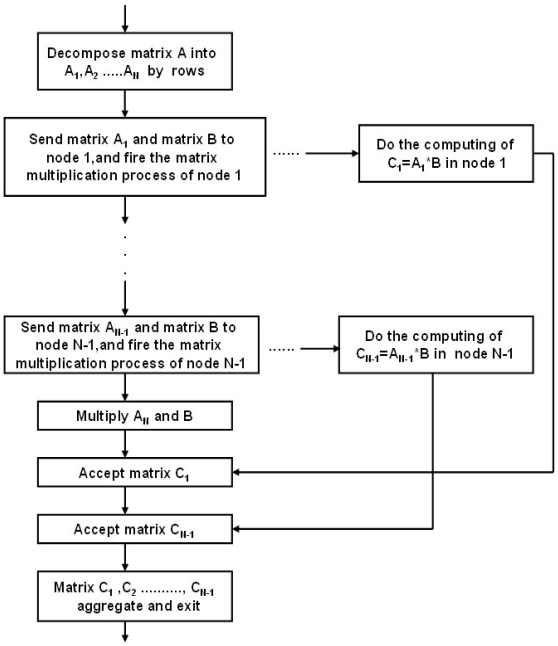


Figure 2.3

From figure 2.3 we can observe that when it computes matrix multiplication parallel, even though the programs been executed by I/O node and each slave node are the same, they execute the different parts of data of matrixes, so it is the parallelism algorithm.

III. THE STRATEGIES OF ALLOCATING MATRIX

In order to reduce the spare time and improve the executing efficiency of each node as much as possible, the demands in terms of time is like this: slave node 1 completes computing and the I/O node accepts its computing results, matrix C_1, now, slave node 2 happens to just complete its computing and after its results being accepted by I/O node, slave node 3 happens to complete its computing , and so forth, till I/O node finishes accepting the result matrix C_{N-1} from slave node N-1. If it can meet this time demand, then all the nodes in system will be in busy state the most probably and so it can exert the parallelism of hardware the most significantly.

According to the constraints that in terms of time described above, we combine program block diagram 2.3, the following equations set can be listed:

$$\begin{cases} T_1 = T_{S_2} + T_{S_3} + \cdots + T_{S_{N-1}} + T_{cal_N} \\ T_2 = T_{S_3} + T_{S_4} + \cdots + T_{S_{N-1}} + T_{cal_N} + T_{r_1} \\ \quad\vdots \\ T_{N-1} = T_{cal_N} + T_{r_1} + \cdots + T_{r_{N-2}} \\ S_A = S_{A_1} + S_{A_2} + \cdots + S_{A_N} \end{cases} \quad (1)$$

Where, S_{A_1} , S_{A_2} , ..., S_{A_N} are the numbers of elements of matrixes of A_1 , A_2 ,... ,A_N respectively.

T_K (1<=k<=N-1) is the total time being needed for slave node k's computing. T_{S_K} is the communication expense of the *k*th slave node's accepting matrixes A_K and B from node I/O and activating the process of matrix multiplication.

T_{cal_N} is the time being needed for node I/O itself doing matrix multiplication. We can obtain the following formula from figure 2.3:

$$T_K = z*S_{A_k}*t_{k_{in}} + z*S_B + w*S_{A_k}/y/f_k + z*w*S_{A_k}*t_{k_{out}}/y \quad (2)$$

In this formula, z is the element length of each matrix (the number of bytes they occupied).

$T_{K_{in}}$, $T_{K_{out}}$ are respectively the time being needed for the matrix multiplication process of the *k*th slave node's reading one element into the local memory of itself from common area (shared-memory) and that of writing it back to common area (shared-memory) from its local memory.

S_B is the number of elements of matrix B.

f_k is the matrix multiplication speed of the *k*th node (completing once addition and once multiplication each second).

$$t_{S_k} = z*(S_{A_k} + S_k)*1/u_k + t_{act_k} \quad (3)$$

In this formula, u_k is the communication ratio (byte/s) between I/O node and the *k*th slave node.

t_{act_k} is the time being needed for node I/O activating the matrix multiplication process of the *k*th node.

So: $z*(S_{A_k} + S_k)*1/u_k$ is the time being needed for node I/O sending matrix A_k and B to the communication area (shared-memory) of the *k*th slave node.

$$T_{act_N} = w * S_{A_N} / y / f_N \quad (4)$$

$$T_{r_k} = z * (S_{A_K} + S_B * 1 / u_k \quad (5)$$

We substitute (2)-(5) into the equation set (1), after arranging we obtain equation set (6):

$$\begin{cases} (z*t_{1_{in}} + w/y/f_1 + z*w*t_{1_{out}}/y)*S_{A_1} + z*S_B*t_{1_{in}} \\ = \sum_{k=2}^{N-1}(z*(S_{A_k}+S_B)*1/u_k + t_{act_k}) + S_{A_N}*w/y/f_N \\ (z*t_{2_{in}} + w/y/f_2 + z*w*t_{2_{out}}/y)*S_{A_2} + z*S_B*t_{2_{in}} \\ = \sum_{k=3}^{N-1}(z*(S_{A_k}+S_B)*1/u_k + t_{act_k}) + S_{A_N}*w/y/f_N + z*(S_{A_1}+S_B)*1/U_1 \\ \quad . \\ \quad . \\ \quad . \\ (z*t_{N_{in}} + w/y/f_{N-1} + z*w*t_{N-1_{out}}/y)*S_{A_{N-1}} + z*S_B*t_{N-1_{in}} \\ = S_{A_N}*w/y/f_N + \sum_{k=1}^{N-2} z*(S_{A_k}+S_B)*1/U_k \\ S_A = S_{A_1} + S_{A_2} + \cdots + S_{A_N} \end{cases} \quad (6)$$

In the equations set (II) , the parameters are known or can be measured in addition to X 1,x2,x3 are for variables. So solving this equation set, we can get each node should be assigned the number of matrix elements.

IV. DISCUSSION ABOUT THE STRATEGIES OF ALLOCATING MATRIX

Now we discuss three points about the equations set:

(1) The mathematical model of strategies of allocating matrix that equation set (6) has described considered the elements of the communication overhead, the computing speed of each process unit and so on. However, in practical system, these elements can be overlooked so that the strategies of allocating can be simplified.

For example, let us consider the following condition: In the multiple processors system, the communication overhead and the proportion that the time being needed for each slave node reading in matrix date to memory from communication buffer and writing date back to the communication buffer from memory possesses in the whole multiplication computing is so small that they both can be overlooked.

Besides the above, the overhead of the I/O node's activating the matrix multiplication process of each slave node can be overlooked too. Considering only the floating point computing ability of each node is an ideal condition, under this circumstance, the equation set (6) can be simplified into equation set (7) :

$$\begin{cases} S_{A_1} = (f_1 / f_N) * S_{A_N} \\ S_{A_2} = (f_2 / f_N) * S_{A_N} \\ \quad . \\ \quad . \\ \quad . \\ S_{A_{N-1}} = (f_{N-1} / f_N) * S_{A_N} \end{cases} \quad (7)$$

Solve this equation set we obtain:

$$S_{A_1} = (f_1 / \sum_{i=1}^{N} f_i) * S_A, S_{A_2} = (f_2 / \sum_{i=1}^{N} f_i) * S_A, \cdots S_{A_N} = (f_N / \sum_{i=1}^{N} f_i) * S_A \quad (8)$$

That is, the numbers of matrix elements we distribute to nodes depend on their portions of computing ability.

(2) If we still overlook each kind of extra expense, and consider the computing ability of each node to be isomorphic, just like we have described above, then equation set (7) can be further simplified like following: $S_{A_1} = S_{A_2} = \cdots = S_{A_N} = S_A / N$, that is, the numbers of elements of matrix distributed to each node are the same.

(3)Because in practice, the communication overhead of every two nodes is always should not be neglected, or the computing ability of each node is different, or that if smaller matrixes do multiplication, it is possible that not every node takes part in the computing, or even the computing in single processor is quicker than allocating matrix to each node does. This problem which embodies in the solution of set of equation is: each node k of which the S_{A_k} is smaller than the number of elements of one row of matrix A is not necessary to be distributed matrix to.

From the discussion about this point, we can observe that multiprocessor system is smarter than parallel processors system, and the former is more versatile.

V. THE END

This paper analyses the many elements of allocating matrix that will affect matrix multiplication algorithm in multiple processors system, and list the mathematical model of the strategy of allocating matrix. However, there are so many parameters in this model that the solving process is very sophisticated, so in practical application, we can idealize some of them in order to make the solving process of equation set easier.

REFERENCES

[1] L.E. Cannon, A cellular computer to implement the Kalman Filter Algorithm, Ph.D. dissertation, Montana State University, 1969

[2] G.C. Fox, S.W.Otto, A. J. G.Hey, Matrix Algorithm on a hypercube I: Matrix multiplication, Parallel Computing, vol. 4, pp17-31. 1987

[3] A. Gupta, V. Kumar, Scalability of Parallel Algorithms for Matrix Multiplication, Proc. ICPP, 1993

[4] A. Grama, A. Gupta, G. Karypis, Introductin to parallel Computing, Addison Wesley, 2003

[5] Z.A.Qadi, M. Aqul, Performance Analysis of Parallel Matrix Multiplication Algrithms Used in the Image Processing, World Applied Sciences Journal 6(1):45-52, 2009

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 140-143

Numerical simulation of cold plasma jet by Lattice Boltzmann method

Yabing Dang [1], Yanzhou Sun[2]
[1] Henan Polytechnic University, School of Electrical Engineering and Automation , Jiaozuo, China
Email: dangyabing@163.com
[2] Henan Polytechnic University, School of Electrical Engineering and Automation , Jiaozuo, China
Email: sunyz@hpu.edu.cn

***Abstract*—A model for calculating the velocity and temperature field of plasma jet by Lattice Boltzmann (LB) method was established in this paper. The numerical simulation of plasma jet was derived by selecting two opportune equilibrium distribution functions, and the results obtained are compared with those of the experiment and in literature. It is found that the LB method is simpler and more efficient than the traditional finite difference method.**

***Index Terms*—atmospheric plasma jet, Lattice Boltzmann, numerical simulation, discharge**

I. INTRODUCTION

Atmospheric Pressure Discharge, particularly arc discharge plasma torch and corona discharge, has been widely used in the field of materials processing and waste treatment. In recent years, the study found: atmospheric pressure cold plasma jet has many advantages [1], such as small volume, low temperature, low cost electricity and the electron density is higher than $10^{13} cm^{-3}$ and so on, as well as low pressure glow discharge.

Atmospheric pressure cold plasmas have received increased attention recently because of several emerging applications such as in material processing, aerodynamics (drag reduction, shock wave mitigation) [2]-[3], bio-medicine, and radar communications (absorption and reflection of microwaves). At atmospheric pressure, some sophisticated diagnostic techniques developed for low-pressure plasmas are not applicable. For example, at atmospheric pressure, the electron density and temperature, two of the most important plasma parameters, cannot be measured by the Langmuir probe because the electron mean free path is shorter than the Debye distance. Because of high temperature (up to 10^4 K), high flow velocity (up to several hundred meters per second) and great gradients in plasma jets, it is not easy to exactly diagnose the characteristics of plasma jets. For this, numerical simulation is necessary for atmospheric pressure cold plasma jet. At present, the finite difference method or the finite element method is commonly applied to discrete macroscopic equations at home and abroad, through SIMPLE and its improved algorithm iterative solution, and has achieved many results[4]-[6]. There are many advantages in the LB approach, such as the direct physical nature and a simple idea. It also provides easily implemented, fully parallel algorithms and the capability of handling complicated boundaries. It can be forecast that the simulation by LBM is faster than that by traditional methods, for there is no need to solve the large equations and no need to eliminate the unsuitable pressure and no iterative procedure in the calculation process. For this, an attempt of studying the plasma by the Lattice Boltzmann method at the microscopic level was performed in this paper.

II. NUMERICAL SIMULATION

A. The LB model

In this paper, as shown in Fig. 1, the model of two-dimensional nine vectors is discretized into a square. And according to the theory of the lattice Boltzmann method, it consists of two steps: a streaming step and a collision step.

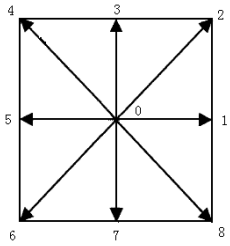


Figure1. D2Q9 lattice configuration.

In this model, each node can have three particles stationary particle, orthogonal direction of movement of particles and the diagonal direction of movement of particles, respectively. The discrete velocities for the D2Q9 model are defined as $e_i \ \ i = (0,1,2,3,4,5,6,7,8)$. And the nine vectors of the lattice links are documented as follows:

$$e_i = \begin{cases} (0,0) & i = 0 \\ (\cos\frac{i-1}{2}\pi, \sin\frac{i-1}{2}\pi) & i = 1,2,3,4 \\ \sqrt{2}(\cos(\frac{i-5}{2}\pi + \frac{\pi}{4}), \sin(\frac{i-5}{2}\pi + \frac{\pi}{4})) & i = 5,6,7,8 \end{cases} \quad (1)$$

The macroscopic density and velocity can be calculated from:

$$\rho = \sum f_i(x,t) \quad (2)$$

$$\rho u = \sum e_i f_i(x,t) \quad (3)$$

Where $f_i(x,t)$ is particle distribution function, and it is a real variable, represent the particle speed and direction of motion. According to the kinetic theory, the mass, momentum and energy of particle is transferred by migration and collision. The key of the lattice Boltzmann method is streaming step and collision step. At each time step, particles are moved from the present lattice site to a neighboring lattice site, and it under the rules of the collision tends to local equilibrium. The evolution equation is given by:

$$F_i(x+e_i,t+1)=F_i(x,t)+\Omega_i(F) \tag{4}$$

Where $\Omega_i(F)$ is collision operator, the right items of evolution equation is the particle in discrete time and space move to the neighboring lattice points.

The evolution is consists of two steps: a streaming step and a collision step. And the collision operator can be implemented by many methods, if we adapt the relaxation process as follows:

$$\frac{\partial f}{\partial t}+\nu\cdot\nabla f=(-1/\pi)(f-f^{eq}) \tag{5}$$

Where $f(x,\nu,t)$ is defined as the number of particles with space element $(x,x+dx)$ and speed element $(\nu,\nu+d\nu)$ at a time. And then the evolution equation is as follows:

$$F_i(x+e_i,t+1)=F_i(x,t)-\frac{F_i-F_i^{eq}}{\tau} \tag{6}$$

We can get the equilibrium distribution function of 9-bit lattices is expanded by LB and a multi-scale method [7] and it is follows:

$$f_i^{eq}(x,t)=\rho\omega_i(1+3(e_i\cdot u)+\frac{9}{2}(e_i\cdot u)^2-\frac{3}{2}u^2) \tag{7}$$

The parameter ω_i is a weighting factor specific for each velocity direction, and they are expressed as follows:

$$\begin{aligned}&\omega_i=4/9 \qquad i=0\\ &\omega_i=1/9 \quad i=1,2,3,4\\ &\omega_i=1/36 \qquad i=5,6,7,8\end{aligned} \tag{8}$$

In this paper, the temperature equilibrium distribution function of plasma jet is taken one order accuracy, and the evolution equation is expressed as follows:

$$T_i(x+e_i,t+1)=T(x,t)-\frac{T_i-T_i^{eq}}{\tau_T}-\frac{q\alpha_i}{\rho c_p} \tag{9}$$

Where τ_T and q are the temperature relaxation factor and the radiance per unit volume of plasma jet, α_i and c_p are the percent of the direction particles and specific heat, respectively, and ρ is density of plasma jet.

In the process of deducing the macro function of the plasma jet, the relaxation factors are obtained as follows [8]:

$$\tau_T=0.5+2\Gamma_e/(\alpha C^2\varepsilon),\ \tau=0.5+\mu_e(\rho c_s^2\varepsilon) \tag{10}$$

Where α is the proportion of non-static particles, Γ_e and μ_e are the effective diffusion coefficient and viscosity, respectively.

B. Computational domain and boundary conditions

In this paper, for the symmetry of the plasma jet, a half calculating region is plotted, and the central line is employed to symmetry side; the left side is fixed bound expressed in Eq. (11). The left-up side adopts a non-slip rebound side; the right side and the upper side are regarded as free bound [9].

$$\left.\begin{aligned}u_Z&=u_{Z,\max}[1-(r/R_0)^n]\\ T&=T_W+(T_{\max}-T_W)[1-(r/R_0)]^n\end{aligned}\right\} \tag{11}$$

Where R_0 is the radius of the nozzle, T_w is the wall temperature of the tube, and r is the radial position coordinates. And the velocity and temperature distribution of parameters n is 2 and 4, respectively.

Usually, in the collision dominated plasma, the excitation temperature can be seen as the electron temperature. Atmospheric pressure plasma can be seen as the collision dominated plasma, and for this the electronic excitation temperature of atmospheric pressure plasma jet can be seem as electron temperature. The electronic excitation temperature of atmospheric pressure plasma jet is up to thousands of degrees, while other particles are only a few degrees or at room temperature [10]. In this paper, we only consider the electron temperature, and the temperature of the ion or atom is ignored. In this paper, the grid division is 300×51, and the plasma is spraying in the air, the working gas is Ar and the diameter of the tube is 8mm.

III. CALCULATION RESULTS

Fig. 2 shows that when other conditions are the same, the flow characteristics of the plasma jet change with the jet-inlet velocity. In Fig. 2 T_0 = 14000K, v_0^a = 450m/s, v_0^b = 550m/s, v_0^c = 650m/s. It can be seen from Fig. 2 that when the inlet velocity of the plasma jet changes from 450m/s to 650m/s, the average extent of the length of plasma jet is about 3mm. The temperature distribution of the plasma jet hardly changes though the variation extent of the inlet velocity attains several hundred meters per second. The calculation results accord with the experimental results in the published documents [11]–[13]. Fig. 3 shows the simulation of velocity and temperature.

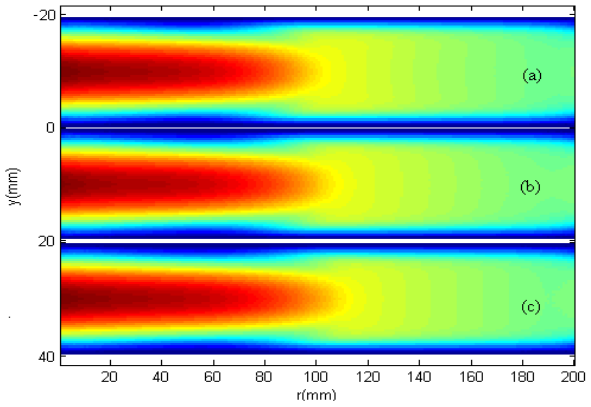


Figure2. Numerical simulation the flow characteristics of plasma jet.

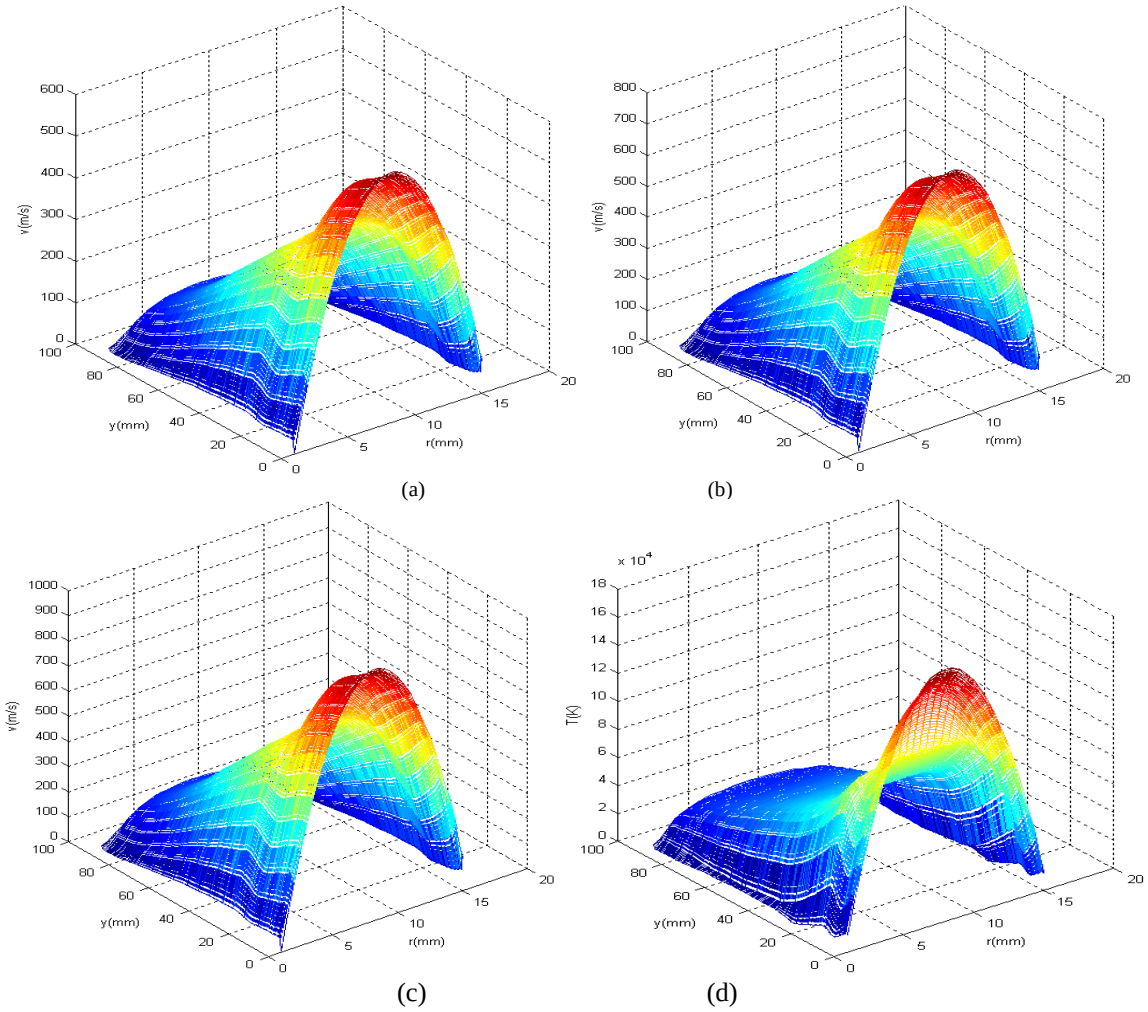


Figure3. a, b and c are the simulation of velocity, and the jet-inlet velocity of a b and c are 450, 550 and 650 m/s, respectively. d is the simulation of temperature, and the jet-inlet temperature is 14000K.

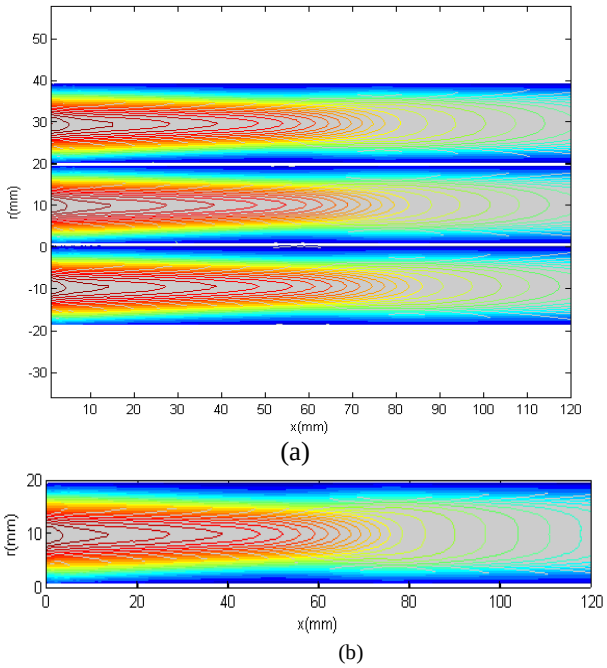


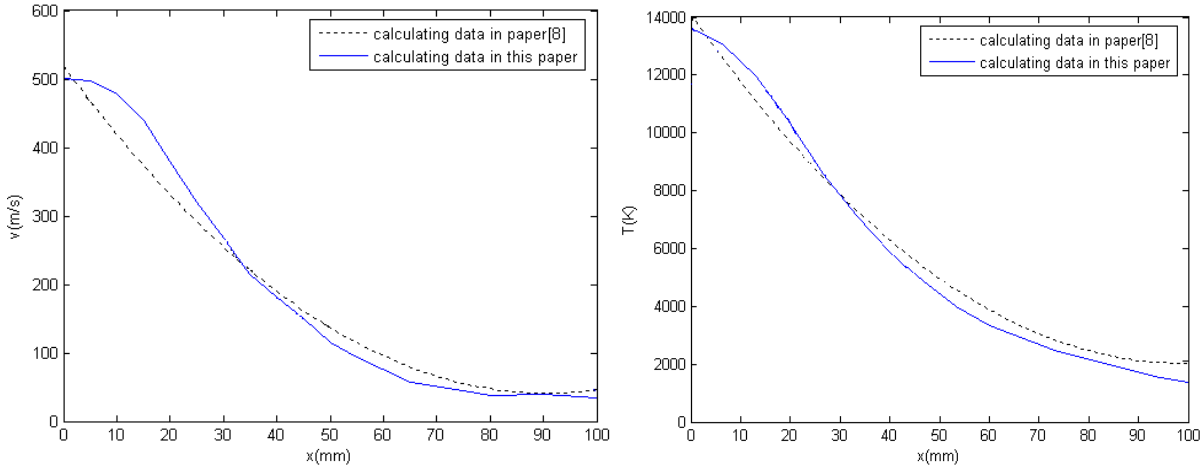
Figure4. (a) Axial velocity distribution with different exit velocities. (b) Temperature distribution. Numerical simulation of plasma jet.

Fig. 4 shows that the axial temperature gradient near the high temperature is higher than 200 K/mm and the velocity gradient attains 10 (m/s)/mm near a high-velocity zone, as in accordance with the experiments and simulations by the conventional method. It is can be seen from Fig. 4 that the speed and temperature contours are both middle-intensive, sides of the sparseness, and at the same position the velocity with higher inlet velocity is higher along the centerline.

The calculation results of the plasma jet along the jet centerline are compared with calculating data in paper [8], shown in Fig. 5. It can be seen from Fig. 5 that the simulation results by D2Q9 model agree with the calculating data and the distribution trend of the velocity and the temperature distribution along the jet centerline are almost as same as in paper [8], respectively. To sum up, the method was valid in this paper.

IV. CONCLUSIONS

In this paper, we can find that the simulation by the LB plasma jet model is faster than that by the conventional methods because there is no iterative calculation in the simulation process. And results show that the method was



(a) (b)

Figure5. a and b are velocity and temperature distribution along the jet centerline, respectively.

valid when the model modeling was made to calculate velocity and temperature field of the plasma jet. Due to the effect of turbulence dissipation, the variety of jet field distribution and length of turbulent plasma are not obvious with increase of the inlet velocity and temperature brought by the enhancement of arc current and gas flow rate.

REFERENCES

[1] Y. H. Wang and D. Z. Wang, "Numerical simulation of dielectric-barrier-controlled glow discharge at atmospheric pressure," Acta Phys. Sin. Vol. 52, pp.1694–1700, Jul. 2003.

[2] E. Stoffels, I. E. Kieft, R. E. J. Sladek, L. J. M. Bedem, E. P. Laan, and M. Steinbuch, "Plasma needle for in vivo medical treatment: Recent developments and perspectives," Plasma Sources Sci. Technol., vol. 15, pp. S169–S180, Nov. 2006.

[3] M. Laroussi and X. Lu, "Room-temperature atmospheric pressure plasma plume for biomedical applications," Appl. Phys. Lett., vol. 87, pp. 112 902/1–112 902/3, Sep. 2005.

[4] Q. Y. Shao, Y. He, W. K. Guo, P. Xu and Zang dehong, "Comparison of results obtained from temperature measurement and numerical simulation of DC plasma arc," Phys. vol. 48, pp. 1–14, Sep. 1999.

[5] X. Chen, "Heat-transferring and Flowing Of The High-temperature and Ionic Gas," Beijing: Science Press, 1993, pp. 50–74.

[6] D. Y. Xu and X. Chen, "Motion and heating of ceramic particles in a 3D laminar plasma jet," Journal of Engineering Thermophysics, vol. 25, pp. 676–678, 2004.

[7] W. P. Shi, Y. Q. Zu, "Evaluation of Fluid Acting Force on the Curve Boundary in the Lattice Boltzmann Method ," Journal of Jilin University: Science Edition, vol.43, pp. 132–136, 2005.

[8] G. L. Wang, J. Y. Zhu and H. O. Zhang, "Numerical simulation of plasma jet by Lattice Boltzmann method," Hua-zhong Univ of Sci & Tech (Nature Science Edition), vol. 31, pp.1–3, 2003.

[9] A. H. Dilawari, J. Szekely, J. Batdorf, R. Detering and C. B. Shaw, "The temperature profiles in an argon plasma issuing into an argon atmosphere: a comparison of measurements and prediction," Plasma Chemistry and Plasma Processing, vol. 10, pp. 321–329, 1990.

[10] A. Fridman, A. Chirokov and A. Gutsol, "Non-thermal atmospheric pressure discharge," Appl. Phys. vol. 38, pp. R1-R24, 2005.

[11] W. X. Pan, "Generation of long, laminar plasma jets at atmospheric pressure and effects of flow turbulence," Plasma Chem. Plasma Process. vol. 21, pp. 23–30, 2001.

[12] M. Xian, W. X. Pan and C. K. Wu, "Temperature and velocity measurement of plasma jet," Eng. Thermophysics. vol. 25, pp. 490–492, 2004.

[13] W. X. Pan, M. Xian and C. K. Wu, "Length change of DC laminar-flow argon plasma jet," Eng. Thermophys. vol. 26, pp. 677–679., 2005.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15, August 2010, pp. 144-147

Component Based Coordination Software Development Method

Qingxin Li, Shufen Liu, WeiFeng Xu
College of Computer Science and Technology,Jilin University,Changchun,China
e-mail: liqingxin@uibe.edu.cn

***Abstract*—Through research on component technology and coordinative model, this paper defines and elaborates the software development method of component based web coordination from the perspective of software reuse theory. It presents an in-depth analysis of source components, component management, component assembly technology, coordinative model, and the use of components and integrated technology. By proposing the component based coordination software architecture that introduces interactions container in the connector, this study sheds light on realizing the plug and play functionality of software component in a heterogeneously distributed environment,, and generating coordinative application software.**

***Index Terms*—Component based, Framework, Software reuse, Coordinative Model, Architecture**

I. Overview of Software Components

A lot of repetitive work has been identified in traditional software development. With the expanding scale of software, it is becoming increasingly challenging to control software development costs, improve the efficiency as well as quality of software development, and ensure the consistency. Since component has the reuse characteristic of software, it is possible to use existing components in application software assembly so as to overcome these difficulties. Coordinative work of the system aims to enable a geographically dispersed group to work together in completing a task through the use of computer network technology[1].

II.Research on Component Technology

During the development of complex enterprise-class software systems, developers often invest a lot of resources into research and practice, and deploy a variety of techniques to improve software quality and reduce development costs. Among these techniques, component based method is proved to be one of the most effective.

A Component definition

The concept of components has been applied to all walks of life, but there is by no means an uncontested definition for software industry. Component itself is software module, which can be delivered to the user with certain functions, and provide an interface that can use the services offered by itself[2]. Based on these characteristics, component can be defined as a functional module that is released independently, and can access its service through its interface. A component should be comprised of two parts at least, namely the functional part offered by itself and the interface elements. Accordingly, the description can be formalized as a binary group, i.e. Component = (CF, CI), where CF is the component function, and CI is the component interface[3].

Component interface is the access point used by the external access to describe the behavior mechanism of the component as the basis of software reuse. A well-designed interface can make it easier for component assembly and replacement[4].

In the process of component development, the component functionality realization and component interface are normally separated, which makes the component design and implementation oblivious to the users. And all services provided by component are registered in the interface, so that the component designers only need to pay attention to the realization of interface functions, while the users only need to be concerned about the interfaces they depend on.

B Component categories

Components can be classified as service component and business component based on the reuse granularity size and concern point[5].

In accordance with its level, service component could be divided into present component, logic component, computing component, etc. For example, computing component mainly completes the business logic calculation; present component mainly provides view, logic operation and other functions to interact with the user for show interface; and logic component mainly realizes more complex logic functionality.

Business component is the most coarse-grained reusable module in component based conceptual framework and consists of a series of different service components. It is a large-scale reusable distributed information system which involves all the necessary software work products for describing, implementing, and deploying specific business logic. Business component directly provides business logic functionality for the system, thus it is not only a component but also a direct expression in solution space to the business concept within problem space.

III.Component Based Software Development Method

Component based software development is a feasible way to realize software reuse. It can be deployed to

AP-PROC-CS-10CN007

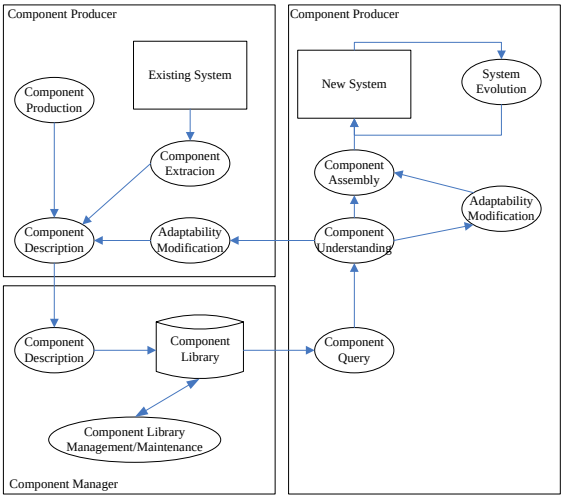


Figure1. Component Library and Component Based Software Development

improve the capability and extent of software reuse, sort out development cycle, reduce development cost and enhance development quality. Component based software development incorporates the following three processes: component obtaining, component management and component based assembly and integration.

A Component obtaining

In the development of component based systems, component obtaining usually adopts the following methods: one is to access the necessary components directly, that is we can directly or indirectly obtain the components that meet the requirements of the component library, for example, buying commercial off-the-shelf components (COTS) from the market. The other deal with developing new components to meet the requirements.The way to obtain components is determined by the system required functionality. Meanwhile, developers must take into account the costs which not only involve developing or purchasing components, but also upgrading and maintenance. However, regardless of the paradigm taken to obtain the components, it must undergo rigorous evaluation and testing, and be incorporated into the component library for unified management.

B Component management

In component based software development process, component management plays a connecting role between component production and component reuse. Component management is achieved by the component library system. Therefore, component library system which supports component classification, organization, storage and query plays a critical role in component based software development. Component library is a reusable set of components based on certain semantics and structures. In general, it is classified into two major parts: component library and component management system. Component library is deployed to store components, while component management system realizes adding, deleting, searching, browsing component and other functionality that manage components. Overall, component library provides component storage, and management offers effective services for developers. Figure 1 illustrates the support from component library for component based software development.

C Component assembly

Component assembly technology is critical in software component technology. It describes the assembly mechanism deployed to build components on the basis of component model, and the process of constructing application by connecting reuse component interfaces. Its essence is to establish static relations between components and coordinate component behaviors through relations to make the components form an organic whole.

Component assembly methods are mainly divided into black-box, white-box and gray-box assembly. Black-box assembly describes an implementation mechanism by which developers only need to assemble component interfaces rather than have an understanding of component interior. It makes component users focus on the system architecture to be built and the choice of components. However, if developers adopt black-box assembly, they would not know the component internal principles and the implementation methods. Consequently, they could not deploy those components that do not meet the requirements completely. Compared with black-box assembly, white-box assembly provides the user with greater flexibility and adaptability, since all the internal components are open, and developers can modify components in accordance with requirements during the development process. However, white-box assembly has also raised a few issues. One is that it places high demands on the users, who are not only required to understand the component interfaces, but also the component implementation clearly. Another issue is that modification may make new components inconsistent with the norms and thus faulty. Gray-box assembly stands in-between black-box and white-box assembly. It merely requires components to offer modifiable source codes of related interfaces, and the developers do not need to understand the internal implementation. Components provide their own expansion language or application programming interface, by which developers can modify the components. The advantage of gray-box assembly is that it meets the assembly requirements by adjusting the assembly mechanism rather than directly modifying the internal structure of components.

Currently, gray-box assembly is separated into four categories: framework-based assembly method, architecture-based assembly method, connector-based assembly method and glue code-based assembly method.

This study mainly focuses on the framework-based assembly method. Generally, it utilizes the component as a black box, and directly inserts components into the framework. However, due to the uncertainty of domain model, it is essential to integrate the components into the framework flexibly. Consequently, framework

configuration files-based socket assembly method is put forward. With this method, one can achieve component pluggable assembly according to requirements by modifying framework configuration files.

Architecture-based assembly method is to study and implement display of system architecture, which deploys reusable components on high level of abstraction to deal with overall organizational structure and control structure, functionality of calculation elements distribution, as well as high-level interaction between calculation elements and other design problems [6].

Connector-based assembly method shifts the focus of development from coding towards selecting and integrating components and connectors. This development method places same priority on components and connectors, and puts all existing and newly developed connectors in connector library for reuse. Connector library is the key of connector-based assembly method.

Glue code-based assembly method assembles appropriate components after the developers get them, and, if necessary, adds glue codes to connect various components reasonably. Recently, some scholars have

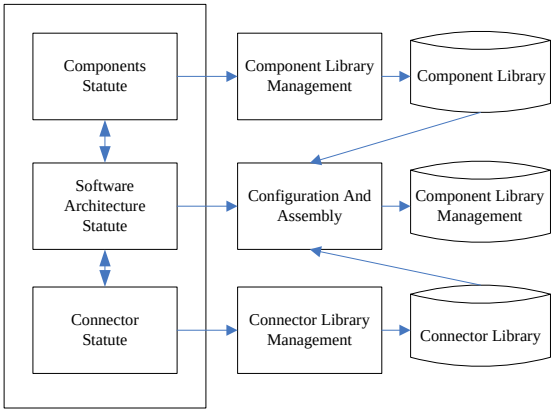


Figure2. Architecture-Based Component Assembly Framework

put forward the concept of glue layer, that is, the components are glued on a layer and interact with each other through this layer. As can be seen, glue code-based assembly method is applicable to the case of large coupling between components[7].

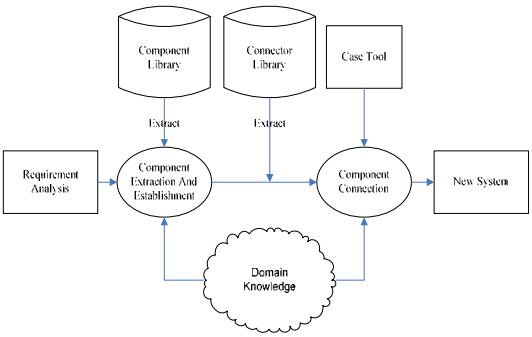


Figure3. Connector-Based Assembly Method

IV. Coordinative Model

The coordinative model provides high-level abstraction to interactions in software system. It consists of three parts, coordination entity, coordination media and coordination law. Coordination entity participates in the coordinative activities; coordination media provide interactive space for coordination entity; and coordination law is packaged in the coordination media to describe the specifications of the coordination media operation[8].

V. Realization

Figure 4 demonstrates the component based coordination software architecture. It introduces interactions container in the connector to provide operating environment for coordination media and the ontology of abstract service norm components, while coordination media is reusable to achieve the interactive access of components.

The remaining elements of the architecture are described as follows:

Component: calculation entity in the architecture, corresponding to the collection of interacting objects (synergistic entity) in the coordination model.

Interface: a collection of components methods, defining the transfer method agreement, with the basic functionality of initializing and executing service instances.

In the above-mentioned architecture, the components adopt triggered, loosely coupled, and asynchronous

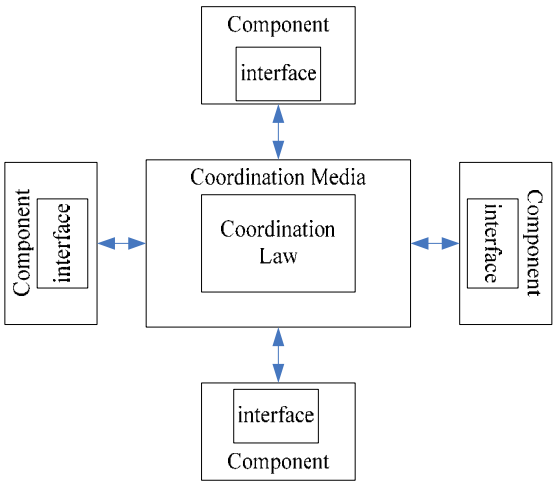


Figure 4 Component based Coordination Software Architecture

interaction with the event notification mode instead of specific binding between the visits. Components trigger exchange of information through the coordinative media event, and components could make asynchronous communication without the knowledge of each other's presence. Adding, deleting or modifying components in the process of operation will not affect the behavior of other components.

VI .Conclusion

The use of component technology in developing coordination application system will realize the plug and play functionality of software component in a heterogeneously distributed environment, thus bring existing applications and system software to rational and efficient reorganization and customization. Component based coordination software development will play a significant role in Internet technology in the future.

REFERENCES

[1] Ciancarini P. Coordination models and languages as software integrators. ACM Computing Surveys, 1996, 28(3): 300-315.

[2] Fuqing Yang, Bing Zhu, Hong Mei. Software reuse. Journal of Software, 2005, 6(9).

[3] Yu Liu, Shikun Zhang, Lifu Wang, Fuqing Yang. Component based software framework and role expansion research, Journal of Software, 2005, 14(8).

[4] Xingping Huang, Xiangming Long, Peng Xu, Fangchun Yang. COSFoTS: Component based telecommunications software framework. Journal of Communication, 2007, 28(1).

[5] Wei Zhang, Hong Mei. One feature-oriented domain model and its modeling process. Journal of Software, 2003, 14(8).

[6] Horstmann, C. S. Cornell, G. Core Java 2 Volume : Fundamentals Prentice Hall PTR. 2006.

[7] Eclipse Rich Client Platform: Designing, Coding, and Packaging JavaTM Applications, Addison-Wesley Professional, 2005: 65-225.

[8] Yuan Yuan & Guoyin Wang. Coordination model of component driven by service ontology. Computer Engineering and Applications，2009, 45 (8), 1-4

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 148-150

Automatic Verification of Acquisti Voting Protocol in Formal Model

Bo Meng[1], Wei Huang[2], and Dejun Wang[2]
[1] School of Computer, South-Center University for Nationalities, Wuhan, China
Email: mengscuec@gmail.org
[2] School of Computer, South-Center University for Nationalities, Wuhan, China
Email: {hwaoding2002@yahoo.com.cn, wdj_1001@yahoo.com.cn}

***Abstract*—In this paper Acquisti voting protocol is modeled in applied pi calculus. Soundness and coercion-resistance are verified with the automatic tool ProVerif. The result shows that Acquisti protocol has the soundness and coercion-resistance in some conditions. To our best knowledge, the first automatic analysis of Acquisti protocol for an unbounded number of honest and corrupted voters is provided.**

***Index Terms*—automatic proof, remote internet voting, applied pi calculus, ProVerif**

I. INTRODUCTION

With the development of internet and information technology, electronic government has got serious attention from government, enterprise and academic world. Owning to advantages of remote internet voting, it plays an important role in electronic government. In order to increase confidence of the voters in remote internet voting system, many researcher focus on design and verification of secure remote internet voting systems and protocols. Remote internet voting protocol is a key part of internet voting system. So how to develop and verify a practical secure internet voting protocol are challenging issues.

In the last twenty years many remote internet voting protocols [1~12], claimed on their security, have been proposed. In order to verify security properties of remote internet voting protocol there are two model can be used: one is formal model (or Dolev-Yao, symbolic model) in which cryptographic primitives are ideally abstracted as black boxes, the other is computational model(or cryptographic model) based on complexity and probability theory. Firstly each model formally defines security properties expected from security protocol, and then develop methods for strictly proving that given security protocols satisfy these requirements in adversarial environments. Computational model is complicated and is difficult to get the support of automatic tools. In contrast, formal model is considerably simpler than the computational model, proofs are therefore also simpler, and can sometimes benefit from automatic tools support, for example: Revere[13], Casper[14], SPIN[15], SVM[16], NRL[17], Brutus [18], Scyther [19], Isabelle[20], Athena[21], ProVerif[22].

ProVerif is an automatic cryptographic protocol verifier based on a representation of the protocol by Horn clauses and applied pi calculus. It can handle many different cryptographic primitives, including shared- and public-key encryption and signatures, hash functions, and Deffie-Hellman key agreements, specified both as rewrite rules and as equations. It can also deal with an unbounded number of sessions of the protocol and an unbounded message space. When ProVerif can not prove a property, it can reconstruct an attack that is, an execution trace of the protocol that falsifies the desired property .It can prove the following properties: secrecy, authentication and more generally correspondence properties, strong secrecy, equivalences between processes that differ only by terms. ProVerif has been tested on protocols of the literature with very encouraging results (http://www.proverif.ens.fr/proverif-users.html).

Owning to analysis manually of security properties of Acquisti protocol in the paper [9], this method depend on experts' knowledge and skill and is prone to make mistakes, we use automatic tool ProVerif to verify security properties of Acquisti protocol.

II. MODELING ACQUISTI PROTOCOL WITH APPLIED PI CALCULUS

Acquisti protocol [9] promises that it can protect voters' privacy and achieves soundness, universal verifiability, receipt-freeness, and coercion-resistance without ad hoc physical assumptions or procedural constraints. It mainly applies threshold Paillier cryptosystem, bulletin board that is a public broadcast channel with memory where a party may write information that any party may read, Mix net that guarantees privacy is a distributed protocol that takes as input a set of messages and returns an output consisting of the re-encrypted messages permuted according to a secret function, proof of knowledge that two ciphertexts are encryption of the same plaintext, designated verifier Proof of knowledge. Acquisti assumes that the private key is private and that an attacker cannot control every possible communication between the voter and an authority.

In Acquisti protocol there are five entities: registration authority, issue authority, bulletin board, voters, tallying

Corresponding author: Bo Meng, School of Computer, South-Center University for Nationalities, Wuhan, China, 430074
This work was supported by Natural Science Foundation of South-Center University for Nationalities (YZZ09008)

authority. Registration authority is responsible for authenticating the voters. Issue authority takes charge of issuing the related key and credentials. Voters register for voting, get their credentials and post a vote. Tallying authority is responsible for tallying ballots.

Acquisti protocol consists of preparation phase, voting phase and tallying phase. In preparation phase the related keys and ballot are generated. Issuer authority creates the voting credential shares and posts copies of the shares of credentials encrypted with Paillier cryptosystem to a bulletin board. The same credential shares encrypted with different Paillier public keys and attach a designated verifier proof of the equivalence between the encrypted share and the one the voter has received to its message are also provided to voters. Issuer authority also creates the ballots shares which are encrypted with the two different Paillier public keys. Both the sets of encrypted ballots shares are posted on the bulletin board together with zero-knowledge proofs that each pair of ciphertexts are encryptions of the same ballot share, and are then signed by issuer Authority. In voting phase the voter vote his favor ballot and post it to bulletin board. Each voter multiplies the shares she has received from issuer authority together with the encrypted shares of the ballot. Because of the homomorphic properties of Paillier cryptosystems, the resulting ciphertext includes the sum of those shares and the ballot's shares. The resulting ciphertext is sent to the bulletin board. In the last phase, tallying phase, the tallying authority tallies the ballot and publishes the result in bulletin board

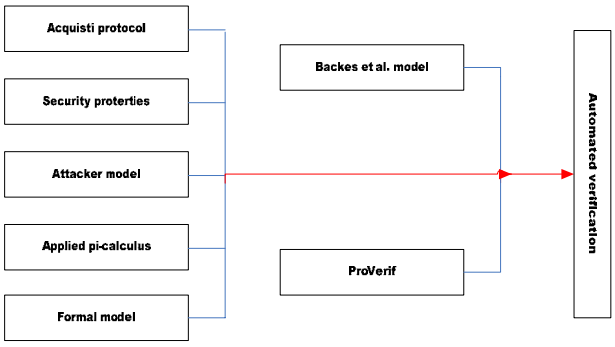


Figure 1. Model of automatic verification of Acquisti protocol

Acquisti protocol is modeled with applied pi calculus [23]. Our choice is based on the fact that applied pi calculus allows the modeling of relations between data in a simple and precise manner using equational theories over term algebra. There, the security properties model is equivalence between processes, while the attacker is thought as an arbitrary process running in parallel with the protocol process representing the adversary model, which is the parallel composition of the protocol participants' processes. The considered attacker is stronger than the basic Dolev_Yao attacker since it can exploit particular relations between the messages by using particular equational theories stating the message relations

III .AUTOMATIC VERIFICATION OF ACQUISTI PROTOCOL WITH PROVERIF

In order to prove the security properties including soundness and coercion-resistance in Acquisti protocol the applied pi model are needed to be translated into the syntax of ProVerif and generated the ProVerif code. The proof of soundness and coercion-resistance in Acquisti protocol is finished by ProVerif. Due to a lack of space, here we only provide the result in Figure 2 and in Figure 3 of our analysis. Acquisti protocol has the soundness property. According to definition of coercion-resistance in Backes et al. model, coercion-resistance is composed of one hypothesis and four conditions. According to our result of analysis we can found that Acquisti protocol has the coercion-resistance with the assumption that the channel between registration authority and coerced voter is private channel. If the channel is public then the coercer could easily distinguish real from fake registration secrets, thus the conditions of coercion-resistance is not satisfied.

IV .CONCLUSION AND FUTURE WORK

Internet voting protocol play an important role in remote voting system. Acquisti protocol is the famous typically remote internet voting protocol that claims to satisfy formal definitions of key properties, such as soundness, individual verifiability, as well as receipt-freeness and coercion resistance without strong physical constrains. But in Acquisti' paper the analysis of its claimed security properties is finished by hand which depends on experts' knowledge and skill and is prone to make mistakes. Recently owning to the contribution of Backes et al, Acquisti protocol can be analyzed with automatic tool. In this paper Acquisti protocol is modeled in applied pi calculus and security properties, including soundness and coercion resistance, are verified with ProVerif. Acquisti protocol has the soundness property in Figure2. At the same time it has the coercion-resistance property in Figure 3 in constrains that the attacker can not eavesdrop the channel between voters and registration authority. To our best knowledge, we are conducting the first automated analysis of Acquisti protocol for an unbounded number of honest and corrupted voters.

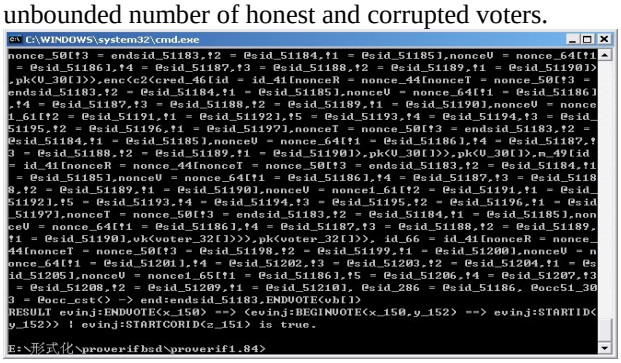


Figure 2. The result of soundness

As future work, we plan to analyze other remote internet voting protocols, such as Meng et al. protocol [12] recently proposed and the protocol claimed that has the

soundness, receipt-freeness and coercion-resistance without physical assumption. It would also be interesting to formalize the security properties in wireless communication protocol in the formal model with ProVerif.

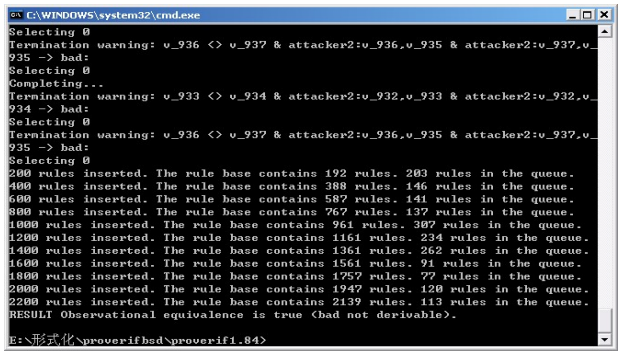


Figure 3. The result of coercion-resistance-condition2 with in constrains that the attacker can not eavesdrop the channel between the voter and the registration authority.

REFERENCES

[1] J.Benaloh, and D.Tuinstra, "Receipt-free secret-ballot elections," *In Proceeding of the Twenty-Sixth Annual ACM Symposium on Theory of Computing*. May23-25, 1994. Montréal, Québec, Canada.pp.544–553, 1994.

[2] A.Juels, and M.Jakobsson, "Coercion-resistant electronic elections," 2002. http://www.vote-auction.net/VOTEAUCTION/165.pdf.

[3] D. Chaum, P.Y.A. Ryan, and S. Schneider, "A Practical Voter-Verifiable Election Scheme," *In Proceeding of ESORICS 2005*. September 12 - 14, 2005. Milan, Italy. pp.118–139, 2005.

[4] R.L.Rivest, "The Threeballot voting system," 10/1/2006. http://people.csail.mit.edu/rivest/Rivest-TheThreeBallotVotingSystem.pdf.

[5] J. Cichoń, M. Kutyłowski, and B. Weglorz, " Short Ballot Assumption and Threeballot Voting Protocol," *In Proceeding of SOFSEM: Theory and Practice of Computer Science*. January 19-25, 2008. Nový Smokovec, Slovakia. Pp.585-598, 2008.

[6] E. Magkos, M.Burmester, and V. Chrissikopoulos, "Receipt-freeness in large-scale elections without untappable channels," *In Proceeding of the IFIP Conference on Towards The E-Society: E-Commerce, E-Business, E-Government*. October 3 - 5, 2001. Zürich, Switzerland. pp.683–694,2001.

[7] D. Chaum, "Secret-Ballot Receipts: True Voter-Verifiable Elections," *IEEE security and privacy*.2004, 2(1):38-47, 2004.

[8] A.Juels, D. Catalano, and M. Jakobsson, "Coercion-resistant electronic elections," *In Proceedings of Workshop on Privacy in the Electronic Society*. November 7, 2005.Alexandria, USA.pp.61-70, 2005.

[9] A. Acquisti, "Receipt-Free Homomorphic Elections and Write-in Voter Verified Ballots," *Technical Report* 2004/105, International Association for Cryptologic Research, May 2, 2004, and Carnegie Mellon Institute for Software Research International, CMU-ISRI-04-116, 2004.

[10] B.Meng, "An Internet Voting Protocol with Receipt-free and Coercion- resistant," *In Proceeding of IEEE 7th International Conference on Computer and Information Technology*. October 16-19, 2007. University of Aizu, Fukushima Japan.pp.721-726, 2006.

[11] B. Meng, "A Secure Internet Voting Protocol Based on Non-interactive Deniable Authentication Protocol and Proof Protocol that Two Ciphertexts are Encryption of the Same Plaintext," *Journal of Networks* .2009, 4(5): 370-377,2009.

[12] B.Meng, Z.M. Li, and J. Qin, "A Receipt-free Coercion-resistant Remote Internet Voting Protocol without Physical Assumptions through Deniable Encryption and Trapdoor Commitment Scheme," *Journal of Software* .2010, In press.

[13] D.Kindred, *Theory Generation for Security Protocols*. Doctoral Thesis. UMI Order Number: AAI9935996. Carnegie Mellon University, 1999.

[14] G. Lowe, "Casper: A complier for the analysis of security protocols," *Journal of Computer Security*,1998,6 (1):53-84, 1998.

[15] P. Maggi, and R. Sisto, "Using SPIN to Verify Security Properties of Cryptographic Protocols" *In Proceedings of the 9th international SPIN Workshop on Model Checking of Software* (April 11 - 13, 2002). LNCS, vol. 2318. Springer-Verlag, London, pp.187-204, 2002.

[16] K. L.McMillan, *Symbolic Model Checking: An Approach to the State Explosion Problem*. Doctoral thesis .Carnegie Mellon University, 1992.

[17] C.A.Meadows, "The NRL Protocol Analyzer: an overview," J. Logic Programming 26 (2) (1996) pp.113-131, 1996.

[18] E.M.Clarke, S.Jha, and W.Marrero, "Verifying security protocols with Brutus," *ACM Trans. Softw. Eng. Methodol.* 9, 4 (Oct. 2000), 443-487, 2000.

[19] C.Joseph, F.Cremers, "Scyther - Semantics and Verification of Security Protocols,"http://alexandria.tue.nl/extra2/200612074.pdf

[20] L.C. Paulson, "Isabelle Reference Manual," http://isabelle.in.tum.de/doc/ref.pdf,2009.

[21] D.X.Song, "Athena: a New Efficient Automatic Checker for Security Protocol Analysis," *In Proceedings of the 12th IEEE Workshop on Computer Security Foundations* (June 28 - 30, 1999). CSFW. IEEE Computer Society, Washington, DC, 192.

[22] B. Blanchet, "An Efficient Cryptographic Protocol Verifier Based on Prolog Rules," *In Proceedings of the 14th IEEE Computer Security Foundations Workshop* (CSFW-14), pp. 82-96, Cape Breton, Nova Scotia, Canada, June 2001. IEEE Computer Society.

[23] M. Abadi, and C. Fournet, "Mobile values, new names and secure communication," *In Proceeding of the 28th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages*, 2001, London, UK. pp.104-115, 2001.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 151-154

Recent Advances in Cloud Storage

Jiyi Wu[1,2], Jianlin Zhang[1],Zhijie Lin[2,3], Jiehui Ju[3]
1. Key Lab of E-Business and Information Security, Hangzhou Normal University,Hangzhou,China
Email:Dr_PMP@yahoo.com.cn
2.School of Computer Science and Technology, Zhejiang University,Hangzhou,China
3.School of Information and Electronic Engineering, Zhejiang University of Science and Technology,Hangzhou,China
Email: zhangjohn@vip.sina.com, jianqing_fu@zju.edu.cn, lin@zju.edu.cn

***Abstract*—As the latest development of the distributed storage technology, Cloud storage is the product of the integration of distributed storage and virtualization technologies. Cloud storage is a method that allows you to use storage facilities available on the Internet. There are ten critical common denominators that must be considered to make cloud storage valuable. A typical cloud storage system architecture includes a master control server and several storage servers. With the advent of cloud computing, multi-tenancy has simply been extended to include any cloud architecture—that supports multiple tenants.**

***Index Terms*—Recent Advances, Cloud Computing, Cloud Storage, Multi-Tenancy, reference model**

I. INTRODUCTION

One of IT's biggest expenses is disk storage. ComputerWorld estimates that in many enterprises storage is responsible for almost 30% of capital expenditures as the average growth of data approaches close to 50% annually in most enterprise. Amid this milieu, there's strong concern that enterprise will drown in the expense of storing data, especially unstructured data.

To address this need, Cloud storage services have started to become popular. Ranging from Cloud storage focused at the enterprise to that focused on end users, Cloud storage providers offer huge capacity cost reductions, the elimination of labor required for storage management and maintenance, and immediate provisioning of capacity at a very low cost per terabyte.

Cloud storage, though, is not a brand new concept. The central ideas for Cloud storage are related to past service bureau computing paradigms and to those of application service providers and storage service providers of the late 90's.

This time, however, the economic situation and the advent of new technologies have sparked strong interest in the Cloud storage provider model. With on-premises storage costs already high and rising in many IT departments, Cloud storage providers can lower cost by off-loading the burden of storage management and shielding enterprises from other costs as well, such as storage and network hardware changes. Cloud storage providers deliver economies of scale by using the same storage capacity to meet the needs of many organizations, passing the cost savings to their customer base.

Cloud Storage is part of a wider definition called Cloud Computing which, according to the National Institute of Standards and Technology, is "a model for enabling convenient, on demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction".

The service models are divided in Cloud Software as a Service (SaaS), Cloud Platform as a Service (PaaS) and Cloud Infrastructure as a Service (IaaS).

Computing resources like servers and network can be replaced, but the core of most of the organizations is the information, usually stored in data centers. For this reason security and availability are the first issues when companies are deciding to migrate part of their data to the cloud, generally by the internet.

This kind of precaution is not so different from the one when data is stored in private data centers, but there are some analysis concerned to this migration to public cloud that need to done by corporations and service providers.

II. CLOUD STORAGE INFRASTRUCTURE REQUIREMENTS

When you combine the technology trends such as virtualization with the increased economic pressures, exploding growth of unstructured data and regulatory environments that are requiring enterprises to keep data for longer periods of time, it is easy to see the need for a trustworthy and appropriate storage infrastructure. Whether a cloud is public or private, the key to success is creating a storage infrastructure in which all resources can be efficiently utilized and shared.

Because all data resides on the storage systems, data storage becomes even more crucial in a shared infrastructure model. There are ten critical common denominators that must be considered to make cloud storage valuable. These include:

A. Elasticity

Cloud storage must be elastic to rapidly adjust the underlying infrastructure to changing subscriber demands and comply with Service Level Agreements (SLAs).

B. Automatic

Cloud storage must have the ability to be automated so that policies can be leveraged to make underlying infrastructure changes such as placing user and content

Corresponding Author: Jiyi WU

management in different storage tiers and geographic locations quickly and without human intervention.

C. Scalability

Cloud storage needs to scale quickly and to tremendous capacities. This translates into scalability across objects, performance, users, clients, and capacity with a single name space across all storage capacity being critical for low Opex reasons.

D. Data Security

For private clouds, security is assumed to be tightly controlled. For public clouds, data should either be stored on a partition of a shared storage system, or cloud storage providers must establish multi-tenancy policies to allow multiple business units or separate companies to securely share the same storage hardware.

E. Performance

A proven storage infrastructure providing fast, robust data recovery is an essential element of a cloud service.

F. Reliability

Enterprise users also want to make sure that their data is reliably backed up for disaster recovery purposes and that it meets pertinent compliance guidelines.

G. Ease of Management

The need for improved manageability in the face of exploring storage capability and costs is a major benefit enterprises are expecting from cloud storage deployment.

H. Ease of Data Access

Ease of access to data in the cloud is critical in enabling seamless integration of cloud storage into existing enterprise workflows and to minimize the learning curve for cloud storage adoption.

I. Energy Efficiency

IT datacenters are growing bottlenecks and approaching ceilings on available power, cooling and flooring space. Green storage technology is the technology that enables energy efficiency and waste reduction in storage solutions leading to an overall lower carbon footprint.

J. Latency

Not all applications are suitable for a Cloud storage model. It is important to measure and test network latency before committing to a migration. Virtual machines can introduce additional latency through the time-sharing nature of the underlying hardware and unanticipated sharing and reallocation of machines can significantly affect run times.

III. MULTI-TENANCY CLOUD STORAGE REFERENCE MODEL

A. Typical cloud storage system architecture

A typical cloud storage system architecture includes a master control server and several storage servers, as shown in Fig 1.

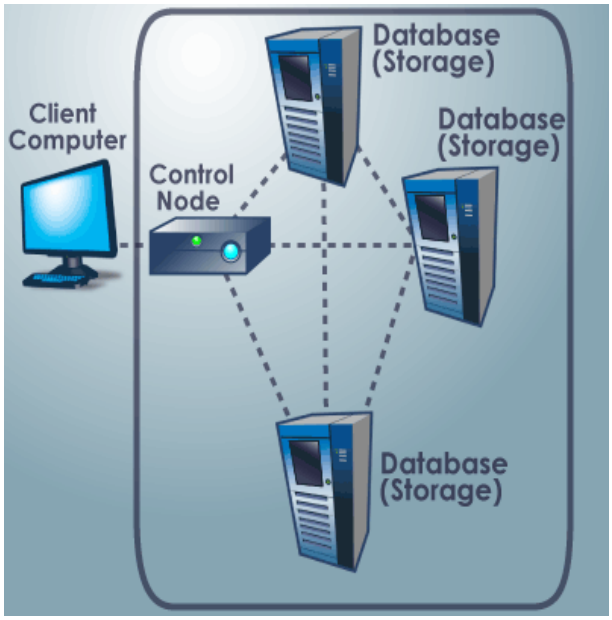


Figure 1. A typical Cloud Storage system architecture

For some computer owners, finding enough storage space to hold all the data they've acquired is a real challenge. Some people invest in larger hard drives. Others prefer external storage devices like thumb drives or compact discs. Desperate computer owners might delete entire folders worth of old files in order to make space for new information. But some are choosing to rely on a growing trend: cloud storage.

While cloud storage sounds like it has something to do with weather fronts and storm systems, it really refers to saving data to an off-site storage system maintained by a third party. Instead of storing information to your computer's hard drive or other local storage device, you save it to a remote database. The Internet provides the connection between your computer and the database.

On the surface, cloud storage has several advantages over traditional data storage. For example, if you store your data on a cloud storage system, you'll be able to get to that data from any location that has Internet access. You wouldn't need to carry around a physical storage device or use the same computer to save and retrieve your information. With the right storage system, you could even allow other people to access the data, turning a personal project into a collaborative effort.

So cloud storage is convenient and offers more flexibility, but how does it work? Find out in the next section.

B. Cloud Storage reference model

The appeal of cloud storage is due to some of the same attributes that define other cloud services: pay as you go, the illusion of infinite capacity (elasticity), and the simplicity of use/management. It is therefore important that any interface for cloud storage support these attributes, while allowing for a multitude of business cases and offerings, long into the future.

The model created and published by the Storage Networking Industry Association™ ,shows multiple

types of cloud data storage interfaces able to support both legacy and new applications. All of the interfaces allow storage to be provided on demand, drawn from a pool of resources. The capacity is drawn from a pool of storage capacity provided by storage services. The data services are applied to individual data elements as determined by the data system metadata. Metadata specifies the data requirements on the basis of individual data elements or on groups of data elements (containers).

As shown in Fig 2, the SNIA Cloud Data Management Interface (CDMI) is the functional interface that applications will use to create, retrieve, update and delete data elements from the cloud. As part of this interface the client will be able to discover the capabilities of the cloud storage offering and use this interface to manage containers and the data that is placed in them. In addition, metadata can be set on containers and their contained data elements through this interface.

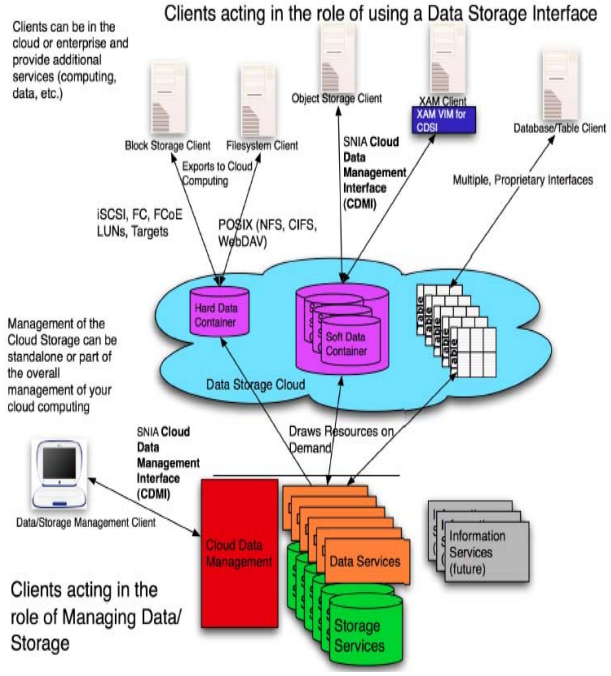


Figure 2.Cloud Storage reference model

It is expected that the interface will be able to be implemented by the majority of existing cloud storage offerings today. This can be done with an adapter to their existing proprietary interface, or by implementing the interface directly. In addition, existing client libraries such as XAM can be adapted to this interface as show in Figure 2.

This interface is also used by administrative and management applications to manage containers, accounts, security access and monitoring/billing information, even for storage that is accessible by other protocols. The capabilities of the underlying storage and data services are exposed so that clients can understand the offering.

Conformant cloud offerings may offer a subset of either interface as long as they expose the limitations in the capabilities part of the interface.

C. Multi-Tenancy Cloud Storage

The terms multi-tenant and multi-tenancy are not new; both have been used to describe application architectures designed to support multiple users or "tenants" for many years. With the advent of cloud computing, this terminology has simply been extended to include any cloud architecture—or infrastructure element within that architecture (application, server, network, storage)—that supports multiple tenants. Tenants could be separate companies, or departments within a company, or even just different applications.

To provide "secure" multi-tenancy and address the concerns of cloud skeptics, a mechanism to enforce separation at one or more layers within the infrastructure is required:

- **Application layer.** A specially written, multi-tenant application or multiple, separate instances of the same application can provide multi-tenancy at this level.
- **Server layer.** Server virtualization and operating systems provide a means of separating tenants and application instances on servers and controlling utilization of and access to server resources.
- **Network Layer.** Various mechanisms, including zoning and VLANs, can be used to enforce network separation. IP security (IPsec) also provides network encryption at the IP layer (application independent) for additional security.
- **Storage Layer.** Mechanisms such as LUN masking and SAN zoning can be used to control storage access. Physical storage partitions segregate and assign resources (CPU, memory, disks, interfaces, etc.) into fixed containers.

Achieving secure multi-tenancy may require the use of one or more mechanisms at each infrastructure layer.

While mechanisms to support multi-tenancy and enforce separation exist at every infrastructure layer, this paper is primarily concerned with storage and the requirements for secure and effective storage multi-tenancy in a cloud environment. To understand the full set of storage requirements, it is necessary to consider cloud storage from both the perspective of the tenant (user) and the provider of cloud services.

Cloud computing services can be broken down into a variety of types, ranging from Software as a Service (SaaS)—in which the provider delivers specific application services to each tenant—to Data storage as a Service (DaaS) —which is virtualized storage on demand over a network. Regardless of the type of cloud service, from a tenant perspective there will be specific requirements that apply directly or indirectly to data storage.

Tenant requirements are typically defined in terms of service level agreements (SLAs), which cover a variety of capabilities including:

- Security
- Performance
- Data protection and availability
- Data management

From the provider's perspective, multi-tenant storage should provide convenient mechanisms for satisfying these and other tenant SLAs as well as supporting additional capabilities such as:

- **Accounting.** The ability to monitor usage by each tenant for billing or other purposes.
- **Self service.** The ability to allow a tenant to perform a defined set of management tasks on their data and the storage they use, thereby offloading these functions from the provider.
- **Non-disruptive upgrades and repairs.** Downtime in multi-tenant environments may be difficult or impossible to schedule, so maintenance activities must be possible without incurring downtime from the point of view of the tenant.
- **Performance management.** The ability to balance cost and performance as the lifecycle requirements of data changes over time.

Designed to enable multi-tenant storage offerings, the SNIA's Cloud Data Management Interface (CDMI) for cloud storage and data management integrates and is interoperable with various types of client applications. CDMI offers a standard approach to data portability, compliance and security, as well as the ability to connect one cloud provider to another, enabling compatibility between cloud vendors.

Using this approach, a client will be able to discover the capabilities of cloud storage and use this interface to manage data containers and the data elements that are placed in them. CDMI makes extensive use of metadata to simplify application access and enable multiple levels of service as required by a diverse set of users.

In the storage layer, the CDMI interface can simplify management since data system metadata can be applied to container hierarchies. For the functional data path interface for data storage, CDMI assigns each data object a separate URI (Uniform Resource Identifier). Since objects can be fetched using the standard HTTP protocol employing RESTful (REpresentational State Transfer) operations, each data element can be managed as a separate resource. In this way, it is possible to separate and classify data elements and containers for secure access as well as service levels. The result is a level of isolation suitable to tenant based, on-demand data access.

VI. Conclusions and Future Work

Cloud Storage with a great deal of promise, aren't designed to be high performing file systems but rather extremely scalable, easy to manage storage systems. They use a different approach to data resiliency, Redundant array of inexpensive nodes, coupled with object based or object-like file systems and data replication (multiple copies of the data), to create a very scalable storage system.

This article gives a quick introduction to cloud storage. It covers the key technologies in Cloud Computing and Cloud Storage,several different types of clouds services, and describes the advantages and challenges of Cloud Storage after the introduction of the Cloud Storage reference model.

Acknowledgment

Funding for this research was provided in part by the Scientific Research Program of Zhejiang Educational Department under Grant No.20071371. We like to thank anonymous reviewers for their valuable comments.

References

[1] Luis M.Vaquero,Luis Rodero-Merino, Juan Caceres,Maik Lindner. A Break in the Clouds: Toward a Cloud Definition. ACM SIGCOMM Computer Communication Review,2009,39(1):50-55.

[2] Wu Jiyi,Ping Lingdi,Pan Xuezeng.Cloud Computing: Concept and Platform, Telecommunications Science, 12:23-30, 2009.

[3] Jonathan Strickland.How Cloud Storage Works[OL], http://communication.howstuffworks.com/cloud-storage.htm, 2010.

[4] Storage Networking Industry Association.Cloud Storage Reference Model,Jun.2009.

[5] Storage Networking Industry Association.Cloud Storage for Cloud Computing,Jun.2009.

[6] Luiz Andre Barroso,Jeffrey Dean,Urs Holzle.Web search for a planet: The Google cluster architecture. IEEE Micro, 2003,23(2):22−28.

[7] Sanjay Ghemawat, Howard Gobioff, and Shun-Tak Leung. The Google file system. In: Proc. of the 19th ACM SOSP. New York: ACM Press, 2003. 29−43.

[8] Robert L.Grossman, Yunhong Gu, Michael Sabala,Wanzhi Zhang. Compute and storage clouds using wide area high performance networks. Future Generation Computer Systems, 2009,25(2):179-183.

[9] Yunhong Gu and Robert L.Grossman. Sector and Sphere: the design and implementation of a high-performance data cloud. Philosophical Transactions of the Royal Society. A(2009)367:2429-2445.

[10] Robert L Grossman,Yunhong Gu. Data Mining Using High Performance Data Clouds: Experimental Studies Using Sector and Sphere. In Proc. of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining,2008, 920-927.

[11] Daniel J. Abadi. Data Management in the Cloud: Limitations and Opportunities. Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, 2009,32(1):3-12.

[12] Peter Mell and Tim Grande. NIST. Retrieved from http://csrc.nist.gov/groups/SNS/cloud-computing/cloud-def-v15.doc,2010.

[13] S Lesem. Cloud Storage Strategy Retrieved from http://cloudstoragestrategy.com/2009/07/security-and-cloud-storage-everybody-talksabout-it-but-is-it-really-all-that-different.html,2010.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 155-158

Electromagnetic thrust angle controller for Permanent Magnet Linear synchronous Motor drive system

Wang Fu-zhong[1], Kang Hong-chao[2]
[1] School of Electrical Engineering and Automation of Henan Polytechnic University,Jiaozuo,China.
e-mail: wangfzh@hpu.edu.cn
[2] School of Electrical Engineering and Automation of Henan Polytechnic University,Jiaozuo,China.
e-mail: kang880318@163.com

***Abstract*—According to the analyzing the thrust-angle characteristics of PMLSM for vertical movement, we obtained the power angle control strategy and designed the electromagnetic power angle controller. The simulation results indicate that the stability, credibility and running efficiency of the vertical transportaion of PMLSM are impoved significantly, and the loss of synchronization of PMLSM is prevented efficiently.**

***Index Terms*—f vertical movement, PMLSM, power angle controller, inverter**

I. INTRODUCTION

Permanent Magnet Linear Synchronous Motor is the first choice for servo system directly driven by linear motor because of its simple structure, high efficiency and great ratio of thrust to volume. But currently,research and use of PMLSM in our country in still in its beginning stage, its theory and control is not yet perfect. In addition, the structure and control mechanism of PMLSM for vertical movement is different from Rotary Motor and PMLSM for horizontal movement. So larger error and poor stability will get if we adopt the control method of Rotary Motor and PMLSM for horizontal movement to control the PMLSM for vertical movement. The electromagnetic thrust angle controller designed in the paper controls the speed of the motor by controlling the electromagnetic thrust angle directly, the response is fast and the stability and reliability of PMLSM for vertical movement is improved significantly.

THE THRUST-ANGLE CHARACTERISTICS OF SYNCHRONOUS MACHINE

When syschronous machine get into the steady state,the most important characteristic is the power–angle, namely, the relationship between electromagnetic thrust and electromagnetic power-angle, which is equivalent to the angle between No-load EMF $\dot{E}_0$ and phase voltage $\dot{U}$ of synchronous machine.

Equivalent circuit of PMLSM is shown in Fig.1

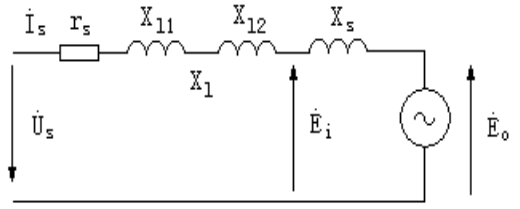


Fig. 1 Equivalent circuit of PMLSM

where $\dot{U}_s$ is phase voltage given to the armature windings, $\dot{I}_s$ is the armature current, X_s is the armture reactances, $\dot{E}_i$ is EMF inside, $\dot{E}_0$ is EMF outside, X_l is leakage reactance of armature, which is composed of slot leakage reactance X_{l1} and terminal leakage reactance, namely, $X_l = X_{l1} + X_{l2}$ .According to the motor practice,the voltage equation of the equivalent circuit is:

$$\dot{U}_s = -\dot{E}_0 + \dot{I}_s r_s + j\dot{I}_s X_l + j\dot{I}_s X_s \qquad (1)$$

Where X_T is Synchronous reactance, and $X_T = X_l + X_s$,From the equ.(1) we can obtain:

$$\dot{I}_s = \frac{\dot{U}_s - (-\dot{E}_0)}{r_s + jX_T} = I_P + jI_Q \qquad (2)$$

Where I_P is the active component of armature current, I_Q is reactive component of armature current. Synchronous impedance $Z = \sqrt{r_s^2 + X_T^2}$.

According to Equ.(2),we can know the electromagnetic power and the electromagnetic force of PMLSM are:

Wang Fuzhong, born in mengzhou, Henan in 1961, is now professor in Henan Polytechnic University and associate of Electrical Engineering and Automation, Master's tutor, The main research directions are the industrial process microcomputer control,electric system automaion and electric machine and electric appliance. (phone:0391-3987555; fax:-0391-3987552; e-mail:Wangfzh@hpu.edu.cn)

Kang Hongchao, is now master in Henan Polytechnic University. The main research direction is industrial process contol.. (phone:15939134059; e-mail:kang880318@163.com).

This work was supported by Provincial open laboratory for control engineering key disciplines(KG2009-05)

$$P_m = 3E_0 I_P$$
$$= \frac{3}{Z^2}\left[(E_0 U_s \cos\theta - E_0^2) r_s + X_T E_0 U_s \sin\theta \right] \quad (3)$$

$$F_x = \frac{3E_0}{v_s Z^2}(U_s r_s \cos\theta + X_T U_s \sin\theta - E_0 r_s)$$
$$= \frac{3E_0 U_s}{v_s Z}\sin(\theta+\alpha) - \frac{3E_0^2 r_s}{v_s Z^2} \quad (4)$$

where θ is electromagnetic power-angle, which is the angle between No-load EMF $\dot{E}_0$ and phase voltage $\dot{U}$. $\alpha = arctan\frac{r_s}{X_T}$, Eq.(4) is a sine curve, when $\theta = 90^\circ$,the electromagnetic trust is greatest.

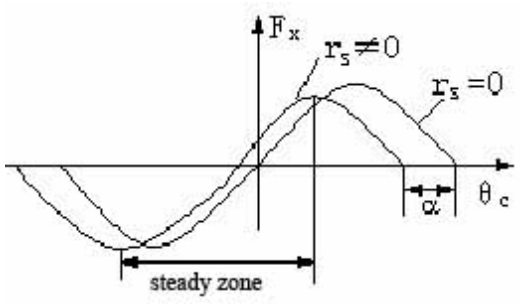


Fig.2 thrust-angle curve

Generally, the resistance of armature circuit is much less than synchronous reactance. namely, $r_s << X_T$, r_s neglected, the electromagnetic trust will be:

$$F_x = \frac{3E_0 U_s}{v_s X_T} sin\theta \quad (5)$$

The corresponding curve is shown in Fig.2.

When the working frequency of PMLSM is significantly decreased, the resistance drop on the armature will be significantly increased, leading to the thrust calculated by Eq.1 unaccurate. For example, when the working frequency is 10Hz, $X_T = 9.83\Omega$, $r_s = 2.985\Omega$, here the effect of the resistance can not be neglected.

If we move the thrust-angle curve when $r_s \neq 0$ to the right a placement α , then move up a placement $3E_0^2 r_s / v_s Z^2$,we will get the curve when $r_s = 0$. Considering the role of armature resistance, the maximum thrust and the angle when the thrust is maximum will decrease, the zone of the thust where the mover get forward rotation is narrower than reversal rotation.

THE CONTROL PRINCIPLE OF ELECTROMAGNETIC THRUST ANGLE FOR PMLSM

The electromagnetic thrust of PMLSM is:

$$F_x = \frac{3E_0 U_s}{v_s X_T} sin\theta \quad (6)$$

$$\theta = \omega t - \frac{\pi}{\tau} x \quad (7)$$

where τ is Polar distance, w is angle velocity, x is the actual distance (the position of the mover) between axis of the pole and axis of the A phase windings.

When $\theta = \pi/2$, the thrust a current of 1 ampere induced is the greatest, but for PMLSM, if it is working in this point ,the machine will get into the unstable region. when disturbance appears, then the mover will slipped down accelerating,which is dangerous. To ensure the stable running of the system, the reaction of the system must be fast and the Adjusted Time must be short,but it is difficult to realize. However, the control strategy, adopted in the paper, sets a constant angle $\theta_N = 33^\circ$ in accordance with the demand of the stability, to keep the machine running in the steady zone and the thrust is greastest. In fact, we should judge whether the machine get into the unsteady zone at any stage by the angle measured. Because once the the systm works near the critical stability point, the machine may get into the unstable region, leading loss of synchronization. The traditional control method, at this time, prevents the problem from happening by increasing the field current or the amplitude of armature current. But since the mover of testing machine is permanent magnet, it is not practicable to increase field current, and it is also not easy to change the amplitude of armature current adopting transducer to adjust the speed. Besides, it will bring about a series of problems by increasing the amplitude of armature current, such as increasing the loss of the stator, high temperature of windings and lower insulating capability, which will threaten the safety of the machine. According to literature 3, when the supply voltage remains the same, the maximum thrust will increase by $1/K$ while the supply frequency increase by K . Therefore, when the machine is getting into the unstable region because of disturbances, we can reduce the power supply frequency in order to increase the elextromagnetic force to prevent loss synchronization.

In Equ.(7), ωt can be regard as the expecting positon angle of Permanent Magnet Linear Synchronous Motor,

$$\omega t = \pi x_i / \tau \quad (8)$$

$$\omega = v_s \pi / \tau \quad (9)$$

where v_s is the preset value of speed of PMLSM, x_i is the expecting position of mover, so Eq.(7) can be changed as:

$$\theta = \frac{\pi}{\tau}(\upsilon_s t - x) = \omega t - \frac{\pi}{\tau} x \quad (10)$$

Therefore, having known the distance x , time t and angular velocity ω the mover travelled, we can get the position(power angle) of PMLSM according to Eq.(7). Then comparing the preset value of electromagnetic thrust-angle with measured value to produce the preset frequency we need, by which we can keep the linear

motor working in the stable region and electromagnetic thrust meeting the load requirement.

Ⅳ THE DESIGN AND SIMULATION OF ELECTROMAGNETIC FORCE ANGLE CONTROLLER

Since the operating frequency of testing machine is low, the frequency can not be too low when the supply voltage is constant. otherelse, the machine will heated, which will effect the safely running of the system.. At this time, we can improve the thrust to prevent the loss of synchronization by changing the ratio between supply voltage and frequency. Meanwhile, increasing the intensity of voltage compensation in the low frequency region must be also considered. The reference ratio between supply frequency and voltage of testing machine is set to 6Hz / 270V. According to theoretic calculation, the thrust can be increased by 3 times when the ratio is 0.6Hz/137, the thrust keeps constant when the ratio is 1Hz / 112V, the thrust can be increase d by 2 times when the ratio is 2Hz / 206V，the thrust can be increased by 1.5 times when the ratio is 3Hz / 222V and keeps constant when the ratio is 5Hz / 238V.

Mitsubishi FR-A241 inverter have the feature of adjustable five-point V/f, namely, it can set the V/f(frequency voltage/frequency) to five points, V/f1-V/f5,in which linear interpolation is applied. By this five different points, we can change the V/f characteristics to meet the requirements of power angle control.

Generally,when the working frequncy of inverter is smaller than the fundamental frequency, the V/f characteristics must be constant,as is shown in Fig.1. However, the resistance and inductance of the stator windings will produce big voltage drop at low frequency of low voltage, which will make the motor torque(thrust) much smaller than working in fundamental frequency voltage. Thus we need to increase compensation for voltage to increase output torque, voltage compensation V is called torque promotion, which can be set value from 0 to 30% by Pr.0 of inverter, according to calculations, we set Pr.0 to 30%, taking characteristics of testing machine at low frequency into consideration. The curve compensated is showm in Fig.3 .Adjustable five-point V/f characteristics and V/f values setted is illustrated in fig.4 and Table 1.

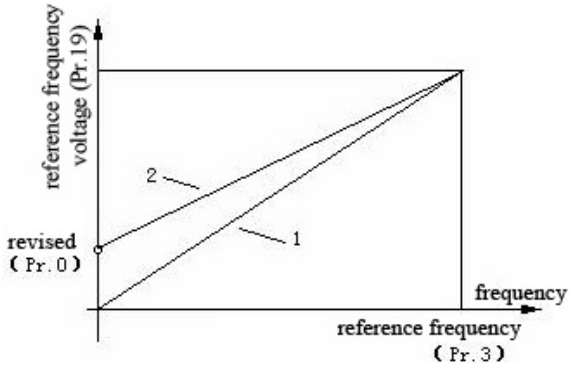


Fig. 3 The V/f characteristic of inverter

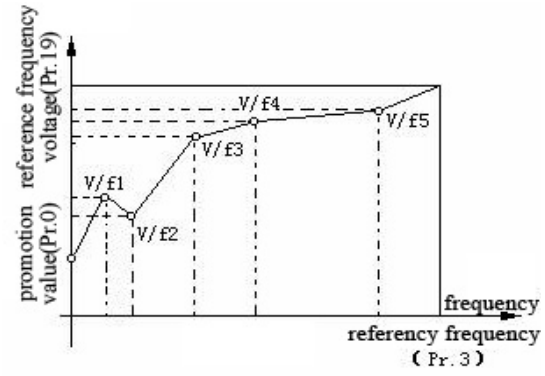


Fig. 4 the V/f 5 points adjustability of inverter

Table 1 the inverter's V/f parameters setting

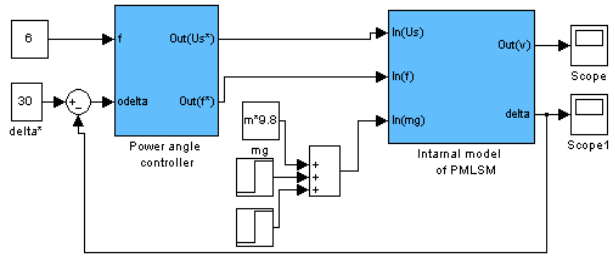
Parameter	Function	Preset
Pr.100	V/f1(frquency 1)	0.6Hz
Pr.101	V/f1(frequency voltage 1)	137V
Pr.102	V/f2(frequency 2)	1Hz
Pr.103	V/f2(frequency voltage 2)	112V
Pr.104	V/f3(frequency 3)	2Hz
Pr.105	V/f3(frequency voltage 3)	206V
Pr.106	V/f4(frequency 4)	3Hz
Pr.107	V/f4(frequency voltage 4)	222V
Pr.108	V/f5(frequency 5)	5Hz
Pr.109	V/f5(frequency voltage 5)	238V

When the rated voltage frequency is set to 6Hz, the air-gap is set to 8mm，the maximum thrust can be obtained when the electromagnetic power angle is 63 degrees. we should always detect the speed and power angle of the motor when it is running, then according to the difference between actual value and preset value of power angle.adjusted the supply frequency to meet the requirement of the thrust and prevent the problem of out of step. The rule of frequency modulation and control based on theoretical calculations and control experience is shown in Table 2.

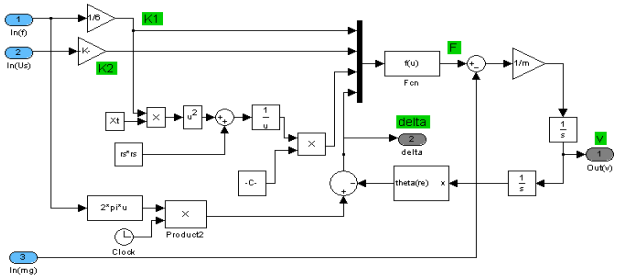
Table 2 the rule of frequency adjustment

f^* \ $\Delta\theta$ / f	$\Delta\theta<0$	[0, 10)	[10, 25)	[25, 30)	$\Delta\theta\geq 30$
$0\leq f<0.6$	f	stop	stop	stop	stop
$0.6\leq f<2$	f	0. 6	0. 6	stop	stop
$2\leq f<3$	f	2	2	0. 6	stop
$3\leq f\leq 6$	f	3	2	0. 6	stop

Matlab simulation model by electromagnetic force angle control strategy is shown in Fig.5, the simulation results in shown in Fig.6

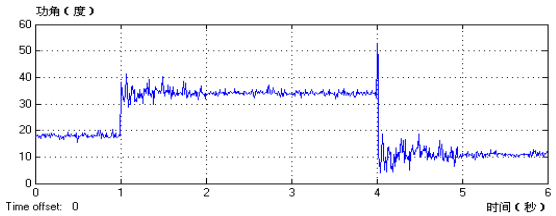


(a) whole system

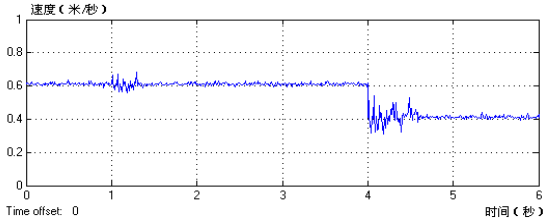


(b) internal model of PMLSM

Fig. 5 the block diagram of power-angle control strategy



(a)the curve of power-angle



(b) the curve of speed

Fig.6 the simulating curve of power-angle control strategy

V CONCLUSION

（1）From the simulation curve, we can see that the power angle controller designed enhanced the stability and reliability of PMLSM for vertical movement.

（2）The controller designed in this paper controls the electromagnetic force angle directly, when the power angle is reduced, the speed of the motor can be almost slowed down at the same time. so we can prevent the out of step of PMLSM for vertical movement significantly by using the relationship between supply frquency and force-angle characteristics.

REFERENCES

[1]Lee,Seung-Hoon;Janq,Ki-Bong;Kim,Gyu-tak,The study of dynamic characteristic of PMLSM according to variable load,[J].Transactions of the Korean Institute of Electrical Engineers,2008.

[2]Yu Yunyue, Principle and Application of Linear Motor,[M] Beijing: China Machine Press, 2000.

[3]Wang Fu Zhong, Wang Xu Dong, Jiao Liu Cheng. Research on control strategy of electromagnetic force angle and the maximum of thrust of permanent magnet linear synchronous motor for vertical movement [J]. JOURNAL OF CHINA COAL SOCIETY 2001,26 (3):307312

[4]Guo Qingding,Wang Chenyuan, Zhou Jiangwen,Sun Tingyu,Linear AC Servo System for Precision Control Technology. Beijing: China Machine Press, 2000

[5]Jiao Liu Cheng, Yuan Shi Ying Study on operating characteristics of permanent magnet linear synchronous motor for vertical movement [J].Proceedings of the CSEE,2002, 22 (4):3740.

[6]Shanguan Xunfeng, Li Qingfu, Yuan Shiying, "Analysis on Characteristics of Permanent Magnet Linear Synchronous Machines with Large armature Resistances and Small Reactance,"in Proceedings of the Eighth International Conference on Electrical Machines and Systems(Volume 1), The Eighth International Conference on Electrical Machines and Systems, Nanjin,2005,pp.434-437.

[7] Zhang Hongwei, Jiao Liucheng, Wang Xinhuan, Wang Fuzhong, Kang Runsheng, ;Research on the control strategy of power angle of permanent magnet linear synchronous motor,"Journal of China Coal Society, Vol.30,No.4.pp.529-533,Aug,2005.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp.159-162

The Comparison of RBF and BP Neural Network in Decoupling of DTG

Luo Yufeng[1], Xu Chao[2], Fan Yaozu[2]
[1]School of Electrical Engineering and Automation, Henan Polytechnic University, Jiaozuo, Henan 454000, China
zhanglyf@126.com
[2]School of Automation Science and Electrical Engineering, Beijing University of Aeronautics and Astronautics, Beijing, Beijing 100191, China
xchao@asee.buaa.edu.cn, fanyaozu@126.com

***Abstract*—In order to improve the precision of gyroscope, two decoupling method of DTG(Dynamic Tuned Gyroscope) were analyzed, the BP neural network and RBF network. The BP neural network has many advantages Compared to the traditional decoupling method, but still some drawbacks such as the over training, the congress process is very slow, and the hidden layer is also hard to determined. The paper introduced the RBF network as the new decoupling method compared with BP network. The simulation result verified that the RBF network is faster than BP, and also the accuracy is much higher.**

***Index Terms*—DTG(Dynamic Tuned Gyroscope), decoupleing, RBF neural network, BP neural network**

I. INTRODUCTION

DTG(Dynamic Tuned Gyroscope) is a dual-input and dual-output inertial device[1], which can measure the two input angular velocities along the axes in the state of force feedback loop. But there is coupling between the two measuring axes and such coupling is twofold because of the mechanical structure, which means that an input angular velocity along one axis will produce rotor angle and torque feedback in two axes. In inertial system, it not only hoped the small drift of the DTG and good linearity, but also no coupling between the two axes, which means that the input angular velocity only produce one output along the corresponding output axis to ensure the accuracy of the system. Although the coupling only has little effect in static test, the survey data will have a greater error when moving base test. Therefore, the decoupling design must be done in torque feedback loop [2]. From the view of control, it is hoped that the input angle had a single corresponding relationship with the output angle along the corresponding axis, thus the tracking of the rotor angle is smooth and rapid. In order to fulfill this requirement, it is needed to control the rotor angle decoupling, and it is also hoped that the input angle had the single corresponding relationship with the current of the moment. So, it is needed to decouple the feedback current, which is named full decoupling[3].

The traditional decoupling method is building the relationship between the input and output signals, but the method is so complicated and low accuracy. BP neural network was introduced to the research of decoupling of DTG and achieved much better result compared with the traditional decoupling method[6]. But, there is still drawbacks of BP neural network, such as the over training, the congress process is very slow, and the hidden layer is also hard to determined. The RBF(Radial basis function) neural network has overcome these drawbacks in many situations, which has only very few nerve cells in local area to determine the output. So, the training of RBF network is very fast and can avoid the iterative calculation process of BP network. Still, there is also no the possibility of local extremum. The learning process is as fast as one thousand time of BP network. This paper will decouple the DTG by RBF and will compare the capabilities of the two networks to verify the ascendance of RBF.

II. ANALYSIS OF DTG COUPLING

From the function viewpoint, the DTG is classified to position gyroscope, the measurement range of the angular displacement is small. So, the DTG is only suitable to platform-INS. There must be a balance loop to make the DTG as a dynamic tuning rate gyroscope[4]. It can deriver the following formula from the transfer function of the DTG:

$$\begin{bmatrix}\beta(s)\\ \alpha(s)\end{bmatrix} = -\frac{1}{s}\frac{1}{C^2+D^2}\begin{bmatrix}C & -D\\ D & C\end{bmatrix}\begin{bmatrix}\dot{\varphi}_x(s)\\ \dot{\varphi}_y(s)\end{bmatrix} \quad (1)$$

$$\begin{bmatrix}I_x(s)\\ I_y(s)\end{bmatrix} = -\frac{1}{s}K\cdot K_{PC}(s)\cdot K_P\cdot\begin{bmatrix}C & -D\\ D & C\end{bmatrix}\begin{bmatrix}\dot{\varphi}_x(s)\\ \dot{\varphi}_y(s)\end{bmatrix} \quad (2)$$

where

$$C = 1 + K_T\cdot K\cdot K_{PC}(s)\cdot K_P\cdot G_1(s)$$

$$D = K_T\cdot K\cdot K_{PC}(s)\cdot G_2(s)$$

$$G_1(s) = \frac{J}{s^2J^2+H^2}$$

$$G_2(s) = \frac{H}{s(s^2J^2+H^2)}$$

$$K_{PC}(s) = K_P\cdot K_1\cdot K_2\cdot K_D\cdot K_L\cdot K_{dc}\cdot K_{dc}(s)$$

Here, K_1 means preamplifier gain, K_2 --Exchange amplifier gain, K_D --Demodulator transfer coefficient, K_L -- Band stop filter gain, K_{dc} -- Correction network gain, $K_{dc}(s)$ --Correction network frequency characteristics,

K --PWM-Transfer Function, K_p --Sensor scaling factor, K_T --Torque scaling factor, J --Gyro equatorial moment of inertia, and H --Gyro angular momentum.

It can be seen that both the rotor angle and the feedback current moment are cross-coupling according to equation (1) and (2).

III. THE TRADITIONAL DECOUPLING METHOD OF DTG

Define,

$$D(s)=\begin{bmatrix} D_{11}(s) & D_{12}(s) \\ D_{21}(s) & D_{22}(s) \end{bmatrix}$$

is the decoupling network transfer function matrix. There is equation (3) according to the DTG closed-loop feedback control system

$$\begin{bmatrix} \beta(s) \\ \alpha(s) \end{bmatrix} = -\frac{1}{s}[I + K_T K K_{PC}(s) K_P G(s) D(s)]^{-1} \begin{bmatrix} \dot{\varphi}_x(s) \\ \dot{\varphi}_y(s) \end{bmatrix} \quad (3)$$

Where

$$G(s) = \begin{bmatrix} G_1(s) & -G_2(s) \\ G_2(s) & G_1(s) \end{bmatrix}$$

Because the DTG two axes is symmetric, the decoupling of the main diagonal matrix elements should be equal. At the same time, in order to make the gyroscope tracking the constant angular acceleration input with the constant bias and track the constant angular velocity input with zero declination, the controlled decoupling matrix should be:

$$D(s)=\begin{bmatrix} \frac{J}{H} & \frac{1}{s} \\ -\frac{1}{s} & \frac{J}{H} \end{bmatrix} \quad (4)$$

Then,

$$\begin{bmatrix} \beta(s) \\ \alpha(s) \end{bmatrix} = -\frac{1}{s}\frac{1}{1+\frac{K_T K K_{PC}(s) K_P}{Hs^2}} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \dot{\varphi}_x(s) \\ \dot{\varphi}_y(s) \end{bmatrix} \quad (5)$$

$$\begin{bmatrix} I_x(s) \\ I_y(s) \end{bmatrix} = -\frac{1}{s}\frac{K K_{PC}(s) K_P}{1+\frac{K_T K K_{PC}(s) K_P}{Hs^2}} \begin{bmatrix} \frac{J}{H} & \frac{1}{s} \\ -\frac{1}{s} & \frac{J}{H} \end{bmatrix} \begin{bmatrix} \dot{\varphi}_x(s) \\ \dot{\varphi}_y(s) \end{bmatrix} \quad (6)$$

According to (5), it can be seen that the control decoupling eliminates the coupling component in the rotor angle declination and recurs the ideal slip angle response. But according to (6), the coupling elements in the current feedback still exist. In order to eliminate the coupled components in current feedback loop, select the output decoupling network under the premise of choosing a decoupling network control. From the principle of decoupling diagonal matrix, there exists:

$$P(s)=\begin{bmatrix} \frac{(H/J)^2}{s^2+(H/J)^2} & \frac{Hs/J}{s^2+(H/J)^2} \\ -\frac{Hs/J}{s^2+(H/J)^2} & \frac{(H/J)^2}{s^2+(H/J)^2} \end{bmatrix} \quad (7)$$

At this time, there is

$$\begin{bmatrix} I_x(s) \\ I_y(s) \end{bmatrix} = -\frac{1}{s}\frac{K K_{PC}(s) K_P}{1+\frac{K_T K K_{PC}(s) K_P}{Hs^2}} \begin{bmatrix} 0 & \frac{1}{s} \\ -\frac{1}{s} & 0 \end{bmatrix} \begin{bmatrix} \dot{\varphi}_x(s) \\ \dot{\varphi}_y(s) \end{bmatrix} \quad (8)$$

From (8), the coupling component in the current feedback loop is eliminated, and the goal of output decoupling has achieved.

IV. DTG NEURAL NETWORK METHOD OF FULL DECOUPLING

It is known that the traditional decoupling method can not achieve the goal of control decoupling and output decoupling at the same time from the upper analysis. This article will design a neural network to make the gyroscope system to achieve the objective of full decoupling.

A. BP neural network selection and training

The reference[6] has pointed out that the two-hidden network is the best one among all the networks mentioned in it. The two-hidden network is the one that has two entrance neurons, and s1 neurons in the first hidden layer, note the corresponding activation function is f11; s2 neurons in second hidden layer, the corresponding activation function is f12. The output is O, Y is the target vector, select f11 and f12 as Sigmoid function; f2 as purelin function.

B. RBF neural network theory

RBF neural network has 3 layers, Fig 1 is the structure. There is no transfer function in input layer, which is just used to input. The transfer function in hidden unit the radial basis function and the function in output layer is linear function. The mapping relationship has two parts, the first one is the non-linear transfer layer from input space to hidden layer space. The output of J-th hidden unit is:

$$h_j(x) = \phi(\|x - c_j\|, \sigma_j) = \exp(-\frac{\|x - c_j\|^2}{2\sigma_j^2}) \quad (j = 1, \cdots, p) \quad (9)$$

Here, $\|a\|$ is the norm of a; x is the input vector, which has n dimensions,

$$x = [x_1, \quad x_2, \quad \cdots, \quad x_n]^T$$

c_j is the center vector of j-th non-linear transfer unit, and has the same dimension with the input vector;

$$c_j = [c_j^1, \quad c_j^2, \quad \cdots, \quad c_j^n]^T$$

c_j^k is the k-th input vector corresponding to the j-th center; σ_j is the width of j-th non-linear transfer unit.

Part II: the combination layer from the hidden layer space to the output layer space. The j-th output is:

$$y_j = \sum_{i=1}^{p} h_i w_{ij}, (1 \le j \le m) \quad (10)$$

In equation (10), w_{ij} is the connection weight of i-th hidden unit with the j-th output; m is the dimension of output; p is the number of hidden unit.

The learning proces including the learning of the units in hidden layer and one of the units in output layer. Nearest neighbor-clustering algorithm is a method of self-adapt clustering learning method, which needn't the number of hidden unit before learning, thus, the RBF network after learning is the optimum one.

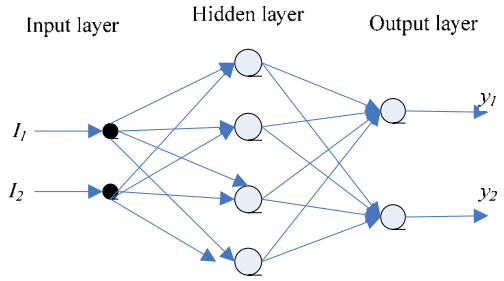


Fig 1 the structure of RBF network

C. The training process and the simulation result of neural network decoupling

The neural network is simulated by the eight locations experimental data(no decoupling). Treat the output of traditional decoupling method as the training data of the neural networks, there are 5000 groups of numbers.

Luo[6] has selected out the best structure of BP neural network in all the possible structure of the network to solve the decoupling problem. Here we will take the best on in reference [6] as the reference to verify the performance of RBF neural network. The network has two hidden layers, the first hidden layer has 3 nerve cells and the second layer has 2, after 7.78s training the max training error in x-axis is 0.000628 and 0.0080967 in y-axis. Correspondingly, the training time of RBF network is 6.7s, and the max training error of x and y axis is 0.000239 and 0.0084 respectively.

The trained neural network is tested by the eight locations experimental data, and the output curve are listed below from Fig 2 to Fig 4.

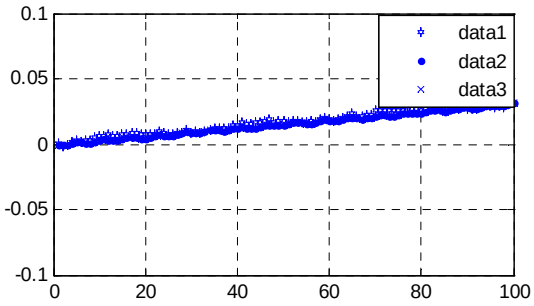


Fig 2. Comparison of X-axis output in Q1x input

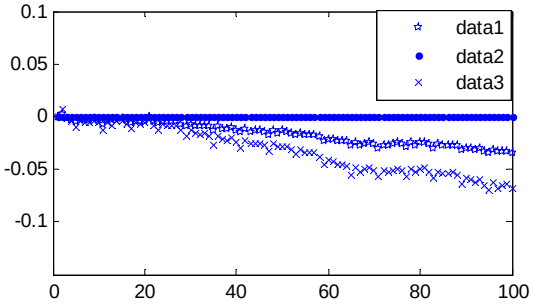


Fig 3. Comparison of Y-axis output in Q1x input

data1: data of RBF Neural Network decoupling
data2: data of traditional decoupling
data3: data of BP Neural Network decoupling

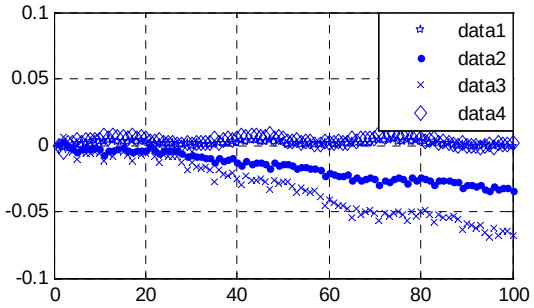


Fig 4. Errors of X and Y axies output in Q1x input

data1: error of X-axis output of RBF Neural Network
data2: error of Y-axis output of RBF Neural Network
data3: error of Y-axis output of BP Neural Network
data4: error of X-axis output of BP Neural Network

The simulation results listed above have proved that the RBF neural network can achieve the goal of DTG decoupling, and the network has much improvement over the BP network in accuracy. But the best result among the three is still the traditional decoupling method, so, there is still research work ahead.

V. CONCLUSION

DTG had a very serious phenomenon of coupling, which should be taken measures to decouple the X,Y axis

signals. Compared with BP neural network, the RBF network is faster, and also the accuracy is much higher. Both the two network can be realized by software, which is very simple and easy to realize. But some measures should take to further improve the accuracy of the neural network, such as training the neural network by much precise data collected from a more effective data processing.

REFERENCES

[1] Hao Ying. Study on Crucial Technologies of Dynamically Tuned Gyroscope, Harbin Engineering University, Harbin , 2006

[2] Ma Yunfeng. "Development of Gyroscope drift measurement and system", Systems Engineering and Electronics, Vol.23, No.2, 2001.

[3] Dai Shaozhong, Wang Bo. "Decoupling and robust control on servo loop of DTG", Missile and Space Vehcile, No.4 , Sum No.284, 2006.

[4] Huang Yexu, Shi Zhongke, Li Rong. "Optimization Design of Rebalance Loop for Rate System of Dynamical Tuned Gyroscope", In: *Industrial Technology, 2006. ICIT2006. IEEE International Conference* pp. 1829-1833, 2006.

[5] Fan Chunling, Jin Zhihua , Tian Weifeng. "A Hybrid Grey-based Model for Drift Signal of DTG", In: *IEEE 1nt. Conf. Neural Networks and Signal Processing* Nanjing, China, pp:1702-1705, 2003.

[6] Luo Yufeng, Xu Chao, Fan Yaozu. "Application of the Neural Network Theory in Fully Decoupling of DTG", In: *IEEE conference on Intelligent Networks and Intelligent Systems*, Wuhan, China, pp: 64 – 68, 2008

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15, August 2010, pp. 163-167

Digital Copyright Protection-Oriented EPD Electronic Teaching Materials Design and Implement

Yonghua Fu, Yong Liu
Zhengzhou Institute of Aeronautical Industry Management / Department of Information & Science, Zhengzhou,China
Email: { fuyonghua_12,y_liu}@zzia.edu.cn

***Abstract*—This paper analyzed the current development of electronic teaching materials, and probed into the restricting elements of the development: digital copyright protection, readability and multi-format support. Also, it reviewed the current readers' digital rights protection technology, and introduced the basic principles and properties of e-paper，and synthesized the advantages of the LCD screen and e-paper, designed and implemented the EPD electronic materials, which is an available way for digital copyright protection, described in detail the structure and implementation process of EPD electronic materials, including: protections facing the digital copyright, the migration and implementation of the operating system, multi-access mode and multi-format support, GUI implementation, EPD display control and power management. It has showed that the EPD electronic materials is highly operable and readable , strong in support of a variety of formats, content updated by wireless and portable, environmental-friendly and energy-saving, which has a good prospect.**

***Index Terms*—digital copyright protection; electronic paper; electronic materials; system structure; implementation flow**

I. THE PRESENT SITUATION AND BOTTLENECK OF ELECTRONIC MATERIALS' DEVELOPMENT

A. The present situation of electronic teaching materials

Traditional paper textbooks are expensive, bulky, unwieldy, not environmental-friendly, can not interact and update, only graphic display, and hard to learn by oneself, which can not fully meet the demands of teaching and learning requirements at any times and places in technological age and satisfy the party's target of "building learning society", which is stressed by the CPC Comrade Hu Jintao[1], the "to strengthen the Learning Society process "[2].

Electronic materials [3] can meet some of these needs, such as renewable, portable, learning anywhere and anytime. Currently, there are many domestic and foreign enterprises have invested into the research and sale of electronic teaching materials or books, mainly in three directions, but also represents the development trend and the level of technology:

*1) Content or software:*The "Electronic" mainly transformed from traditional teaching materials, which is multimedia courseware, or e-books. Such as domestic Superstar Group, which has made 3.7 million e-books, and e-Video publishing House of Higher Education has published 756 kinds of electronic publications since 1999. Foreign intelligence program company are mainly producing electronic version of the college textbooks , which had provided teaching materials around 6000, unified format for students to download and use of 180 days freely.

*2) Hardware:*Many companies mainly research on the reading hardware platform. Such as Hanwang Technology, Nankai Tianjin Branch, Foxit software company has launched its own brand of EPD electronic materials. Foreign company like Apple has developed ipad, which could support the color version of e-books.

*3) Mixed mode:*It not only develops the content, but also produces the reading hardware platform. Such as the domestic Hanwang invested a large sum to promote "electronic book" reader in whole country, the user can be randomly presented 500 genuine books, and built "Kingship Book City", which can provide users with nearly 20,000 genuine books for free download of sales approach. And in 2010, the sale-online profits of books are more than physical book for the first time, encouraging the development of e-book.

B. The bottleneck of electronic materials

However, with the thriving development, there are also many problems of electronic materials, mainly including digital copyright protection, readability, and multi-format support. Through the literature researching and first investigation, the current electronic teaching materials mainly depend on the LCD screen, which causes poor reading, inconvenient carrying, high-cost, high energy consumption, environmental damage, fewer formats supported, and lack of digital copyright protection.

*1) Digital copyright protection:*Compared to the paper materials, electronic materials copyright protection is a hard issue .Electronic materials must pass the digital rights management technology (DRM) to protect the content, in foreign countries, there are mainly Microsoft's DAS technology and Adobe's ACS technology ,while in domestic ,there are Founder Apabi technology, Superstar' PDG technology, and the scholar of SEP technology. Despite the technologies are numerous, but mostly are

This study is imbursed by Science and Technology projects of Henan province (No.092102210394):On Research & Realization of Digital Copyright Protection-Oriented EPD mobile reading terminal.

closed, and can not apply to other systems. Electronic materials need to face a number of publishers. it is necessary to find or design a mobile reading materials for electronic devices, which is easy to be implemented, with the opening of DRM and its key technology (including information security, data encryption, hardware fingerprint, digital watermarking techniques), and as an independent module embedded in electronic materials.

*2) Readability:*Good readability is the key of electronic teaching materials. Paper materials' biggest advantage is that we are used to it, therefore .So, patterning the "feel" of paper materials and achieving volume thin, flipping effect, are the existing electronic materials' bottlenecks- as electronic materials mainly display on LCD screen, although with the fine and smooth, it is not easy to carry, more power consumption, the effect under natural light is poor, the angle of reading limits, which is not suitable for "anywhere" learning.

*3) Multi-format support:*The type of many existing electronic teaching materials differs. There are text file, graphics and video files, sound files, video files, animation files. Even if the same type of documents, such as text files, also including ".txt", ".doc",".pdf", ". caj", etc. therefore, how to be compatible with a variety of content formats to the maximum is the key elements, also affecting the use of electronic materials.

II. E-LEARNING MATERIALS BASED ON EPD

According to the bottleneck and needs, the readability should be first considered. It introduces the basic principles and characteristics of EPD (Electronic Paper Display), designs electronic teaching materials based on EPD, and compares with the current widespread use electronic materials of LCD (Liquid Crystal Display).

A. E-paper's theory and characteristics

E-paper is generally implemented through electronic ink. Electronic ink is a kind of special materials processed into thin films and used with electronic display equipment; it is the comprehensive application of chemistry, physics and electronic technology. Electronic ink is composed of millions of tiny-sized microcapsules, whose diameter is as thin as hair. Each microcapsule contains white particles and black ones, respectively, with a positive charge and negative charge, they suspended in the clean liquid[3], See figure 1.

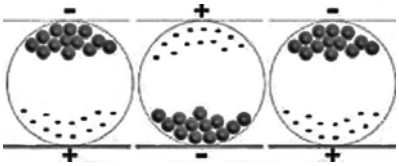


Figure 1. The Working condition of electronic ink

At the top of E-ink film is a layer of transparent material, which is used as the electrode of power, the bottom is another electrode of the electronic ink, microcapsules sandwiched between the two electrodes. When microcapsule is motivated by negative charge, white particles with positive charge move to the top of microcapsules, the corresponding position is shown in white.

Black particles with negative charge arrived at the bottom of micro-capsules due to the electric field force; the user can not see its black color. If the role of the electric field is in opposite directions, it shows the opposite effect, which is displayed in black and the white hide. So, as long as we change the direction of electric field, we are able to show the switch between black and white, white parts correspond to parts of the paper which is not written, and the black parts would correspond to a part which is written, See figure 2.

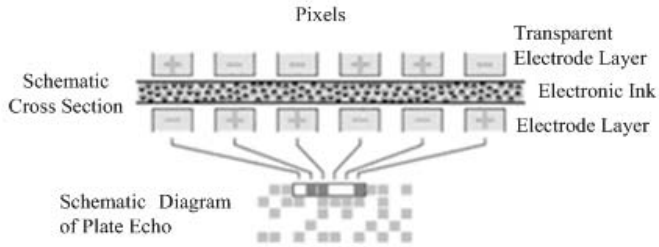


Figure 2. The displaying principle of electronic paper

This e-paper, composed by the e-ink, has the following characteristics: (1) it is rewritable. With EPD, we can rewrite and update the text or images, and browse for the large amount of information. (2) Suitable for the naked eyes. As a result of a higher contrast ratio, the text and the pictures are clear, and would never change in all directions, so you can read it in a comfortable way. (3) Portability. The film has light weight, easily portable character and can be appropriately folded and curled. (4) Even in the case of power failure, it can keep showing a long time, and runs on very little electricity.

However, the current e-paper has the following defects, compared with the LCD screen, e-paper has a lower Refresh rate, which is not suitable for complex, interactive and faster applications, it is more applicable to static display[4].

B. Compared with LCD

*1) Readability:*The reflectivity rate of EPD is 3 to 6 times that of the LCD, the contrast ratio is 2 times that of the LCD. EPD's much more readable than the liquid crystal display devices, and commensurate with newspaper. Therefore, E-paper can be treated as a display screen at the time of reading[5].

*2) Power consumption:*Compared with liquid crystal displays, EPD greatly reduced power consumption, which means extending battery life under the same conditions, or you can use smaller batteries, so enhance the system capacity of battery life[6][7].

*3) Availability:*E-paper displays have excellent Availability; it is not only thinner than liquid crystal display, but also light-weighted, durable and flexible. But it has a lower Refresh rate, and it is not helpful to human-computer interaction. LCD screen display refresh faster and has rich colors, and touch-screen can be used to achieve good human-computer interaction[5][6].

Comprehensively considering the respective advantages and disadvantages of the electronic paper and

liquid crystal display, The reader of this article runs the operating system GUI, control the human-computer interaction by the LCD screen, touch screen and keyboard, and determine whether the content displayed on LCD screen can be shown on the electronic paper through the buttons. At the same time, the LCD display can be shut down to save power consumption when in reading, the user can control the content in the electronic paper and turn page by pressing the page up-down keys, as the system structure shown in figure 3[8][9].

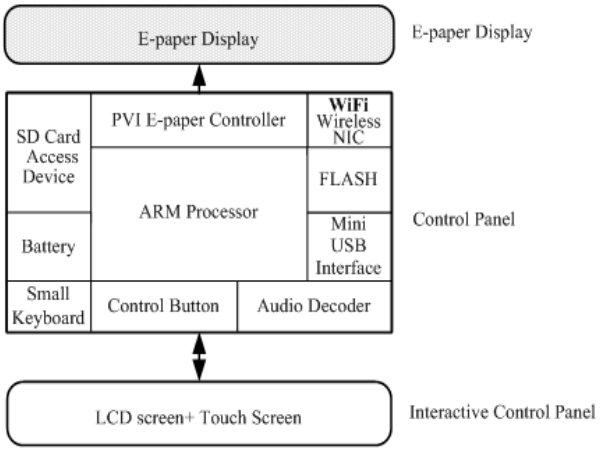


Figure 3. Electronic Readers system

In the electronic Readers system, the functions of each module are as follows.

*1) The modules of contents stores:*store e-books, documents are up to read in the SD card, connected to the Internet, or reading online through browser.

*2) Input Module:*The small keyboard keys, you can conduct various kinds of input; open the power with power key, page control while reading; window operations in graphical interface with touch screen.

*3) Output Module:*LCD screen and electronic paper. Graphic interface shown in the LCD display, mainly for the content of the search to find, while contents which needed to read displays in the electronic paper. When reading in the electronic paper can turn off the LCD screen, the system can still be in operation; through the control button we can display the contents in electronic paper directly.

4) Power modules: battery and USB power supply, as while as USB power supply, the battery charge[10].

. THE IMPLEMENTATION OF EPD ELECTRONOIC MATERIALS SYSTEM

A.DRM (digital copyright protection)

In the digital copyright protection, we should ensure that the content of each reading be legitimate. One important way is the equipment, through which to obtain and read the contents, should be legitimacy. Each device must apply to the Certification Center, which issues a unique certificate. Issued certificate is a key for compliant hardware or software. When the equipment has been signed by the certificate, you can pay to the content server and download freely and read the contents, at the same time of download, the content server communicates with the authentication center to confirm whether the device is a legitimate reading device. Each download are encrypted by the device certificate, so you can ensure that each book can be read on only one device, can not be copied. The module shown as figure 4.

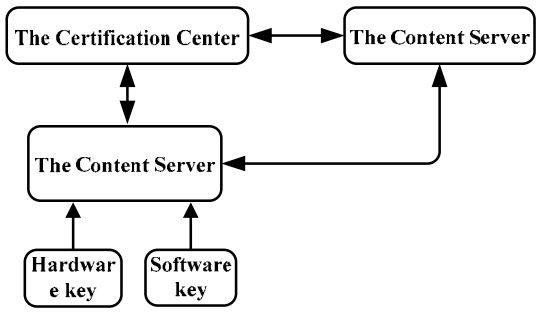


Figure 4. the module of digital copyright protection

There are two ways of reading content, online and offline reading. Online, content servers and devices can dynamically consult and change the key at specified times, guarantee not to be preserved and reproduced. Offline, in the case of disconnected networks can still guarantee read, each download content is not only encrypted by device certificate, but also ensured not be copied and altered.

B.Access to Multi- document Format

The current document formats are diverse. In order to simultaneously display documents in various formats, the properties of the document should be abstracted. The user interface should be unified. A document should include the following properties, such as author, publisher, publication time; text; content encoding format; content directory; text chapters and paragraphs; text attributes, including color, font size and format. After extracting the contents of these formats, the document layout can be carried out by the layout engine, displaying the content.

The core component is the virtual contents extracting interfaces (VCI, Virtual Content Interface). VCI provide access to the layout engine of obtaining document content and unified attributes interface. For each specific document, just through signing up for a format parser (Parser), the interfaces of registration standards contents can be analyzed and displayed. In this way, it ensured the unity of the operating interface and supported the various documents' extensibility. See figure 5.

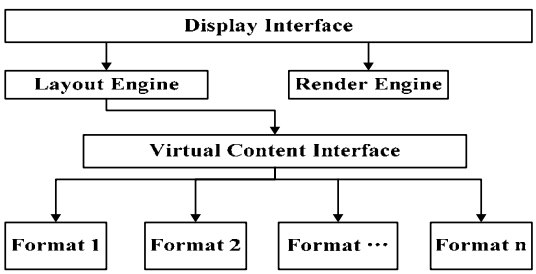


Figure 5. Accessing to multiple document formats

C.Operating system migration and implementation

The whole hand-held e-reader is a typical embedded devices, its system software consists of three parts: the Bootloader, Linux kernel and file system. The Bootloader introduces open source u-boot, according to hardware design, in the configuration file modified in various hardware-based address, the implementation of the compiler make the binary file, the Bootloader through JTAG programmer to the system in the FLASH. The Linux essence uses 2.6.25 editions, carries out make menuconfig to choose processor ARM in the disposition menu, and disposes each hardware module the actuation, make production compression essence document zImage. The filing system uses ext3 form Ramdisk, the application procedure (including the GUI master routine) places in Ramdisk[10][11].

As for the drive program, the majority of them may use the driver in Linux, work well after carrying on the few revision, most important electronic paper controller's actuation design. Electronic paper image's demonstration flow, see figure 6, the processor carries on the display control to the electronic paper controller routing directive, each picture's data size is fixed, regarding 800×600 the resolution display monitor, the data quantity is the 800*600*2/8 byte, each picture element is 4 gray scales (2 bits data) [7-8].According to the Linux actuation's frame, designs the electronic paper controller's actuation a character equipment, mainly realizes in the electronic paper controller's initialization and the driver struct the file_operations structure content, it is the application procedure and the hardware interactive connection[9].

```
struct file_operations epd_fops = {
.open = epd_open,
.close = epd_close,
.write = epd_write,
.ioctl = epd_ioctl,                    };
```

In the driver initialization is completed in the main control chip set of the register to request the allocation of space for data transmission. epd_open and epd_close are respectively responsible for opening and closing operation of the driver, epd_ioctl carries on a variety of control over the controller, such as reset, screen clearing. It is crucial to achieve epd_write, it must be implemented in strict accordance with the order of electronic display to achieve a direct impact on the refresh rate of electronic paper.

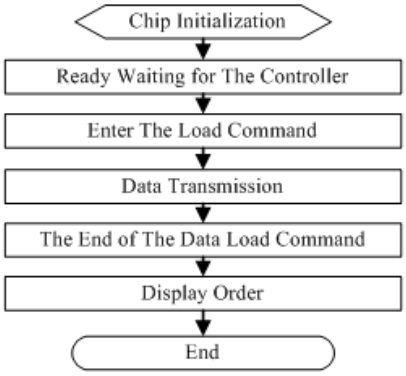
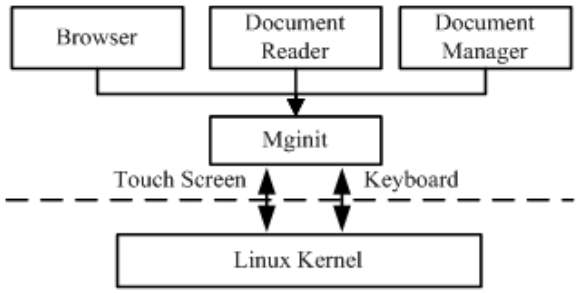


Figure 6. E-paper image display process

D.The realization of GUI

GUI introduces the open source MiniGUI. MiniGUI is based on customer/server structure embedded GUI, has a server advancement (mginit), it is responsible for the initialization of some input device, and transmit the input device's news to the onstage customer advancement through UNIX Domain sleeve joint character. The application procedure construction based on MiniGUI, see figure 7. Carries on the cross compiling after the MiniGUI source code produces the dynamic storehouse as well as configuration files MiniGUI.cfg, its copy to the

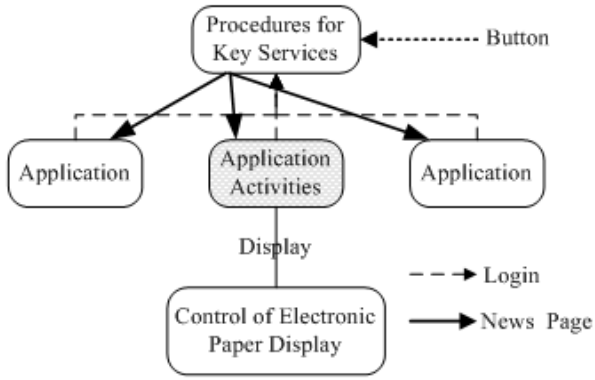


filing system, and carries on the correct disposition, guaranteed that the systems operation is normal[12][13].

Figure 7. Based on the application framework MiniGUI

E.EPD control

In the system, it is necessary to design a background service program for dealing with key events page. Each program requiring displaying the relative content on EPD must apply to the service program for registering the matter of turning the page event, then when the service program captures the key events page for all registered applications send key event broadcasting, and only in the current the activities of the application window to display their content sent to the EPD on the show. See figure



8[14].

Figure 8. Display Controls

F.Power Management

Battery-powered skill is used to operate the handheld embedded devices, so a good design of the power

management of handheld devices is a key technology. In this system, in addition to the normal shutdown and sleeping mode, the LCD screen is also power-consuming. In order to save energy and, if necessary, close the LCD screen. The system-defines four states:

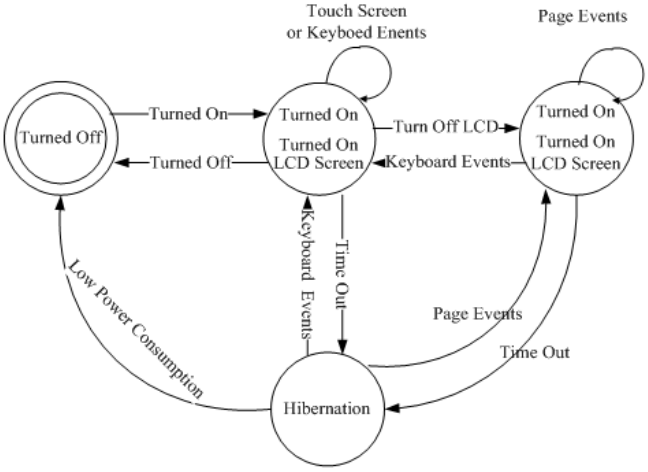
The state of Turning off: the processor enters the deep sleep state, only the RTC clock is running.

The state of Turning on: the processor is in normal operation, man-machine interaction, the LCD screen and the touchscreen are all running.

LCD screen turned off: the electronic paper is on for reading, the LCD screen is closed, but the system is still dealing with the normal operation of the events page.

Hibernation: the processor is move to hibernation mode, but maintain memory refresh, all the memory data is not lost. Wake-up keys will help quickly return to the pre-hibernation state.Four state of the conversion, see figure 9.

Figure 9. the four state transitions



Ⅳ.CONCLUSION

EPD is an energy-saving display device. From the perspective of digital copyright protection, this article combined with the advantages of LCD and EPD, researched on the hardware and software key technology, which supported the EPD's mobile e-learning materials, designed and implemented a specific type of hand-held EPD electronic textbook prototype. The electronic materials can provide open digital rights protection mechanisms, support a variety of common data formats. It is low power, lightweight, easy to carry, good readability, graphical interface to support high-speed wireless network to download, multi-channel content acquisition, control flexibility , easy to use, energy saving. This thesis content is becoming electronic information field, especially the hotspot of new display technology.

REFERENCES

[1] Learningcommunity.http://baike.baidu.com/view/3318.htm

[2] the Central Committee of the Communist Party of China on strengthening and improving Party building of the decision of a number of major issues under the new situation: http://www.gov.cn/jrzg/2009-09/27/content_1428158.htm

[3] LIU Ying, XU Yun-fei, On the Design and Development of Electronic Teaching Materials,China Educational Technology,pp 85-87, Feb. 2008.

[4] Makoto Omodani, “Electronic Paper:Concept and Expectations” , Chinese Journal of Scientific Instrument, Vol 25, No. z2, pp. 67-71, Aug. 2004.

[5] Minoru Koshimizu, and Xiaojin Zhang, “The Past,Present,and Future of Electronic Paper” , Advanced Display , No. 6, pp. 13-16, Jun. 2008.

[6] Huibo Xu, Xiaowei Niu, Xincheng Lu, and Yueming Sun, “The Research Progress on Microcapsulated Electrophoretic Ink”, Chemical Industry Times, Vol 22, No. 5, pp. 51-55, May. 2008.

[7] Chaoyang Zuo, Jianping Wang, Dengwu Wang, and Xiaopeng Zhao, “Recent Development and Applications of Microencapsulated Electrophoretic Ink” , Materials Review, Vol 20, No. 4, pp. 18-21, Apr. 2006.

[8] YongHua FU,The design and implement of a new Handle Electronic Reader.2009 international conference of management engineering and information technology, No. A, pp. 111-115,Aug. 2009

[9] Jonathan Corbet, Alessandro Rubini, And Greg Kroah-hartman, Linux Device Drivers, 3rd Edition, O'Reilly & Associates Inc, Sebastopol, Jul. 2005.

[10] Shah J, and Brown RM, “Towards EPDs made from microbial cellulose” , Applied Microbiology and Biotechnology, Vol 66, No. 5, pp. 352-355, May. 2005.

[11] Karim Yaghmou, And O'Reilly Taiwan company, Building Embedded Linux Systems, China Electric Power Press, Beijing, Apr. 2003.

[12] Philips.Apollo, “Electrophoretic display controller Device Specification” , China Academic Journal Electronic Publishing House, Vol 29, No. 4, pp. 952-978, Feb. 2008.

[13] ligong Zhou, ARM Embedded MiniGUI example of the initial development and application, Beijing University of Aeronautics and Astronautics Press, Beijing, Jan. 2005.

[14] Andrew Sloss, Dominic Symes, And Chris Wright, ARM System Developer’s Guide:Designing and Optimizing System Software, Elsevier, Amsterdam, Apr. 2004.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 168-170

Security Research of VPN Technology Based on MPLS

Chen Lin [1], Wang Guowei[2]
[1] School of Computer Science &Technology, Henan Polytechnic University, Jiaozuo, China
Email: chenlin@hpu.edu.cn
[2] School of Computer Science &Technology, Henan Polytechnic University, Jiaozuo, China
Email: wangguowei@hpu.edu.cn

***Abstract*—The VPN technology based on MPLS is the current mainstream VPN technology that uses isolations of routing and address or other information technologies to resist attacking and marking spoofing, in which the security of data transmission are guaranteed to a certain extent. In this paper, Securities of MPLS VPN are analyzed in three levels, in which the overall securities of MPLS VPN network are enhanced by deploying the appropriate security measures.**

***Index Terms*—MPLS; VPN; MPLS VPN; network security**

I. INTRODUCTION

With the development of computer network technology and internet, it is increasing that requirements of the flexibility and efficiency and security in network, virtual private network (VPN) technology was widespread concerned. The VPN technology Based on MPLS is the current mainstream VPN technology that uses isolations of routing and address or other information technologies to resist attacking and marking spoofing, in which the security of data transmission is guaranteed to a certain extent. But as an IP-based network technology, it is not solving illegal access of a protected network element and the error configuration as well as internal attacks and other security issues which widespread in the management of shared network.

II. AN OVERVIEW OF MPLS VPN TECHNOLOGY

Multi-Protocol Label Switching(MPLS) is a new network technology of booting high speed data transmission and exchange by utilizing fixed-length label in open communication network[1]. It is powerful to overcome packet forwarding technology limitations of the traditional IP for the performance characteristics of a perfect combination of flexible routing functionality in the network layer (Layer 3) and the high-speed switching data in link layer (Layer 2).The key of MPLS technology is Label concept which is short and easy-to-handle and only has local significance of information content. Label is short for easy-to-handle which is directly referenced by index. Local significance is designed for easy distribution. The value of MPLS is that connect modes are introduced into connectionless network.

MPLS VPN is a kind of VPN technology Based on MPLS of IP-VPN which is IP virtual private network for using the application of the MPLS technology and simplifying core router's routing on equipments of network routing and switching, the label swapping combining traditional routing technology[2].

Multilevel mesh network structure is generally used in MPLS VPN. It is consist of several different sites collection in VPN which a site can belong to different VPN and sites can be controlled for visits and isolation. MPLS VPN architecture is mainly divided into data and control plane. Data plane defines the VPN forwarding process; The control plane defines the establishment of the Label Switched Path (LSP) and routing information distribution process of the VPN[3]. Network framework of MPLS VPN is shown in the figure 1.

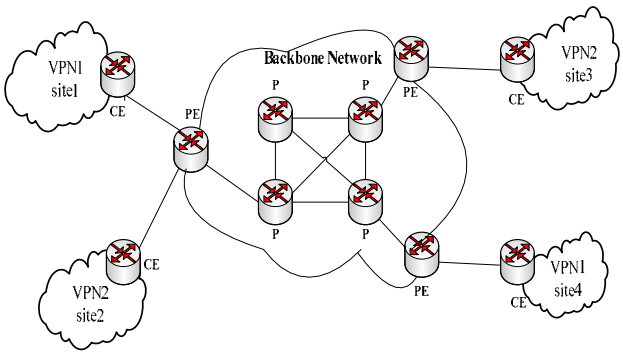


Figure 1. Network Framework of MPLS VPN

The architecture of MPLS VPN consists of three components: CE, PE and P[4].

Customer Edge (CE): the user interface. There are edge devices directly with the service provider network. CE can be either a router or a switch or a host. Typically, CE "perception" does not exist to VPN, also not need to support MPLS.

Provide Edge (PE): the service provider of edge router, which is directly connected with the user's CE. All managements of the VPN occur on PE, which the VPN routing information are maintained that is directly connected to, while all other VPN routing do not need to.

Provide (P): the backbone routers of service provider in the network not directly connected with CE, which have MPLS forwarding capabilities.

III. THE SECURITY OF MPLS VPN

The transfer and processing of information are divided into control, data, and manage three levels in

MPLS VPN network. In the control plane, the exchange and processing of routing information are completed and VPN routing tables are established and maintained. In the data plane, implementation of VPN data is fast forwarded. Configuration of the equipment is completed and appropriate management information is delivery in the management plane. The security threats of MPLS VPN network are also come from these three levels.

A. The security threats of the control plane

VPN routing information are exchanged through P/PE routers providing VPN services in the control plane, in which security attacks are against the two classes of devices. On the one hand, attacks are used on a routing protocol and illegal exchanging information of routers, on the other hand, there is an attack on a routing device which allows P/PE router does not work regularly.

- The attacking to VPN routing protocol

This kind of attack is typically found in members and routing information publication stage. For example, the attacker pretends as a PE equipment to establish a session with other equipments for the routing information. The internal routing information of VPN is disclosed. The attacker can also be forged or altered routing information so that the user data are passed to the wrong direction and user internal information are theft.

- The attacking to P/PE router

Usually P/PE router is attacked for means of squeezing the resources. For example, Denial of Service attacks affect and destroy the routing information to be sent properly, it interferes with the routing information for establishing and maintaining that impact VPN packets transmitted, the user businesses are affected ultimately.

B. The security threats of the data plane

- The security threats of internet

In the case of VPN users are connect with internet, attacks are launched with IP source address spoofing, TCP session hijacking and planting Trojans, in which the user data streams are viewed， modified and deleted non-authorized.

- The security threats of the shared device

In the MPLS VPN network, it is shared of network resources by normal VPN users, such as CE and PE equipment. In this case, although the VPN tunnel system can guarantee the security of information delivery to a certain extent, all security measures only will increase security threshold, the possibility of an attacker illegal capture, forgery and replay the possibility of MPLS label package can not rule out.

C. The security threats of the management plane

- The attacking to network devices through the administrative interface

An attacker accesses network management system remotely through the network access control management interface illegal gained by means of guessing. Configuration management information of the device is viewed, extracted and changed.

- Impact or damage management information delivery through clogging resource

If deliveries of resources are not through a specialized transport channel or the use of a "band" way of delivery, the management information does not pass normally for network resources are excessive extending by an attacker.

- The disclosure of internal information

The proper configuration can guarantee VPN routing information not leak for VPN is strictly isolated between address space and routing space. But as the number of VPN is increasing, a simple mistake of the network administrator may cause the VPN connection between staggered across the VPN to members of the error, which causing the routing information delivery and the leakage of error at cases of configuration management and more complex.

IV. THE IMPROVEMENTS OF MPLS VPN SECURITY

Although the MPLS VPN has the same level of security as ATM, FR virtual circuit for packets are sent in the MPLS domain in the form of label forwarding, it is not secure enough by MPLS technology itself. Therefore, the use of appropriate security measures to protect the MPLS VPN network security is very necessary for the threats of MPLS VPN in three levels. The design of MPLS VPN should be sure of routing information of the control plane are accurate, reliable and guaranteed, data delivery of the data plane are privacy, accuracy and integrity, configuration information of the management is secure.

A. Control plane safety

The safe measures of control plane are mainly guaranteeing the deliverable security of the routing information and isolation of routing.

The routing protocol neighbor certifications are most widely deployed. Neighbor certification allows receiving routing to use key to authenticate routing update source that only it and neighbor router know. The key of authentication between routers does not need transport with using MD5 authentication. The key and message are created into message digests as MD5 hash value to prevent the router receiving unauthorized updates from routing peers. this mechanism are also used to verify tag distribution peers receive updates.

The condition of PE equipments are Overburdened should be prevented for the abuse of routing information through strictly to limit the total number of routing information on the side of PE to CE.

The interface address On the CE site on PE VRF should be strictly prohibited while it is not needed.

These addresses are absolutely forbidden for CE site access in case of it is not required, such as the Loopback address of VPN Routing and Forwarding table (VRF).

B. Data plane safety

- CE-PE data encryption

the transmission path between CE and PE is relatively safe for multiple CE devices are connected into

the PE via Ethernet switches with Virtual Local Area Network(VLAN) which the transmission path is determined by the network administrator, the date are allow to access with non-encrypted way at the case of the consideration of the business costs and simplification of the configuration. If the way of access is wireless or remote, one of the encrypted access methods is necessary.

- PE-PE data encryption

In order to guarantee the security of data transmission, Internet Protocol Security (IPSec) is deployment to authenticate or encrypt the data flow between ingress to export[5]. The transmission of information between the PE is not encryption in general. The reasons are that it has a degree of security for the technology of MPLS VPN tunnels are used to transmit information; it is very complex of the implementation of encryption between PE and expensive of information delivery that heavy burdens of processing are bring to P/PE devices.

- CE-CE data encryption

IPSec tunnel is deployed to provide user data security in mutual communication between sites. This technology is deployed in the CE or between hosts requiring data protection in sites.

C. Management plane safety

- The access control of Network management system

The attack of hacker to network management system is primarily implemented through network management interfaces. In order to prevent the information of management thieving and malicious tampering, access authentication should be deployed at the administrative interface.

- The delivery channel of network management information

In order to prevent information of resource network management abnormal delivering for resource squeezed, management terminal should be used with out-of-band access management interface. The use of the link is isolated physically or logically with other infrastructure in VPN. If manage terminal is in-band access management interface, a filter or firewall must used to limit access to non-authorized users.

- The correctness of device configuration

Network administrators should guarantee the correctness of the VPN device configuration to prevent leakages of user data, which require improving the skillful level of administrator and increasing the moral quality of education at the same time.

V. CONCLUSIONS

As the sign of the network communication, MPLS VPN will gradually replace the traditional circuit communication and become future trends of network for the performance of its flexible, high-speed switching and routing and the high security.

REFERENCES

[1] Jian C, Chin L, “A restorable MPLS-based hose-model VPN network, ” Computer Networks, vol. 5l, pp. 4836-4848, 2007.

[2] Ayan B, “Generalized Multi-protocol label switching: An overview of signaling enhancements and recovery techniques,” IEEE Communications Magazine, vol. 39, pp. 144-151, 2001.

[3] Myoungju Y, Jongmin L, Tai-Won U, “A new mechanism for seamless mobility based on MPLS LSP in BCN,” IEICE Transactions on Communications, vol. 91, pp. 593-596, 2008.

[4] Rosen E, Rekhter Y, RFC 4364 BGP/MPLS IP Virtual Private Networks(VPNs)[S], IETF, 2006.

[5] Yang Yanyan, Martel Charles U, Fu Zhi, Wu Shyhtsun Felix, “IPSec/VPN security policy correctness and assurance,” Journal of High Speed Networks, vol. 15, pp. 275-289, 2006.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 171-174

Personal Spam Filter by Semi-supervised Learning

Zhang Shunli[1], Yin Qingshuang[2]
Department of Computer Science and Engineering, Henan University of Engineering, Zhengzhou, China
Email: zhang_slxx@126.com
Department of Computer Science and Engineering, Henan University of Engineering, Zhengzhou, China
Email: yinqsh@163.com

***Abstract*—An approach of personal spam filtering by semi-supervised learning is a proposal. As a semi-supervised classifier, a transductive support vector machine uses both labeled emails from available public sources and unlabeled emails in individual inbox as the input data. The problem of the generalizing the training data to the test data in traditional support vector machine is solved. It provides a way to combine the ability of generalization and adaptation for the spam categorization. The model and parameter selection is stated here in order to improve the performance of TSVM. The experiments show that the results of filtering with TSVM are better than those of SVM.**

***Index Terms*—semi-supervised learning, transductive support vector machine(TSVM), personal inbox filter, spam categorization, vector space model**

I. INTRODUCTION

With the development of the Internet, the volume of spam is increasing rapidly. The method of labeling spam artificially is a heavy task, so automatic spam categorization has become an important tool of network security [1]. The most common strategy at present is to build a server-based filter to search for spam, which is also called personal spam filtering. Because of privacy issues, training filters can not rely on the labeled emails in individual mailboxes, but should depend on available public training data. Generally speaking, spam categorization is a problem of special text categorization, and can be regarded as two kinds of categorization problem. The result of categorization is whether an email should be identified as a spam or not.

A support vector machine (SVM) is one of the more effective methods of machine learning. It has many advantages in text categorization, and is widely used in mail categorization [2, 3]. However, the traditional SVM is based on inductive reasoning, and its purpose is to look for the generalized rule for a given learning problem. While filtering personal email, it is very difficult to find a valid rule to classify the emails of individual users because of the various forms of personal email. Therefore, during the process of establishing classifiers, and in order to achieve the best categorization performance, we should not only consider the distribution of training samples, but also the distribution of the unlabeled emails in the mailbox. Semi-supervised learning can use both labeled and unlabeled data as the input data of classifiers, and has become more popular in recent years. Transductive support vector machines (TSVM) are an extension of support vector machine theory in the field of semi-supervised learning, and resolve the problem of distribution differences between the labeled data and the unlabeled data. Therefore, TSVM is a better way to solve the problem of individual spam categorization.

In this paper, I used TSVM to solve the problem of filtering spam in personal mailboxes, and analyzed the input model and parameter settings of TSVM in detail, and obtained a very good result in the experiment of email categorization.

II. INDIVIDUAL SPAM CATEGORIZATION

As a kind of special text categorization, spam categorization has the basic features of text categorization. They are as follows:

A. High-dimensional Input Space

There are over 10,000 different words in emails, which have a lot of the characteristics of learning and categorization. However, experiments show that almost all of these characteristics are related, which means that inappropriate methods of characteristic selection will lead to information loss.

B. Text Vectors are Sparse

Although many terms in emails are very large, the number of terms which appear in a specific emails are limited. Due to this situation, the value of most vectors features is 0. The length of the email is usually shorter than that of the text, so email vectors are sparse.

C. Most of the Text Categorization Problems are Linearly Separable

These features are the merits of SVM on the text categorization [4, 5]. They are also the reasons why SVM is widely used in text categorization.

Even though there is enough theoretical evidence to show that SVM can resolve the above problems quiet well, some difficulties still exist while using SVM to classify personal spam. By maximizing the categorization border of training samples, SVM studies the generalization rules of learning classifiers. Because the content of personal email varies, the available public email cannot

Corresponding author: Zhang Shunli

provide adequate information for categorization models. Therefore, the sample distribution of individual mailboxes must be considered in the design of the classifier. In addition, e-mail senders often use various methods to avoid anti-spam systems, which require classifiers to adapt to the changes. TSVM can access the special rules pertaining to the emails which need to be classified instead of the general rule of a whole input sample, so we can use the method of TSVM based on semi-supervised learning to resolve the above problems.

III. TRANSDUCTIVE SUPPORT VECTOR MACHINE

The basic principle of SVM is to maximize the categorization intervals of training data. TSVM was put forward by Vapnik originally. It is a kind of improved SVM. Considering both test and training samples at the same time, it searches for the maximum interval between them and marks the test samples while training. Compared with SVM , Its optimizations are extended as follows:

For a learning task $P(\vec{x}, y) = P(y \mid \vec{x})P(\vec{x})$, learning unit L is a function h in an assumptive space H: $X \to \{-1,1\}$, the sample sets of n training samples are S_{train} :

$$(\vec{x}_1, y_1), (\vec{x}_2, y_2), ..., (\vec{x}_n, y_n) \quad . \tag{1}$$

Each training sample is made of feature vectors. $y \in \{-1,1\}$. Different from inductive reasoning, $\vec{x} \in X$ and two types of Categorization tags, learning unit uses the set S_{test} of k test samples:

$$(\vec{x}_1^*, y_1^*), (\vec{x}_2^*, y_2^*), ..., (\vec{x}_k^*, y_k^*) . \tag{2}$$

In S_{test} , y_j^* is the tag which $\vec{x}_j^*$ needs to identify.

TSVM can be expressed as:

Solve $(y_1^*, ..., y_k^*, \vec{w}, b, \xi_1, ..., \xi_n, \xi_1^*, ..., \xi_k^*)$ to minimize

$$\frac{1}{2}\left\|\vec{w}\right\|^2 + C\sum_{i=0}^{n}\xi_i + C^*\sum_{j=0}^{k}\xi_j^*) , \tag{3}$$

and satisfy (4), (5), (6) and (7).

$$\forall_{i=1}^{n} : y_i\left[\vec{w}\cdot\vec{x}_i + b\right] \geq 1 - \xi_i \tag{4}$$

$$\forall_{j=1}^{k} : y_j^*\left[\vec{w}\cdot\vec{x}_j^* + b\right] \geq 1 - \xi_j^* \tag{5}$$

$$\forall_{i=1}^{n} : \xi_i > 0 \tag{6}$$

$$\forall_{j=1}^{k} : \xi_j^* > 0 \tag{7}$$

C and C^* are the user-specified parameters, and they can balance maximizing categorization interval and wrong-subdividing of training samples or exclude test samples.

TSVM can be seen as the optimization problem to directly access unmarked date tags. The specific training process can be found in reference [6]. Joachims applied TSVM to text categorization problems [6], the result of text categorization is improved significantly compared with that of SVM. Especially for small training sets, TSVM reduces the number of required training samples.

. PERSONAL SPAM CATEGORIZATION

In order to make TSVM get the best categorization performance, we must resolve some important problems, such as : the expression of TSVM input samples, the parameter choices of TSVM (such as the kernel function of C and TSVM) and so on. There are still many open problems here, from which we can obtain some solutions from the previous experimental results [6].

A the Expression of Characteristics

As a kind of text expression method, Vector Space Model (VSM) is currently used most widely in text categorization. For spam categorization, characteristic is the term which appears in the email. Therefore, we can express email as:

$$d_i = \{w_{i(1)}, w_{i(2)}, ..., w_{i(n)}\} . \tag{8}$$

$w_{i(j)}$ refer to the weight of the term j in the email i. j is the index of terms in the dictionary which is established by analyzing all emails. The value of $w_{i(j)}$ is usually counted through tf, tf-idf or Boolean. Because of the different lengths of emails and the informal structure of emails, or incomplete sentences, term frequency can't reflect the inner content easily. Experimental results show [6], that in the mail categorization the Boolean method is better than the others. In the expression of characteristic Boolean, the Boolean value is 1 if the term j appears in the text i, otherwise the Boolean value is 0 .

Whether or not to adopt the table of the stop word is also a factor in influencing the performance of text categorization. Different from many machine learning algorithms which limit the size of feature vectors, SVM uses all of the words without the need for characteristic selection. In addition, it doesn't identify words by the suffix, because the terms in the email message have different forms and are case sensitive.

In order to obtain a better performance from TSVM, the feature vectors should be normalized. It guarantees that all of the feature vectors are in an ultra-sphere. In the spam categorization, we can use the formula (9), $w_{i(j)}$ is the primitive feature vector expressed by the Boolean; $w_{i(j)}'$ is the normalized feature vector; n is the number of the components in the feature vector.

$$w_{i(j)}' = \frac{w_{i(j)}}{\sqrt{\sum_{k=1}^{n} w_{i(k)}^2}} \quad (9)$$

B. Model and Parameter Selection

Parameter selection in SVM is an open question: the choice of parameters depends on the training samples. Some methods of automatic model selection are based on estimating the border of generalization performance. Cross-validation is the most common way to estimate classifier generalization error. Cross-validation method can be found in Ref [6].

The choice of kernel function is a question of SVM. Some kernel function has been used for text categorization [7, 8]. Experimental results show that most of the text sets are linearly separable and linear kernel function is a simple and effective method. In the spam categorization algorithm, we adopt the inner product of data as the linear kernel function.

C. Individual Email Categorization Algorithm TSVM

Based on the above analysis, we propose the personal email categorization algorithm based TSVM be as follows:

1) Express emails as the Boolean vector space model without using characteristic selection and stop word. Use formula (9) to normalize feature matrix.

2) Use cross-validation to determine parameters C in model TSVM, and use linear kernel function.

3) Take labeled training samples and unlabeled samples as TSVM input. Train TSVM classifiers to get the maximum interval between training data and test data, while getting the tags of unlabeled emails in the inboxes.

D. Performance Metrics

Error rate is the most commonly used performance evaluation criteria in the two types of text categorization. For spam detection, we are more concerned about non-spam. Therefore, we adopt the rate of false reports and the rate of missing reports instead of the error rate. They are defined as follows [9]:

RFR refers to the rate of false reports; NFCSE refers to the number of false classification spam emails; TNSE refers to the total number of spam emails; RMR refers to the rate of missing reports; NFCRE refers to the number of false classification regular emails; TNRE refers to the total number of regular emails.

$$RFR = \frac{NFCSE}{TNSE}$$

$$RMR = \frac{NFCRE}{TNRE}$$

The value of AUC is another important evaluation criterion of categorization performance [10]. In categorization problems, TP is the ratio of the targeted samples assigned to targeted categories, FP is the ratio of non-targeted samples assigned to non-target categories. TP is also called sensitivity, and (1-FP) is also called specificity. ROC curve (Receiver Operating Characteristic) is formed by all possible values of TP as vertical coordinate and FP as horizontal coordinate. The value of AUC is the area under ROC. The larger AUC value indicates the better categorization result.

Ⅴ. PERSONAL EMAIL INBOX CATEGORIZATION EXPERIMENT

A. English E-mail Categorization

In the experiments of English email categorization, we adopt the personal mail in 2006 ECML-PKDD discovery challenge to filter data [11]. ECML-PKDD discovery challenge includes two tasks. Labeled training data are obtained from public data sources, while the unlabeled data in the personal mailboxes are taken as the test data. Task A deals with the situation of a large number of training data and many unlabeled available data that each user has. This task has three separate mailboxes. The number of training samples is 4,000, and the number of emails in a mailbox is 2,500. Task B has 15 separate mailboxes, but only a small amount of training data and unlabeled data are available. The number of labeled samples is 100, and the number of emails in each mailbox is 400. The experiment does not provide the original texts of the emails, and the emails are expressed only by adopting the properties of tf in vector space model.

An experiment can be conducted by using the method proposed in this paper. First, convert tf attributes into Boolean attributes, and normalize them according to the formula (9). The value of parameter C is obtained from the cross-validation in data adjustment, the value of C in task A is 0.05, and the value of C in task B is 0.2. As a comparison, another experiment can be conducted by using the traditional method of SVM and comparing their results. TSVM and the false and missing reports of TSVM and the values of AUC are shown in Table Ⅰ.

From the above results, we can see that: the rate of false and missing reports of TSVM is both less than those of SVM, and the values of AUC are larger than that of SVM. In addition, we also get very good results if the number of training samples are small.

B. the Mixed Chinese-English Mails Categorization

We adopt the emails in the mail server to construct the sets of experimental data. In this experiment, the total training data are 500, and all of them are from spam or non-spam, and are selected randomly from personal mailboxes. Test data are the emails in three personal mailboxes on the server, and the number of the emails in

TABLE I.
THE RESULT OF THE EXPERIMENT ECML-PKDD DISCOVERY CHALLENGE DATA SETS

	the Rate of False Reports		the Rate of Missing Reports		AUC	
	SVM	TSVM	SVM	TSVM	SVM	TSVM
TaskA	0.0277	0.0213	0.0421	0.0088	0.8320	0.9321
TaskB	0.7177	0.1917	0.5650	0.1200	0.8531	0.9006

each mailbox is between 100 and 500, and different from each other. Because the emails include both Chinese and English language, the construction of all feature spaces includes Chinese, as well as English terms. Through cross-validation, parameter C is valued 0.15. the rate of false reports, TSVM and SVM, the rate of missing reports, and the value of AUC are shown in table Ⅱ.

Ⅵ. CONCLUSION

Personal E-mail spam filtering is a challenge to the existing categorization method. Because the sample distribution of each individual mailbox is different, email filters need to be adaptable. In this paper, we used TSVM to classify personal spam. TSVM can access the special rules for each mailbox instead of the general rules for the entire sample space. Experimental results show that the classifier based TSVM improves the email categorization performance, and is better than traditional SVM in the handling and filtering of personal emails.

TABLE II.
THE RESULT OF DATA SETS MIXED WITH CHINESE AND ENGLISH

	the Rate of False Reports		the Rate of Missing Reports		AUC	
	SVM	TSVM	SVM	TSVM	SVM	TSVM
Inbox1	0.7700	0.2050	0.0650	0.0600	0.8401	0.9666
Inbox2	0.6400	0.2150	0.1850	0.0950	0.8892	0.8986
Inbox3	0.7300	0.2650	0.2750	0.3600	0.7558	0.7675

REFERENCES

[1] Wang Bin, Pan Wenfeng, Content-based Spam Filter Technology Review [J]. Chinese Information Journal, 2005, vol. 19(5), pp. 1-10.

[2] Youn, S., McLeod D, A Comparative Study for Email Categorization[C]. Proceedings of International Joint Conferences on Computer, Information, System Sciences, and Engineering (CISSE'06), Bridgeport CT, December 2006

[3] Drucker, H., Donghui Wu, Vapnik, V.N. Support Vector Machines for Spam Categorization [J]. IEEE Transactions on Neural Networks, 1999, vol. 10(5), pp. 1048-1054.

[4] Wang Qingxiang, Guan Kai, Pan Jingui. Email Filter Based on Support Vector Machines [J]. Computer Science, 2007, vol. 34(9), pp. 93-94.

[5] JOACHIMS, T, Text Categorization with Support Vector Machines: Learning with Many Relevant Features[R], LS8-Prport 23, Universität Dortmund, LS VIII-Report

[6] JOACHIMS, T, Learning to Classify Text Using Support Vector Machines: Methods, Theory, and Algorithms [M]. Kluwer, 2002.

[7] the Application in Spam Mails Filtering of SVM Based on a Variety of Kernel Function [J]. Computer Applications.2008, vol. 28(2), pp. 424-427.

[8] Ola Amayri, Nizar Bouguila. Improved Online Support Vector Machines Spam Filtering Using String Kernels[C]. Proceedings of the 14th Iberoamerican Conference on Pattern Recognition. Guadalajara, Jalisco, Mexico. 2009, pp. 621-628

[9] Liu Zhen, Yu Kun, Zhou Mingtian. The Spam Filter Technology Based on Multistage Property Sets [J]. Computer Application Research.2005, vol. 22(7), pp. 122-126.

[10] Tobias Scheffer. Email Answering Assistance by Semi-supervised Text Categorization [J]. Intelligent Data Analysis, 2004, vol. 8(5), pp.481-493.

[11] BICKEL, S. ECML-PKDD discovery challenge 2006[EB/OL]. [2010-1-29] http://www.ecmlpkdd2006.org/challenge.html

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 175-178

Organization and optimization of Web server

Tian Bing[1], Wang Jingjing[2]
[1] Henan Polytechnic University/Institute of computer science and technology, Jiaozuo, China
Email: tianbing2001@126.com
[2] NingGuo Central Primary School, Jiaozuo, China
Email: wangwjjing@126.com

***Abstract*— With the rise of Internet technology development,**
Browser Server mode has gradually replaced the traditional client-server mode. In the new architecture, the business logic of software application is implemented mainly by the Web server. System bottleneck will arises when the network traffic surge which cause greater pressure to database access. At this time, server will be slow or even to stop work. In this paper, these problems are analysis, and some useful methods are given which help the server more efficient.

***Index Terms*—Web server, System optimization, Server cluster, Loads balance**

I. INTRODUCTION

B/S (Browser/Server) mode generated from the rise of Internet technology, it improves the old C/S architecture. In the B/S mode, the application server fully realized the software application business logic. Processing of user requests is fully realized in the Web server and users can conduct business deal just by the browser. So this architecture is a new information system technology and has become the most popular way to set up the application software architecture.

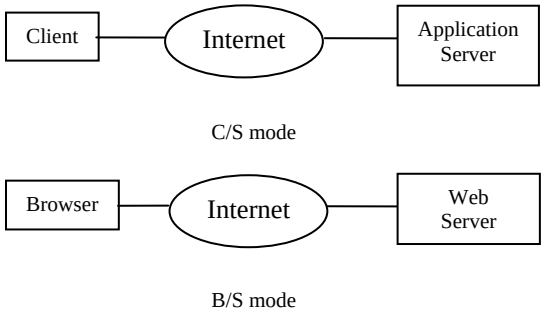


Figure 1. C / S mode and B /S mode

First, let's look at how the server works. Web server can resolve HTTP protocols. When the Web server receives an HTTP request, it will return an HTTP response which called an HTML page back. For example, to handle a request, Web server can respond to a static page or a picture for page jump. The server may also generate the dynamic response to some other procedures. Such as CGI scripts, JSP(Java Server Pages) scripts, ASP(Active Server Pages), server-side JavaScript, or some other server technology.

No matter what their purpose is, these the program of server-side usually produces a HTML response to the browser. You know, Web server, the delegation model is very simple. When a request is sent to the web server, the request just is passed to program which can handle the request process. Web server provides only an environment in which server-side program can be executed and the response generated by procedures returns, but not beyond the scope of functions. Server-side programs typically have transaction processing, database connectivity and messaging functions. Transactions and database connection pool are not supported by web server, but it can be configured a variety of strategies to achieve fault tolerance and scalability, such as load balancing, caching.

Next we look at an example. We know that an online website can provide real-time pricing and availability information. The site is likely to provide a form allows you to select products. When you submit a query, the website will process your query and return the results embedded in HTML pages. I want to introduce how a web server works and this will help you understand the functions of the server.

Let us look at an online store to find out how the server works. If a user visits the website of the online store, he will send a request to the web server. When the server receives the request, it will start the corresponding service process to handle the request. Once finished, the web server will formulate information to HTML form which the browser can receive. At last, the Web server will send this information to user's browser and the user also gets what he needs.

II. PROBLEMS ENCOUNTERED IN THE CURRENT SYSTEM

Web server must be able to achieve higher efficiency and stability which is the Web-based enterprise applications must have. High reliability can be considered as a redundant system configuration. If the application server can not handle a specific request sometime, there should be the other server which can quickly take over this job.

For an efficient system, if a Web server fails, other servers can immediately replace its place on the request for processing. The process should be transparent to the users and the user can be aware of anything.

The stability determines whether the application can support the growing user requests, it is the ability of the application itself. Stability is an effective measurement to evaluate how the other factors affect system

performance, such as the maximum number of users that the system can support, the processing time required for a request.

When the website's load is very large, there will be broadband network congestion and slow response to the problem site. How to solve these problems is of concern to each technician. For hardware, the maximum working load of a single server is limited. Performance of the server can not just rely on the improvement of the system hardware upgrades. Sometimes, the hardware investment is disproportionate and huge financial is not equivalent to performance boost. So we must find a solution.

We know that the bulge in the number of users will lead the server into an overload situation. Typically, the server will encounter two kinds of overload conditions. One is called short-time overload which is temporary. This situation is mainly caused by server load characteristics. A lot of research shows that web requests from the network traffic distribution is similar to that Web request traffic can be significantly large range of changes. This is often making the server in a short time overloading, but it just lasted a very short time. The other is called long-time overloaded which is usually caused by a specific event, such as the server is attacked by Denial of Service attack or a "live lock" phenomenon.

The first situation is inevitable indeed but the second situation can be eased by improving the server. If not considering malicious attacks, careful analysis to the server's process on information package can reveal that the unfair CPU usage by high-priority process is reason that system performances will degrade under overload conditions.

If all the loads were averagely assigned to different internal servers, the server load will be balanced to achieve the purpose of optimizing system performance. Typically, a group of servers in the cluster system will serve the same web application, but from the outside this seems like one server. Thus, when a large number of visits go into the server, the system will distribute them to different processing nodes. In fact, each server received only one part of the whole work and can easily finish it. So such a cluster server mode can achieve the objective of the optimization and decomposition.

III. HOW TO SOLVE THE PROBLEM

The method of diversion load has been mentioned. The problem is how to realize it. We must consider the real situation of web applications and features in optimizing web server. In addition to analyzing the characteristics of web load, we should also consider the environment of the web server. Server can not only take their normal work, but also maintain high throughput in the peak period. However, the server under high load performance is often below expectations.

In the solutions, the appropriate time of load diversion is important. In the existing server load balancing method, two methods are widely used and studied. One is the DNS (Domain Name Server) load balancing method RR-DNS (Round-Robin Domain Name System). In this method, host name will be mapped by the DNS to its IP address. When you type a URL in your browser, the browser will send the request to the DNS and requested return the corresponding IP address of the site, which is known as DNS queries. When the browser access to the site's IP address, it will use the IP address to connect to the site to be visited and display the page to the user. Shown in Figure 2, DNS server usually contains a single IP address with the IP address by mapping the name of the site list.

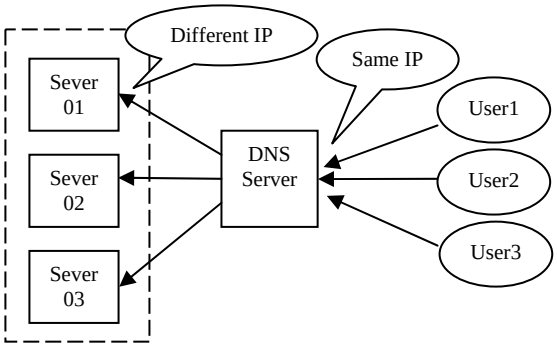


Figure 2. Round-Robin Domain Name System

In the DNS load balanced server, the same website has several different IP addresses in the DNS servers in fact. Each IP address represents different severs in the cluster, and all the servers are logically mapped to the same website name. Through examples, we can better understand this. We will publish three IP addresses to a cluster of three machines and DNS server contains the corresponding map table. When the first request reaches the DNS server, the IP address of the first machine is returned; when the second request arrives, the second machine's IP address is returned. And so, when the fourth request arrives, the first machine's IP address will be returned again, so on the cycle.

Using DNS Round Robin method, all requests will be equally distributed to the group of machines in a particular site. All nodes in the cluster share the task of network services. It is obvious that the possibility of occurrence of the bottleneck will be reduced.

The biggest advantages of DNS Round Robin is easy to implement and inexpensive. If you want to build such a system, you need not require additional auxiliary equipment. The hardware costs are very low and the whole system is easy to set up. To support the rotation schedule, the system administrator only needs to make some changes to DNS servers, and in many newer versions of the DNS server, this function has been added. For web applications, the code does not require any modifications. In fact, Web application itself will not aware of the load balancing configuration, even in front of it. Server performance scalability is also very easy. You can easily improve system responsiveness by adding a new server into the cluster. For the users, the building process is simple and advanced network knowledge does not be need. Even if there will be some problems later, all the maintenance is also very convenient.

This software-based DNS load balancing method has some main shortcomings. One is that it can not ensure

the consistency of the association service provided by the system. The consistency is an important ability which the load balancing server system should have. The system must determine the session information is on the server side or bottom of the database level to guide the user's request to the appropriate server. But you will find that this system does not have such intelligent features, because it is to carry out similar judgments through the cookie, hidden fields and URL rewriting. When a user establishes a connection with the server by the text-based approach, all of its follow-up visits must be connected to the same server. Now the problem arises. The server's IP is being temporarily stored in the browser cache in such system. Once the date is invalid, the connection has to be re-established. But the request of the same user may be assigned to a different server to process, and then all of the previous session information will be lost.

Another problem is that the system does not have high reliability to support high reliability. We know that a cluster has several nodes. If one node fails, then all requests assigned to the node will not be respond. Such situation is not that we would wish to see. So we must periodically check for damage nodes in the routers to avoid above situation. Once found, the damaged nodes will be removed from the list. But ISP (Internet Service Provider) would store the data of DNS in cache on the Internet in order to save time. In fact, update of DNS will be very slow and some users can not avoid visiting some sites that no longer exists, or lack of access to some new site. So the DNS Round-Robin resolved load balancing to some extent, the situation is not very optimistic.

IV. How to improve

We can overcome the shortcomings of the above by improving the load balancer. Virtual IP address is a good way to resolve many problems that the above method has.

The system using a virtual load balancer looks like a server with just a single IP address for visitors. Of course, this IP address is virtual address. Shown in Figure 3, it maps all the address of each machine in the cluster. When the request reaches the load balancer, it rewrites the request's header file, and assigned to one machine of the cluster.

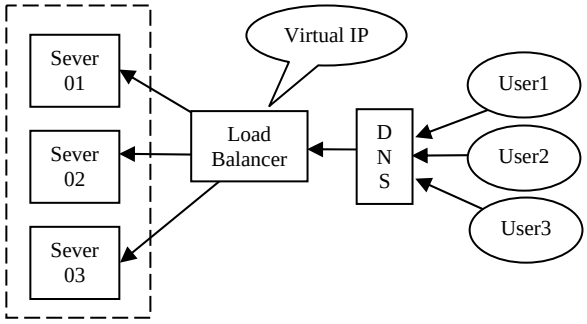


Figure 3. Load Balancer System

If a machine is removed from the cluster, the request will not be sent to the server that no longer exists, because all the machines have the same IP address in the face of it. Even if a node in the cluster is removed, and the address does not change. So the cached DNS on internet are no longer a problem. When a response returned, the client browser can only see the returned results from the load balancer. In other words, the client operation is targeted at the load balancer. For subsequent operations, it will be fully transparent to the client.

This method also ensures a server to be consistent. Load balancer reads the cookies or URL in each client request. Based on the information, the load balancer can rewrite the header and sent the request to the appropriate node which will maintains the corresponding session information requested by the client. Note that there is a problem. In normal HTTP communications, load balancers can provide consistency. But it can not provide the same service through a secure channel. For the load balancer is not allowed to deliver the information hidden in one session when the message is encrypted. So it is difficult to handle session information in the encrypted request. However, there are two ways to solve this problem.

First, a proxy server should be set up before the server cluster. It accepts all encrypted requests and decrypts them. Then the treated requests are re-sent to the appropriate node. This approach does not require additional hardware support, but it will add an additional burden on the proxy server.

Second, we can set up a Special hardware decoder. The requests will be decrypted by the decoder before they reaches the load balancer. In this way, processing speed is faster than the proxy server. But the cost is also high and the complexity of the system is obvious.

Since all of the web application requests must go through the load balancing system, the system can calculate many useful data, such as the number of active sessions, the number of active sessions, response times, peak load times, and so on. All of these statistics can be well used to adjust the whole system performance.

Virtual load balancer has some disadvantages, such as expensive cost and complexity system. For the load balancer is the single hardware that all the requests are via to pass, any failure of the load balancer would lead to the collapse of the entire system. This reduces the system's fault tolerance.

V. Conclusion

In B/S mode it is essential to maintain high-performance of web server. When the system has low efficiency caused by the bottleneck, it must take some measures to reduce system burden. Some ways mentioned in this article to solve these problems and these methods also some shortcomings. In fact, to a specific solution, the characteristics of the specific application environment should be taken into account and advantages and disadvantages of each method have to do a detailed analysis. Only in this way we can really solve these problems.

References

[1] Cai wei, Li Ming-shi. Web site traffic analysis system [J]. Computer Engineering and Science, 2006,28 (3) 26-27, 46.

[2] Lee-Song, room out. Web site traffic statistics based system design [J]. Computer Knowledge and Technology, 2008,1 (5) :875-877, 892.

[3] Steven Douglas Olson. Ajax on Java [M]. Nanjing: Southeast University Press, 2007.

[4] John Rodley. Web and Internet database development [M]. Beijing, China Machine Press, 1997

[5] Xu-Dong Li, Ke-Zheng Huang, Lin Hua. Based on Web, Java and CORBA implementation of the infrastructure of collaborative design [J]. Computer Engineering and Applications ,2001,9:87-89

[6] Kang Xiaojun, Shao Hong, Liu Jitao. Teaching load balancing technology to solve network bottlenecks. China Education Network ,2007-12

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 179-182

Task Scheduling of parallel programming systems using Ant Colony Optimization

Jun Mao
College of Computer Science and Technology
Henan Polytechnic University,
JiaoZuo 454000, China
morejune@gmail.com

***Abstract*— Efficient scheduling of tasks for an application is critical for achieving high performance in heterogeneous computing environment. The task scheduling has been shown to be NP complete in general case and also in several restricted cases. The paper introduces a novel framework for task scheduling problem based on Ant colony optimization (ACO). The performance of the algorithm is demonstrated by a Matlab program for producing effective schedules for random task sets .**

***Index Terms*—Ant Colony Optimization, Parallel programming, Task scheduling**

I. INTRODUCTION

A parallel program ,which offers an effective and promising alternative choice for high performance computing, is a collection of separate co-operating and communicating modules called tasks and processes. Tasks can be executed in sequence or at the same time on two or more processors. An efficient task scheduling avoids the situation that some processors are idle while others have multiple jobs queued up. The task scheduling activity determines the execution order. To meet the computational requirements of a larger number of current and emerging applications, a satisfactory algorithm for task matching and scheduling is able to enhanced the parallelization functions.

One of the key challenges of such heterogeneous systems is the scheduling strategy. Given an application modeled by a dependence graph, the scheduling problem deals with mapping each task of the application onto the available processors in order to minimize makespan.The task scheduling problem has been solved for years and is known to be NP complete [1]. Several heuristic algorithms are proposed in literature to solve this problem. These heuristics are classified into different categories such as list scheduling algorithms, clustering algorithms,but there have been being limitations. For example, the solution quality is not guaranteed for large sized problems.

Reliability of the heterogeneous systems has a vital role in scheduling the application on to the processors. As heterogeneous systems become larger and larger, the issue of reliability of such systems needs to be addressed. This problem can be prevented by a constructive algorithm based approach, called Ant colony Optimization (ACO).

Ant Colony System (ACO)

ACO is originally designed to find a better solution to Traveling Salesperson Problem (TSP), where a salesperson is going to visit each city exactly once so that to travel the minimum distance. Difficulty of solving TSP comes from how to find a better route (shortest route). Computationally, this problem is known as NP-hard thus it is hard to exactly find an optimal solution. Suppose we are standing on a present city and going to find a better trip of visiting a set of n cities. To this end, ACO algorithm decides the next based collective pheromone trails of ants and a probability p_0. If a uniformly generated random number p is greater than p_0 then the next city is decided by the highest probability:

$$\arg\max_{j\in N}[\eta_{ij}]^{\beta}[\tau_{ij}]^{\alpha} \tag{2.1}$$

where N is the set of unvisited cities, η_{ij} is given as the reciprocal of distance between city i and city j, τ_{ij} is the pheromone on the path between cities i and j. The parameters α,β are predetermined who are sometime sensitive to the convergence of an algorithm in practice. If p is less than p_0 then the next city is decided by the following probability:

$$p_{ij}=\frac{[\eta_{ij}]^{\beta}[\tau_{ij}]^{\alpha}}{\sum_{l\in N}[\eta_{il}]^{\beta}[\tau_{il}]^{\alpha}} \tag{2.2}$$

The pheromone on the path is updated after an ant moving to the next city j. The update rule is kept to as follows:

$$\tau_{ij}=(1-\rho)\tau_{ij}+\rho\tau_0, \tag{2.3}$$

where, τ_0 is a predetermined parameter, and ρ is an evaporative rate of pheromone. After a tour completed, ACO will globally update the pheromone on the path of current best tour by the following global update rule:

$$\tau_{ij}=(1-\rho_g)\tau_{ij}+\rho_g V, \tag{2.4}$$

where, V is the best value of objective function, ρ_g is an addition rate.

We summarize the ACO algorithm as follows:

Step 1 initialize all parameters used in the algorithm

Step 2 do the following steps until a tour is completed

2-1 select a city according to (2.1) and (2.2)

2-2 update the local pheromone according to (2.3)

Step 3 global pheromone update on the path of current best tour by (2.4)

Step 4 go to Step 2 until the stop criteria are satisfied.

II. METHODOLOGY

parallel programming systems is based on scheduling the tasks to be executed. Suppose that n tasks from m tasks terminals are submitted to work center. A parallel programming systems has a number s of process terminals to execute the tasks. Our task is to place each task to a right process terminal in order to maximize the outcome for the parallel programming system.The different priority level will also be given to task terminals refering to their past contribution to the system. In this research we do not focus on how the methods of measuring the contribution of a task terminal. Instead we suppose the priority level is given.

The main difficulty is how to build up the path and the route when one applies ACO to solve a certain problem. To express our approach we use the following variables to express the methodology in this research.

Sets of tasks from task terminals: $\{o_1,\ldots,o_m\}$

Prices for different process terminals: $\{t_1,\ldots,t_s\}$, per second (depends on performance ,customers' satisfactions ,response speed etc. of each process terminals)

Maximum available time for each process terminal: $\{u_1,\ldots,u_s\}$

Tasks index by different task terminals: $\{c_{11},\ldots,c_{1n_1};\ldots;c_{m1},\ldots c_{mn_m}\}$,

$n_1+\cdots+n_m=n$.

Priority level of task terminals: $\{e_1,\ldots,e_m\}$

Timetable of processing in time (sec.):

an $s\times m$ matrix (b_{ij}), $i\in\{1,\ldots,s\}, j\in\{1,\ldots,m\}$.

The revenue of a timetable is calculated by

$$f=\sum_i t_i(\sum_j b_{ij}w_{ij}), \qquad (3.1)$$

where, w_{ij} is a $0-1$ binary matrix. The elements of matrix (w_{ij}) are divided to n groups, each group corresponds to a process terminal. All w_{ij} in the k-th group take value 1 if all tasks c_{ki_k} of process terminal o_k are processed. Otherwise, All w_{ij} in the k-th group take value 0.

Now we apply ACO to find a global optimal timetable. We start with allocating an ant to an task c_{ki_k} based on the credit information on the advertiser by a probability

$$p_{c_i}=\frac{c_i}{\sum_{i\in N}c_i} \qquad (3.2)$$

A high priority level is given a higher priority to have its task to process. After a task determined according to the priority level information (3.2), a break is determined to process the task according to rules of (2.1) and (2.2). Then the above process is repeated until a timetable completed. The objective value f is obtained according to (3.1). This value is used to give pheromone η_{kj} on the path from k to j.

The reason for applying the ACO algorithm in this research is to exploit its excellent searching ability to maximize the objective f over all feasible timetables.

The complete algorithm is listed below:

Algorithm AcoMaxOutput

Step 1 form an initial pheromone by set $\eta_{ij}=e_i+t_j$ for c_{in_i} going to time slot j.

And $\tau_{ij}=1.$ Set all parameters .

Step 2 choose an task c_{kn_k} according to (3.2).

Uniformly generate an random value p.

If $p\geq p_0$, compute probabilities according to (2.2) and local update according to (2.3) to determine a break subject to u_i.

If $p<p_0$, compute (2.1) to find a time j to air c_{kn_k}.

Step 3 If a timetable is not completed, go to Step 2.

Step 4 compute f according to (3.1).

If f is currently best, set $V=f$. Global update according to (2.4) on the path of current best timetable.

If f is less that current best, set $V=1$. Global update according to (2.4) on the rest of the path of current best timetable.

Step 5 go to Step 2 and repeat several times.

We will see that the initial values $\eta_{ij}=e_i+t_j$ contribute to obtain a better f within initial iterations. At Step 2 c_{kn_k} is chosen by credit information of advertiser. This setting is very important to control no-show risk.

III. NUMERICAL EXPERIMENTS AND DISCUSSIONS

To investigate the performance of AcoMaxOutput, we conducted numerical experiment using a dataset of task times (in second) and price for different process terminals. Dataset

task times (in second) are showed in table2 according to second. We have 207 tasks in total. Price for different process terminal is is listed below in detail:

Table 1: Price value for different process terminal

[60 40 40 40 40 40 60 60 60 60 60 80 80 60 60 60 100 80 100 100 100 100 80 60]

The tasks come from twelve different tasks terminals. The detailed times are listed as follows:

Table 2: task time in second

Tasksterminal o_1 : 15,13,15,14,15,14,14,14,15,14,15,15,15,15,14,13
Tasksterminal o_2 : 14,16,15,15,15,15,14,14,16,14,17,15,15,16,14,14
Tasks terminal o_3 : 15,14,13,15,14,14,14,14,16,16,16,15,16,16,16,15,17
Tasksterminal o_4 : 17,15,15,14,16,15,15,15,14,15,15,17,15,14,15,15,15,16
Tasksterminal o_5 : 15,14,16,15,14,14,16,16,14,14,15,16,16,15,14,17,15,16,15,
Tasksterminal o_6 : 15,16,15,13,15,14,13,14,15,14,15,15,16,15,15,16,12,16,15,
Tasksterminal o_7 : 14,13,15,16,15,16,15,14,16,13,14,15,15,15,13,16,16,14,16
Tasksterminal o_8 : 16,16,17,15,14,17,14,15,15,14,14,15,15,15,16,16,16,14,15,
Tasksterminal o_9 : 16,16,14,18,15,15,16,14,14,15,15,15,15,15,15,16,16,16,15,
Tasksterminal o_{10} : 15,15,14,15,19,15,14,16,13,15,15,15,14,15,15,14
Tasksterminal o_{11} : 15,15,15,14,15,13,14,14,17,15,14,15,13,14,15,17
Tasksterminal o_{12} : 15,18,16,16,14,15,16,16,14,15

12 tasks terminals provide 219 tasks. The value of t_i in the algorithm is replaced by the data in this table.

We suppose that priority level of 12 tasks terminals as follows.

Table 3: Priority level of 12 tasks terminals

[450 300 200 450 400 280 400 450 200 400 450 200]
[1300 1450 200 400 450 200 400 450 260 400 50 50]

Credit information is measured by many methods. Variety of this dataset influences the revenue for the broadcasting station.

ACO algorithm works with many parameters, some of them are very sensitive to its convergence. Therefore, setting better parameters can be a key to make ACO algorithm work efficiently. After running AcoMaxOutput using the trail parameters with small-scale data, we eventually set the important parameters and other two initial data as follows:

Table 4: initial data setting

α	β	ρ	ρ_g
2	1	0.2	0.2
Maximum time limitation for each_break		total_number_ of_iterations	
50		500	

A wrong initial dataset may make the algorithm not to be convergent. Search ability and convergence are two main criteria. We found that our algorithm provides quite good results with the dataset in this table under the above two criteria.

A program based on AcoMaxOutput was coded using Matlab. The program was executed 10 times with the credit information described in Table 3. We obtained two sets of 10 better output in the following Table 5. The results in 2nd &5throw and 3rd &6throw correspond to the 2nd dataset and 3rd dataset in Table 3, respectively.

Table 5: 10-run results

# of run	1	2	3	4	5
Best Revenue 1	80420	80400	80280	80040	80120
Best Revenue 2	80080	80200	80180	80060	80020
# of run	6	7	8	9	10
Best Revenue 1	80160	80220	80360	80160	80400
Best Revenue 2	80060	80200	80060	80260	80030

The bset output we obtained each time is relatively stable.

One of the 10 runs are depicted in below Figure 1. We see that the algorithms starts with a very good Output value . We believe that this good result is benifited from the setting $\eta_{ij} = e_i + t_j$ in Step 1 of AcoMaxOutput. Actually, we found that the first Output value can be declined to less than 3000 while we set $\eta_{ij} = 1$ as the initial data.

When the credit information is changed slightly (see dataset in 3rd row in Table 3), we see that the best revenues in 3rd row (Best Output2) of the 10 runs in Table 5 were changed slightly as well. Figure 2 portrays the f value in each iteration of one of the 10 runs. We see the stability and higher search ability of AcoMaxOutput.

The numerical experiments show that the optimal output is influenced by the priority level of tasks terminals

IV. CONCLUSION AND FUTURE RESEARCH

We have studied how to apply the ACO algorithm to maximize the outcome for parallel program system . Different with TSP, there do not exist explicit paths and routes, which are needed to put the pheromone on and make ACO work well. The main contributions of this

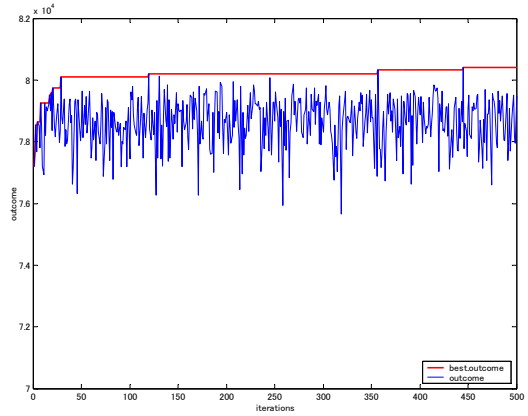


Figure 1: Output and iteratons

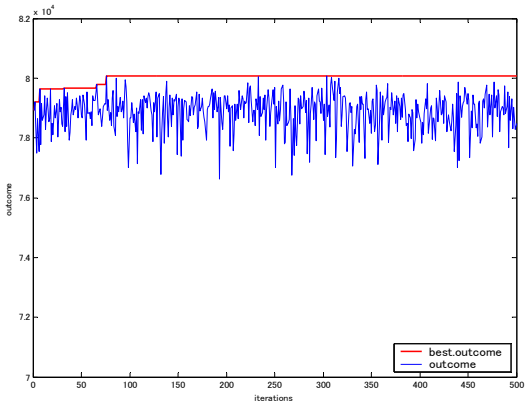


Figure 2: Output using a different dataset of priority level

research are:

(1) proposed a method to build a route in order to store the pheromone to maximize the output.

(2) based on above ideas, we proposed algorithm AcoMaxOutput to solve this revenue problem using credit information of advertiser.

(3) found a good initial data $\eta_{ij} = e_i + t_j$ in Step 1 of AcoMaxOutput, while the most other ACO-based algorithms use $\eta_{ij} = 1$ or a constant number suggested by, such as, Karpenko, Shi and Dai (2005).

The results show that AcoMaxOutput works efficiently and credit information can influence the optimal revenue. We know that substantial experiments should be conducted in this research field. There many topics we are interested in area such as how does the Priority level influent the oOutput and how deep it could be. We consider doing this research in near future.

ACKNOWLEDGMENT

The paper was completed by author when he was in J.M. Laboratory for Mathematical Optimization as a visiting researcher .The authors appreciates Professor Jianming Shi and the fellows in the Laboratory for providing advices and helps in this research.

REFERENCES

[1] Freund, R.F.Siegel, H.J. “Heterogeneous processing”, IEEEComput.Vol.26, no.6.pp.13-17, 1993.

[2] Della Croce, F. and Oliveri, D. (2006) ‘Scheduling the Italian Football League: an ILP-based approach’, *Computers & Operations Research*, Vol. 33, pp.1963–1974.

[3] Solimanpur, M., Vrat, P. and Shankar, R. (2004) ‘Ant colony optimization algorithm to the inter-cell layout problem in cellular manufacturing’, *European Journal of Operational Research*, Vol. 157, pp.592-606.

[4] M.Dorgio, V.Maniezzo & A.colorni “Ant System Optimization by a colony of Co operating agents”,IEEE Trans. Syst. Man Cybern. B, 1996, 26, (91), pp.29–41

[5] Dorigo, M. and Shützle, T. (2004) *Ant Colony Optimization*. Cambridge, MA: MIT Press.

[6] Chi Ho Tsang and Sam Kwong, “Multi-Agent Intrusion Detection System in Industrial Network using Ant Colony Clustering Approach and Unsupervised Feature Extraction”, Proceedings of the IEEE, pp. 51-56, 2005.

[7] C.W. Chiang, Y.C. Lee, C.N. Lee and T.Y. Chou “Ant colony optimization for task matching and scheduling”, IEE Proc.-Comput. Digit. Tech., Vol. 153, No.6,November 2006.

[8] JiQuan Shen, XueFeng, Zheng XuYan Tu. Humanoid Grid Management Model and its Implementation Frame.2006 International Symposium on Distributed Computing and Applications to Business,Engineering and Science.

[9] Dorigo, M. and Gambardella, L.M. (1997) ‘Ant colonies for the travelling salesman problem’, Biosystems, Vol. 43, pp.73-81.

[10] Karpenko, O., Shi, J. and Dai, Y. (2005) ‘Prediction of MHC class II binders using the ant colony search strategy’, Artificial Intelligence in Medicine, Vol. 35, pp.147-156

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 183-185

Research of Pervasive Computing

Li Gang [1] Chen Anfang[2] Yan Junhao [3]
[1]School of Computer Science and Information Engineering, An yang Institute of Technology, An yang, China
lzg0391@163.com
[2] Department of information science and engineering,Wanfang college of science & technology HPU
eric_chen2008@163.com
[3] Modern Educational Technology Center ,Henan Polytechnic University, Jiaozuo, China
yanjh@hpu.edu.cn

***Abstract*—Pervasive Computing has become a hot research field and human computer interaction in Intelligent Space is an important branch, fully reflect the characters of Pervasive Computing. In the scenes where various kinds of resource changes and information interaction occurs at anytime and anywhere, how to achieve computers serve human transparent in their daily lives is an important issue that need focused on. Having analyzed the characteristics of information interaction and studied the function that model required, a dynamic intelligent space four levels model was proposed in this paper. MPLS technology was used in the area of data information transmission to realize low time delay and low packet loss rate in the transfer process of data packets; A privacy protecting model was defined in the area of users' information protecting to realize privacy protection for users in their different identities.**

***Index Terms*—Pervasive Computing, Intelligent Space Model, MPLS Technology, Privacy Protection Model**

I. INTRODUCTION

The idea of Pervasive Computing was put forward by Mark Weiser [1] chief executive of Xerox in 1991, which combines wireless network technology, artificial intelligence technology, human-computer interaction technology, database technology, embedded technology, mobile communications technology [2], small computing devices operating systems and small Computing equipment manufacturing technology, etc. It is embedded into the daily environment or computer tools, which breached the limit of that users sit in front of the computer and made the computer itself in the disappearance of people's line of sight, at last, the center of people's attention return to the task itself completely. Pervasive Computing environment is the integration of information space and physical space, which can obtain digital service transparent anywhere and anytime.

Promoters of Pervasive Computing hope that computing can be embedding to the environment or day-to-day tools to make human feel it is more natural when they interact with computers [3]. And the notable goal of Pervasive Computing is to make computer equipments can sense the change of the surrounding environment, and automatically do the action what users need or they have already set. For instance, a user is meeting and his cell phone can incept the information of this meeting environment and automatically switch a quiet mode, and automatically answer calls, "master is in the meeting".

II. RESEARCH OF INTELLIGENT SPACE MODEL

A. Intelligent Space

The intelligent space based on Pervasive Computing reflects the integration of information space and physical space. All kinds of information equipments, technology and the environment space synchronizes together, which let people enjoy the convenience that all kinds of information bring to them without paying attention to technology or equipment itself, which has space refractivity, we can refer to an office or a classroom, and can also be a building, or can be large to a district, a city, in the areas of the Internet can also refer to the entire world [4].

B. Four levels space model

Through analyzing the interactive feature of intelligent environment information and researching the function what the model need implement, dynamic information space model to achieve different requirements of intelligent environment to information interactive was presented [5]. With the analysis of information units, the model provides information interaction mechanism, dividing into four levels, information environment oriented mechanism, equipment oriented mechanism, Judgment recognition oriented mechanism and ultimate user oriented mechanism. Using different levels of DIU (Dynamic Information Unit), the model can make intelligent environment changing in different scenes.

Given the definition of information unit, DIU= (Iscene, R， Ability， Context, SV). Iscene denotes a type of information, for example, the different form of the temperature information, etc. R denotes the service for Iscene, for example, mobile sound serve based on sound service, etc. Ability denotes the able of equipment, for example, projection equipments have display function and voice function, etc. Context denotes the environment of Iscene, for example, a quiet meeting environment or a driving environment with safety guarantee, etc. SV(Security Verify) denotes the verify mechanism which ensure the information transport in the model on the safe side and the privacy protection mechanism, for example, if the message in information transport aspect is not trusty after verified, the alarm will be sent to the equipment; in the privacy protection aspect, model protects users' privacy information and physical location.

The DIU of dynamic information space is divided into four levels, which are DIU1 (Iscene, R), DIU2 (Iscene, R, Ability), DIU3 (Iscene, R, Ability, SV)和 DIU4 (Iscene, R, Ability, Context, SV).Chart one reflects the relation of DIU:

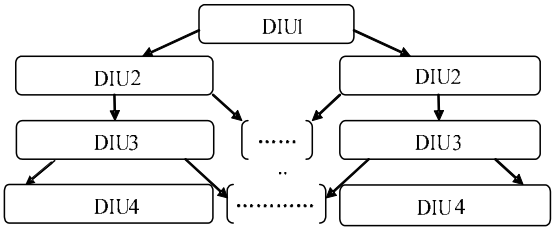


Figure 1. the relation of DIUs

C. *Using MPLS technology in the model*

MPLS (Multi-Protocol Label Switching) is grouping transmitted protocols based on fixed-length labels transmitted, combined with Mobile IP can improve grouping transmitted rate and service quality. In MPLS-based mobile Ipv6 network environment, designed low latency and low packet loss rate and with error recovery ability switch protocol is particularly important to ensure that information of the model is accurate, rapid transmission [6].

1. Transfer switch algorithm

While (MN demand for a switch) (

1) MN sends out the transfer request to CAR, requesting CAR to collect NI;

2) if (the transfer request of MN is the first time | | the interval of two requests > T) / / T is the time of routers update, in this new routing algorithm, generally every 10s update a time.

{CAR collects NI, updated forecast information table, and sends the information table to the MN;} {Else to 6}; / / MN, CAR and TAR need not repeat the work done by the previous.

3) MN chooses a TAR, and send it to CAR;

4) CAR sends the transfer request to TAR;

5) TAR finds CR nodes with the CAR, and establishes LSP between TAR and CR (Figure 2);

6) MN switch. During the MN is out from the CAR to connect to the TAR, the data packets for sending MN is intercepted by CR, who will send the packets to TAR;

7) if (MN switching success) (MN and TAR completes bundled update BU;)

{Else MN returns to the CAR, to Step 1}; / / switch failure

8) TAR answers and sends data packets to MN;

)

2. Analysis the process of transfer switch

In the mobile MPLS network, MN sent the transfer request to CAR and requested CAR to collect network information of other AR. CAR received the transfer request, through the step by step discovery mechanism to collect NI and establish forecast information table, and sent the information to MN. In according to their own service requirements, MN chose a suitable transfer switch node as the goal of access routers TAR from the forecast information table, based on TAR which MN has chose, CAR sent out transfer request to it, which told TAR a mobile node will be here. TAR reverted to CAR firstly after it accepted the request, secondly used the algorithm similar to the literature 7 to find the CR nodes of CAR and TAR , thirdly established the LSP between TAR and CR, then waited the switch of MN.

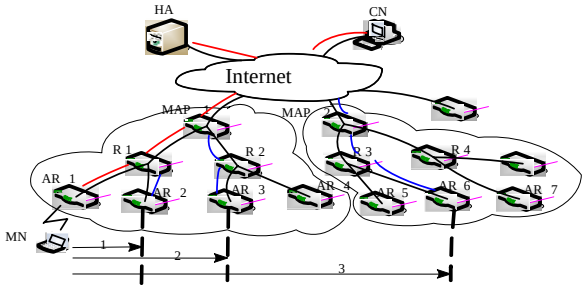


Figure 2. designed network topology

D. *Privacy protect model*

Pervasive Computing environment has dynamic specialty which make it difficult to protect privacy. In this environment the operation must be truly invisible and ubiquitous, it is crucial to users for their behalf. In order to achieve this, the environment should possess enough information of users in every situation where they may be involved. But the pervasive of the system depends largely on gaining and sharing of users' personal information with other entities, sometimes without users' directly consent.[7]

In the Pervasive network environment, users can get enough information from local information equipment in different places where they are, for example, when Mr. M got airport hall, his PDA contacted information server and got the information Mr. M need. In order to guarantee the privacy information of Mr. M and what he inquired not be known by non-authority users, privacy protect model should be embedded in the third level of space model.

1. Proposed new model of privacy protect

Considering the dynamic change of each factor in Pervasive environment, there are too much differences to effect the result of privacy protect, which include different levels and mechanism of privacy protect for people in different place; different restricting degree of social, law and moral criterion in different environment; different information equipments and the information they have, etc.

After synthetically considering the method of accessing control and the factors effect privacy protect, we put forward privacy protect model based on priority level, (Preferred Privacy Level Protect Model)

PPLPM=User ((L,M,S),T,C,UI,IV,IR,IU)

L stands for Law; M stands for Market Moral Norms; S stands for Social Surroundings; T stands for Technology Method; C stands for a set of Contextual Variables; UI stands of User Information; IV stands for Information Veracity; IR stands for Information Receiver; IU stands for the type of Information Use. These parameters are

interrelated and interdependent, through their different matching, the dynamic privacy protection mechanisms can be implemented in real time environment, which provide privacy protection for users in different environments and different time.

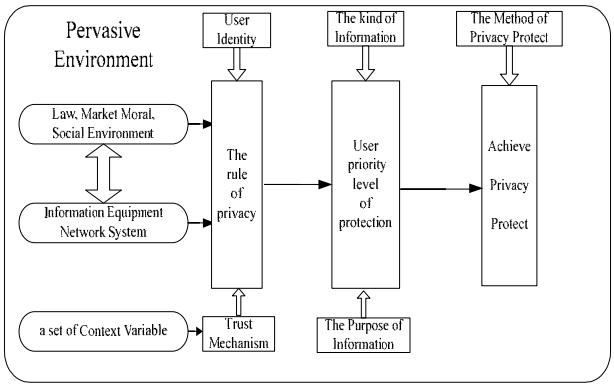


Figure3. Privacy Protection Model

2 Privacy protect method

1). Identity management protocol

In the identity management equipments of users, different identity subsets of users are allowed to show depending on context relation. In Pervasive environment, the equipments should be interconnected with other equipment for interaction in the operating environment, so they can't be restrict to contact with one user or some users. In this identity management based on context guide, users can be aware of the occurrence of the operation which is not practised by them and do the corresponding control to achieve privacy protection.

2). Maze protocol

Maze is a protocol aiming at ensuring the privacy of users and the anonymity of their communications. Through this protocol users communicate with each other and access resources through appropriate authentication technique, at the same time preventing the disclosure of their physical locations. Maze's operation is based on it's portals、routers and circuits. Maze routers are deployed in a hierarchical fashion. Leaf level portal is the beginning of the circuit, which can connect the user directly, but it can't distinguish the true identity of users. When users do operation, they connect leaf level portal firstly, then through circuit route a "storage center" can be get, which is actually a storage of the user's information routers, in where the identities of users can be distinguished, and ultimately realizing users the operation. With this protocol, leaf level portals have real physical address unknown users' identities, and storage center can identify user without known the full address of users, getting the result of privacy protection.

E. summarizes

This paper defined a four levels model of intelligent space in Pervasive Computing environment. The transmission of data and other information used MPLS-based mobile Ipv6 technology in the model, makes the transfer process of data packets achieve a effect of low latency and packet loss rate, makes the transmission of information rapidly and accurately. Privacy protect model is embedded in intelligent space model to make sure the transmission of information safe and accurate and users' privacy can be protected. This model not only can be used in small-scale intelligent space, such as: the smart classrooms and smart conference rooms, but also can be applied in a wider room and domain in Pervasive Environment.

REFERENCES

[1] Weiser M. The computer for the twenty first century. Scientific American, 1991,265 (3): 94 ~ 104. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] Gu Hong-liang, Shi Yuan-chun, Xu Guang-you. Proactive location-awareness service model for smart space. Journal of Tsinghua University (Natural Science) .2006 No. 46 Vol. Four ,35-40.

[3] Yan Xin-Ying, Fu Meng-yin, Liu Wei. Embedded device fault detection and diagnosis of system design. Micro-computer information, 2006,12 (2).

[4] Shi Yuan-chun Smart Space: a harmonious environment HCI Computer World. [J] Journal 2006.9 36 B13-15

[5] Wang Hai-peng, Zhou Xing-she, Zhang Tao. Environment for intelligent information on the activities of space model. Computer science, 2005,32 (12) :72-74.

[6] E. Rosen, A. Viswanathan, R. Callon. Multiprotocal Label Switching Architecture [S]. rfc3031, 2001.

[7] Muhtadi, J.Al., Campbell, R., 2002. Routing through the mist: privacy preserving communication in ubiquitous computing environments. In: Proceedings of ICDCS 2002, Austria.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 186-189

Design of remote automatic meter reading system based on ZigBee and GPRS

Li Quan-Xi[1], Li Gang[2]
[1] School of Computer Science & Technology Henan Polytechnic University, Jiaozuo, China
[1] Provincial opening laboratory for Control Engineering key disciplines
Email: lqx427@163.com
[2] School of Computer Science & Technology Henan Polytechnic University, Jiaozuo, China
Email: lg6080258@126.com

***Abstract*—Because of the traditional way of metering error and low efficiency, we propose household metering system design based on Zigbee and GPRS technologies, using PIC18LF4620 as the core processor and CC2430 chip as close communication function, using SIM300 chip as communication function in distance. Clustering structure of the network to reduce data traffic, Energy-saving sleep cycle has been achieved. Experiments have proved that this system can be safely.**

***Index Terms*—Zigbee; GPRS; Meter reading system; structure of Clusters; CC2430**

I. INTRODUCTION

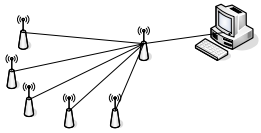
With the development of the computer, wireless communications and the rapid development of microelectronics technology, the people's life standard is constantly enhancing. And the demand of the home automation, building automation is also increasing. For households at the top of high buildings and luxury housing plot, the traditional meter reading has been unable to meet the future residential development needs, traditional metering not only waste labor human power, but also exit man-made meter error. If there is always no body at home, and charging are even more difficult. Smart increasing demands for remote meter reading. Using remote, wireless meter reading system can avoid manual meter reading mistakes, and errors of leakage of metering reading. It can improve efficiency, reduce labor intensity, and liberate labor, force. To meet demand, this paper will make the use of the ZigBee and GPRS technology to design a system that automatically copied in distance.

II. ZIGBEE AND GPRS TECHNOLOGIES

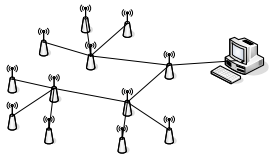
A. ZIGBEE TECHNOLOGY

ZigBee technology is emerging following the Bluetooth. It is short-range, low power, low cost and low complexity of wireless communications technology. The technology is applies value in the home automation, building automation, industrial control and industrial areas of logistics. ZigBee uses FM technology and spread spectrum technology to work in the 2.4GHz (global epidemic), 868MHz (Europe, popular) and 915MHz (U.S. pop), and in these three bands can transit high data rapidly with 250kbps, 20kbps and 40kbps. When using the 2.4GHz band, ZigBee technology can transmit 10 meters in the indoor, while in the outdoor transmission distance can reach 200 meters; in other uses spectrum, the indoor distance is 30 meters, while in the outdoor transmission distance can reach 1000 meters. The actual distance will be based on the size of the transmission power.

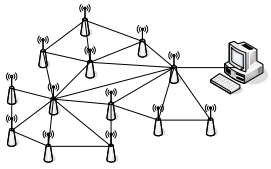
ZigBee technology has a variety of network topologies shown in Figure 1[1]



(a) STAR



(b) CLUSTER



(c) MESH

Figure 1. The third network of topology structure of ZigBee technology

Compared with other networks, ZigBee has the following advantages: low power, low cost, short time delay, network large capacity, reliability and safety.

B. GPRS technology

GPRS, is a short form of General Packet Radio Service. It is the European Telecommunications Association (the GSM system), It is technological innovation about exchanging and transmitting in groups. GPRS shares wireless channel, using IP to PPP to achieve data terminals in high speed and in distance. As the existing of GSM network to the technology of the third generation of mobile communication (2.5G), GPRS, has a significant advantage in many ways. It provides end to

end, wide-area wireless IP connection. It has the characteristics of using existing net works, using resources, always online, high transfer rate, reasonable cost and so on.

III. THE DESIGN OF SYSTEM STRUCTURE AND NETWORK

A. THE MAIN STRUCTURE OF SYSTEM, FUNCTION AND TECHNICAL SPECIFICATIONS

The system of home meter reading is composed of control terminal in distance, GPRS module and user metering module. Shown in Figure 2.

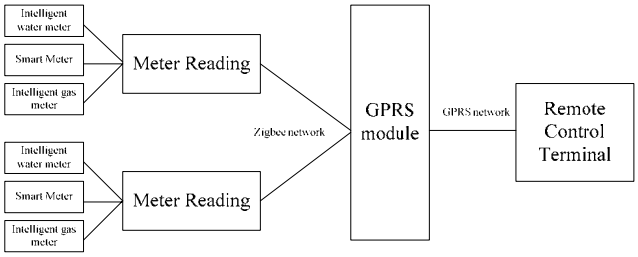


Figure 2. The whole diagram of the system

User meter reading module uses 2.4G free of charge, without application band. By the ZigBee network coordinator, the center of the network node, full-function device (FFD), a reduced functionality device (RFD) to form ZigBee wireless networks[1], to achieve the instrument of data copying and transmission and other functions.

ZigBee network is for short distance communication, and GPRS is networks for remote communication. they can combine with each other, through the net work's gateway to achieve a user meter copying to control and transfer in distance.

Home Meter Reading system should have the following functions:

- Timing, location, automatically copy the user's electricity, water, gas volume.
- Low-power, low cost, reliable, safe.
- To user's meters of electric water and gas working condition monitoring in time. When each of them work abnormally it can report automatically.
- To provide effective operating parameters for the property sector and offer the basic data for the implementation of automatically meter copying.
- To realize the transmission and processing of data in distance.

B. The network structure design

Using the combination of ZigBee and GPRS network, to copy home meters. Show in Figure 3.

The function of communication in short distance uses the CC2430 chip by chipcon company. A simple external circuit can constitute a data transceiver module. Remote communications of GPRS module uses SIM300 chip, its stability is relatively high. In the process of communication, the system goes through the RS232 serial port to send commands and data to SIM300, then SIM300 begin landing GPRS Gateway to get IP address successfully, after that starts to communicate with the remote control terminal, thus establishing a communication link with the Internet.

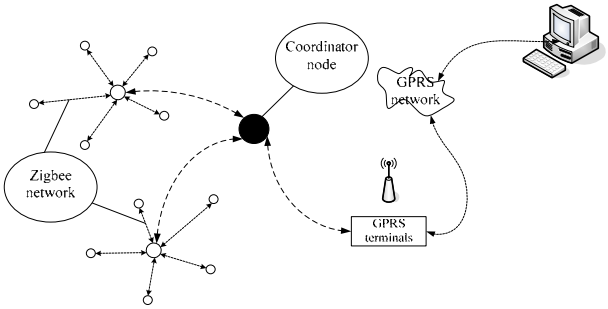


Figure 3. Network Structure

As the close communication within the node in the same region, data transmitting has great redundancy, so taking clustering structure of the network, and distributing centralized nodes to form clusters and elect cluster head. cluster node, among the data distribute to cluster head. cluster The first collection of cluster nodes within the cluster data fusion of information to complete work on packet compression, reducing the flow of data to achieve the purpose of energy saving[2].

IV. THE DESIGN OF HARDWARE

A. THE DESIGN OF USER METER READING MODULE DESIGN

User meter reading module consists of three parts, intelligent instrument data acquisition module, data storage and data transfer module of ZigBee.

ZigBee module uses CC2430 chip. CC2430 accords with IEEE802.15.4 standard, and it is special chip in ZigBee. In addition to the CC2430 including RF transceiver, it also incorporates enhanced flash memory 8051MCU, 32 / 64 / 128KB, and it watchdog of 8KB of RAM, and the ADC, DMA, etc.. CC2430 can operate at 2.4GHz frequency band, using low voltage (2.0 ~ 3.6V) power supply and low power consumption (when receiving data 27mA, send data 25mA), its sensitivity reach to-91dBm, the maximum output is +0.6dBm, the maximum transmission speed is 250kbis/s[3]. CC2430 module data transmission path and the process show in Figure 4.

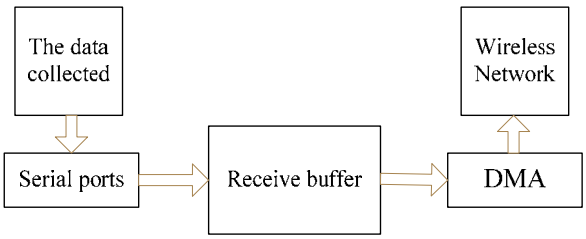


Figure 4. CC2430 module data transmission path and the process

Data collection needs to store the received data, and then select the path to send data out. It needs adequate storage unit. Select PIC18LF4620 MCU as the core processor. In the state of idle and sleep, you can make the system power consumption to a minimum. Interface is simple and little devices, simplifying the hardware

debugging more difficult, increasing the stability of the system. The interface circuit of PIC18LF4620 and CC2430 show in Figure 5[4].

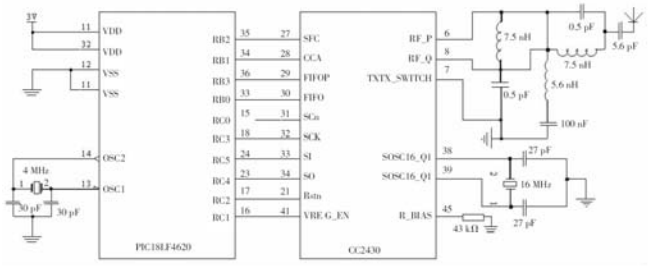


Figure 5. The interface circuit of PIC18LF4620 and CC2430

B. The interface of GPRS data transmitting module and ZigBee network

SIM300 is used in the communications of GPRS data transmitting chip module, SIM300 has TCP / IP protocol stack. It can facilitate the achievement of Internet access. Using the RS-232 serial communication interface to communicate with the chip, to support EGSM 900M, DCS 1800M and PCS 1900M 3 total bands, to be compatible with GSM Phase 2 / 2 +. It has small size, low power consumption, can provide voice, data and messaging functions [5]. GPRS network and ZigBee network hardware connections shown in Figure 6.

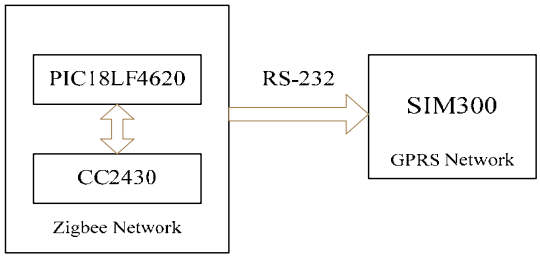


Figure 6. The diagram of ZigBee network and the GPRS network interface connection

V. SOFTWARE DESIGN

About software design, every communication protocol layer has to save energy. The communication between sensor nodes and the network coordinator introduce an example of communication process among the ZigBee modules. Before ZigBee modules making communications, it initializes effectively, ZigBee sensor nodes and the initialization process of the coordinator shown in Figure 7. In the initialization process, the network coordinator issues a request to connect sensor nodes. After the sensor nodes successfully receive and verify a data frame, and MAC command frame, the node returns to the confirmed frame. The ZigBee module of the sensor node is in dormancy. After finishing initialization, ZigBee module information processing as shown in Figure 8, the network coordinator is in working state, waiting for the response of the connection request of sensor nodes, and when the fixed time is up, sensor nodes take the initiative request to connect the network coordinator, and reported to the network coordinator about the collected data of intelligent instrument. Sensor nodes and link nodes, and the communication between the link nodes and network coordinating is like this[6].

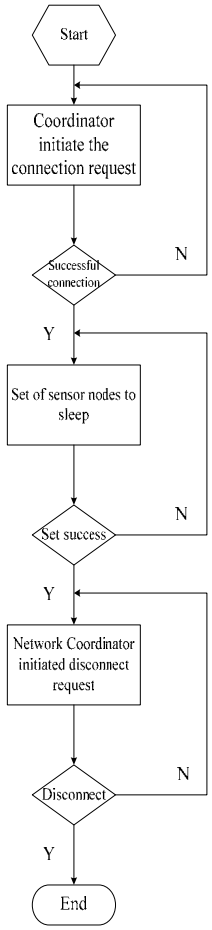


Figure 7 ZigBee module Initialization flowchart

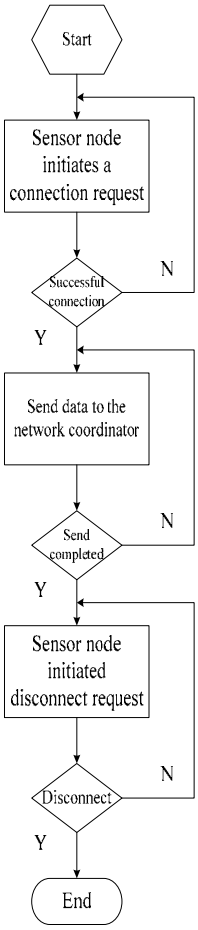


Figure 8 ZigBee module information processing flowchart

The program of the sensor nodes collect in the fixed time, A/D conversion, the fixed time of sending and dormancy, the nodes of router and coordination achieve the function of data forwarding and routing. Note that when programming, data communication between coordinating node and the GPRS modules should follow the appointed message so that the remote control terminal can analysis the reported message better[5].

VI. RESULTS OF EXPERIMENT

During experiment, six sites were selected and formed a small sub-cluster network and elected a cluster head. Every send the collected data in the fixed time to its cluster head. After finishing transmitting data, node was in dormant state. Cluster head finish fusion of data among clusters and compress the data packet, then transmit to routing node, and send data through a series of routing nodes to the GPRS module. GPRS module data will send data to the remote control terminal. Select three experimental points, to test the situation of user using water, electricity, gas meter. The experimental data show in Table 1.

TABLE I. THE RESULTS OF THREE EXPERIMENTAL POINTS

Property / Room number	Month	Water meter（A）	Electric meter（B）	Gas Meter（C）	Total (A*2+B*0.5+C*2.5)
421	*10*	*12.9 tons*	*169 kWh*	*5 m3*	*123.7 yuan*
419	*10*	*19.5 tons*	*361 kWh*	*19 m3*	*267 yuan*
422	*10*	*16.3 tons*	*254 kWh*	*11 m3*	*187.1 yuan*

In the testing process:

- GPRS network transmission effects influenced the system. Research has shown that using GPRS network to transmit data, the data is less than 128 Byte, and the communication delay about 2s. This method can increase the amount of data transmitting, thus it can avoid delaying communication time.
- The capacity of the buffer area of data, affect the quality of sending and receiving. When the capacity is too small, it can close data in transmitting, and even lead to GPRS module automatically reset and make the entire system collapsed. So in making program, it should use flow control.
- Saving energy is important. It can be achieved by increasing the system dormant time.

VII. CONCLUSION

Through study and analysis, the wireless remote meter-reading system is designed. This system combines ZigBee technology with GPRS network. It is using PIC18LF4620 as important processor, by CC2430 to do communication in short distance and SIM300 to achieve communication function in long distance, using RS-232 link communication joint to connect the communication between ZigBee and GPRS technologies. In this way it is convenient to copy the data of water, electricity and gas meter. It can full use the resources of networks. This system has low cost and a little power consumption, while it has great extension and security. It can be used in other areas widely.

REFERENCES

[1] JI Jin-shui. Zigbee wireless sensor network technology based on system design[J]. Computer Engineering and Design. 2009，28（2）：404-408.

[2] LIU Rui-xia, LI Chun-jie, GUO Qing, WEI Nuo, KONG Xiang-long. Cluster Routing Protocol Based on ZigBee Mesh Network[J]. Computer Engineering. 2009，35（3）：161-181.

[3] HE Ming-xing. Based on the ZigBee and GPRS technologies of wireless sensor network gateways design[J]. Indust ry and Mine Automation. 2009，8：106-108.

[4] JU Yu-peng，SHI Wei-bin. Design of remote automatic meter reading system based on ZigBee technology[J]. Network and Communication. 2009，15：38-44.

[5] LI Wen. Design of Remote Monitoring and Control System Based on ZigBee and GPRS[J]. Low Voltage Apparatus. 2009，12：37-44.

[6] WEI Shu-fang, SUN Tong-jing, SUN Bo, GUO Yuan-sheng. The Application of ZigBee - bas ed Wireles s Sensor Network in Coal Mine[J]. Control and Automation Publication Group. 2009.11（2）：65-67.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 190-192

The Medical Image Retrieval Based on the Integration of Corner and Texture Features

Jun-ding Sun[1,2], Xiao-yan Wang[1], Yuan-yuan Ma[1]
[1] School of Computer Science and Technology Henan Polytechnic University, Jiaozuo, China
sunjd@hpu.edu.cn
[2] Provincial Opening Laboratory for Control Engineering Key Disciplines, Jiaozuo, China
Wangxiaoyan10@yeah.net, mayuan200766@163.com

***Abstract*—This paper presents a medical image retrieval method based on the integration of corner and center local binary pattern (CLBP). The CLBP is adopted to describe the texture feature of the medical image firstly. Based on multi-scale curvature product (MSCP), the interest corners are extracted and then a new method is introduced to describe the shape feature. Then, such features are used together as the index for medical image retrieval. Experiment results show that the presented method not only reflects the texture information very well, but also has a rotating shift invariant feature, and it can improve the precision of medical image retrieval and the recall rate.**

***Index Terms*—medical image retrieval, CLBP algorithm, corner, Similarity**

I. INTRODUCTION

With the development of the technology, the application of the medical digital imaging devices in clinical becomes more and more widely and the technology of PACS aspects also receives a continuous development. Large amounts of medical images are produced in the hospital under such condition. Therefore, it has become a serious problem for the doctors to choose useful images from the medical image database for analyzing and diagnosing. It has made the medical image retrieval become one of the hot spots.

Because of the high gray scale resolution, high spatial resolution, high similarity, large quantity information contained in images and other characteristics, medical images are different from general images. Recently, most of image retrieval methods use the underlying image features when describing images [1], but most medical images are grayscale images and the majority of literatures use the texture feature. In recent years, the corner of the image has been widely used in image retrieval. Corner detection is an important basic research topic in the field of computer vision and pattern recognition. Accurate corner detection plays a crucial effect to the completion of many computer vision tasks (such as image matching, object recognition and motion analysis, etc.), and researchers proposed a lot of corner detection methods in recent years [2].

Considering the method of using a kind of feature can only express some extent part properties of the image. A new method is proposed in the paper which combines the corner feature and center local binary pattern for medical image retrieval.

II. TEXTURE FEATURE EXTRACTION

A. Introduction of LBP

Texture reflects the local structure of an image, which denotes the changes of pixel grayscale or color in a neighborhood. LBP (local binary pattern) is a simple and effective method of existing methods, and it has already drawn a lot of attention [3], and several extensions of LBP were also proposed. Both LBP and its extensions has been widely used in many areas of image processing and pattern recognition.

For each pixel in an image, LBP operated with eight neighboring pixels using the center as a threshold. If the gray-level of the neighboring pixels is larger or equal, the value is set to one, otherwise to zero. The final LBP code was then produced by multiplying the thresholded values by weights given by powers of two. The computation of LBP is given in Fig. 1.

B. Texture feature extraction method in this paper

It is obvious that the LBP descriptor produces rather long histograms (256) and is therefore difficult to be used as a region descriptor. To address the problem, the improved LBP algorithm, called Center Local Binary Pattern (CLBP), was proposed to extract the texture feature in this paper.

The main idea of CLBP is that the relation of the center pixel and the center-symmetric pairs of pixels instead of the gray-level difference between the center pixel and its neighborhoods are considered. The computation of CLBP value was denoted by Fig. 2, Eq. (1), Eq. (2) and Eq. (3). The new method can reduce the feature dimension effectively from 256 to 32. Therefore, CLBP is more effective for region description than LBP.

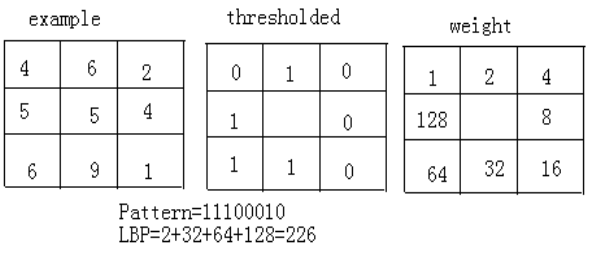


Figure 1. LBP operator

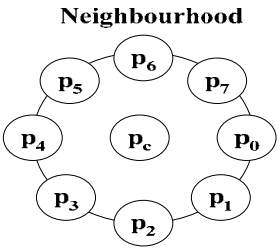


Figure 2. 8-neighbourhood.

$$s_k = \begin{cases} 1, & (p_k \geq p_c \geq p_{k+4}) \,|\, (p_k < p_c < p_{k+4}) \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

$$\mathrm{W} = \begin{cases} 1, & \mu_1 \geq \mu_2 \\ 0, & \mu_1 < \mu_2 \end{cases} \quad (2)$$

where, p_k, p_{k+4} ($k = 0,1,2,3$) and p_c are correspond to the gray-level of center-symmetric pairs of pixels and the center pixel on a circle respectively. The neighborhood's average pixel value is represented by μ_1 , and the entire image's average pixel value is expressed by μ_2 . Then CLBP can be calculated by Eq. (3).

$$CLBP = \sum_{k=0}^{3} S_k \cdot 2^k + W \cdot 2^4 \quad (3)$$

III. SHAPE FEATURE EXTRACTION

A. Corner detection and extraction

Corner detection is an important research area in computer vision and it plays an important role in describing object features for object recognition and identification. Furthermore, most contours based relevant point detection methods rely on detecting corners on the contour of a shape. Most of literatures about corner detection are single-scaled and work well if the object has similar size features, but are ineffective for objects with multiple-size features.

The image analysis theory based on scale space was proposed by Witkin in literature [4]. After that, the multi-scale curve analysis became the main method of the solution to detect the corner.

Mokhtarian et al. put forward curvature scale space (CSS) -based corner detection algorithm [5], and obtained good detection results. On the other hand, it is not sensitive to noise. However, because of using a single high scale and a global threshold in the process of corner detection, the method may miss the true corner and detect the false corner. Therefore, Mokhtarian and Mohanna used different scale to adopt different lengths of the outline to reduce the case of missing true corners [6]; Zhang Xiaohong, et al.[7] proposed the corner detection method based on multi-scale curvature product (MSCP). Though some true corners can be enhanced effectively by MSCP, it smoothes out some other corners. A novel algorithm for detecting corners was presented based on Curvature Scale Space (CSS) and Multi-scale Curvature Product (MSCP) by Jun-ding Sun and Qi-qiang Guo[8]. Firstly, the corners of an image are detected at different curvature scale space. Then, a multi-scale curvature polynomial is defined as the sum or multiplication of the curvature of the contour at each scale. This method can enhance the corner maximum, suppress noise and prevent some corners from being smoothed in the high scale.

The corner detection algorithm based on multi-scale curvature polynomials was used to extract medical images corners in this paper

B. The method of describing the corner characteristics

The corners of the image are detected firstly, then the distance between every corner and the centre of the image is calculated using Euclidean distance. Because the number of extracted corners in each image is not the same. We introduced relative difference, denoted by *VC*, to represent the corner feature. The new method is given by the following equation.

$$VC = \sigma / \mu * 100\% \quad (4)$$

where, σ 、 μ are the standard deviation and mean of the distance from corners to the center of images respectively.

IV. SIMILARITY MEASURE

Suppose H_1 and H_2 be the texture spectrum histograms of the query image and one image in the database. The distance of the two histograms can be written as,

$$dis1(H_1, H_2) = \sum_{i=1}^{K} |h_{1i} - h_{2i}| \quad (5)$$

where $K = 32$.

For the corner feature, the distance is defined as,

$$dis2(VC_1, VC_2) = |vc_1 - vc_2| \quad (6)$$

where VC_1 and VC_2 denotes the corner feature of the query image and the image in the database.

Finally, The distance of the two images is written as,

$$d = \alpha_1 \times dis1(H_1, H_2) + \alpha_2 \times dis2(VC_1, VC_2) \quad (7)$$

where α_1 and α_2 are thresholds and $\alpha_1 \in [0,1]$, $\alpha_2 \in [0,1]$, $\alpha_1 + \alpha_2 = 1$.

V. EXPERIMENT RESULTS

In this section, 156 CT scan images of the lung fixed to 512×512 pixels were used as test images. Experimental software environment is matlab 7.0.

The result of corner detection of an example image is shown in Fig. 3. Fig. 4 gives the retrieval results and the first image is the example image.

In order to deeply proved the method, 12 categories of images are selected from the image library as

example images and 12 retrieval results are obtained. Precision Recall is accepted as the evaluation criteria of retrieval algorithms in this paper. The image contrast curve of CLBP & VC algorithm, a single feature CLBP and the VC algorithm under the same environment was displayed in Fig. 5. It is indicated that the retrieval precision of the algorithm in this paper are better than the other two algorithms.

VI. CONCLUSIONS

The integration of CLBP and corner features for medical image retrieval was proposed in this paper, and experiments were done to compare this feature fusion algorithm with a single feature algorithm. It can be seen that this feature fusion algorithm gains better effect from the experimental results.

Figure 3. The result of corner detection

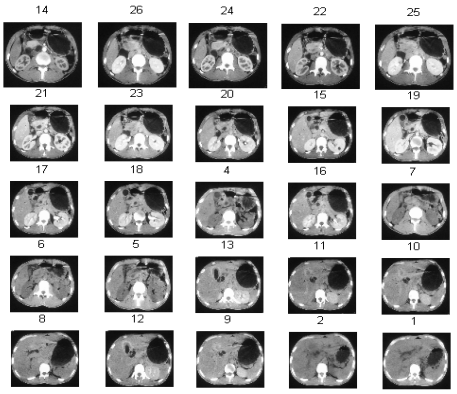


Figure 4. Retrieval results.

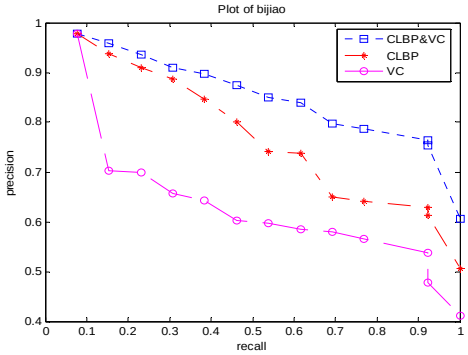


Figure 5. Contrast curve of the three algorithms.

REFERENCES

[1] MULLER H, MICHOUX N, BANDON D, and GEISSBUHLER A, “A Review of Content-Based Image Retrieval Systems in Medical Applications-Clinical Benefits and Future Directions [J],” International Joumal of Medical Informatics, 2004, 73(1): 1-23.

[2] Baojiang Zhong, and Wenhe Liao, “Corner detection techniques based on the refined curves of cumulative chord length,” Computer Aided Design and Computer Graphics, 2004, 16 (7) :939-943.

[3] Ojala T, Pietikainen M, and Maenpaa T, “.Multiresolution gray-scale and rotation invariant texture classification with local binary patterns [J],” IEEE Transaction on Pattern Analysis and Machine Intelligence, 2002 ,24(7) :9712987.

[4] Witkin, A.P, “Scale-space filtering,” In International Joint Conference on Artificial Intelligence, Karlsruhe, Germany, 1983, 1019-1022.

[5] Mokhtarian F, and Suomela R, “Robust image Corner Detection Through Curvature Scale Space,” IEEE Traps Pattern Analysis and Machine Intelligence, 1998, 20(12): 1376-1378.

[6] Mokhtarian F, and Mohanna F, “Enhancing the curvature scale space corner detector [C],” Proceedings 12th Scandinavian Conference on Image Analysis SCIA 2001, Bergen, Norway, June 11-14, 2001: 145-152.

[7] ZHANG Xiao-hong, LEI Ming, and YANG Dan, “Robust image corner detection based on multi-scale curvature product [J],” Journal of Image and Graphics，2007，7(12): 1270-1275

[8] Sun Junding, Guo Qiqiang, and Zhang Zhaosheng, “Contour Corner Detection Based on Curvature Scale Space,” Opto-Electronic Engineering, July, 2009: 0078-05.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 193-195

A power flow algorithm based on distributed computing

Zhao Zhi-Min[1], Gui Wei-Feng[2]
[1] College of Electric Information Engineering Pingdingshan University, Pingdingshan, P.R.China
Email: zzm0375@163.com
[2] Wanfang College of Science&Technology Henan Polytechnic University, Jiaozuo, P.R.China
Email: guiweifeng@hpu.edu.cn

***Abstract*—For power system, power flow analysis is very important to enhance the system security and efficiency. An algorithm based on distributed computing is proposed for power flow analysis. In this algorithm, the coefficient matrix transformed into a bordered block diagonal form (BBDF) matrix, and each block is sent to corresponding client to calculate. In IEEE standard power systems, this algorithm has been proved more effective than Newton-Raphson iteration and decoupled algorithm.**

***Index Terms*—distributed computing, power flow, Jacobian matrix, matrix partitions, client**

I. INTRODUCTION

In power engineering, the power flow analysis is an important tool involving numerical analysis applied to a power system. The great importance of power flow analysis is in the planning the future expansion of power systems as well as in determining the best operation of existing systems. So precise power flow analysis is foundation for power system planning and investment, and fast power flow analysis is precondition for power system structure and operation [1].

With the development of society and economy, and power grid becomes more and more huge, the power flow analysis is a difficult work because the computation is enormous. So the novel algorithms are needed to improve conventional power flow algorithm.

Distributed computing is a field of computer science that studies distributed systems. In distributed computing, a problem is divided into many tasks, each of which is solved by one computer [2].

Power system has many nodes, such as generator, substation, these nodes have many own computers, and these computers have been in a network [3]. Equipped with distributed program, these computers can make up of a distributed system to calculate the power flow.

In power flow analysis, distributed computing has four methods: partition method, multiple factoring method, sparse vector method and inverse matrix method. For partition method, it has specific physical meaning, and its program is easy [4]. In this paper, the distributed computing used partition method is applied to the power flow algorithm, this algorithm has been proved more effective than Newton-Raphson iteration and decoupled algorithm in IEEE standard power systems.

II. BASIC PRINCIPLES

A. Power flow equations

The power equations give the relationships between bus powers and bus voltages in term of the admittance parameters of the transmission system. In power system, there are following types of buses: there are load (P, Q) buses, generator (P, |V|) buses, and a slack or swing bus.

Power flow equations have two variables, one is voltage magnitude |V|, and the other is voltage phase θ. In polar coordinate, the power flow equation along a line (i, j) as follows:

$$\begin{aligned} P_i &= \sum_{j=1}^{n} |V_i||V_j|(G_{ij}\cos\theta_{ij} + B_{ij}\sin\theta_{ij}) \\ Q_i &= \sum_{j=1}^{n} |V_i||V_j|(G_{ij}\sin\theta_{ij} - B_{ij}\cos\theta_{ij}) \end{aligned} \quad (1)$$

In (1), the G_{ij} is conductance, and B_{ij} is susceptance. And the equations are nonlinear equations.

At the P, Q buses, the complex voltage is unknown because we do not know |V| and θ. At the P, |V| bus, Q and θ are unknown. The |V| and θ variables are implicit variables in the power flow equations and iterative solution methods are required.

B. Solution of power flow

There are several different methods of solving the resulting nonlinear system of equations. One method is Gauss iteration or its variation, Gauss-Seidel iteration. The complex form of the power flow equations is used. The iteration formula remains unchanged through the entire calculation.

Another method is Newton-Raphson iteration. The real form of the power flow equation is used. The iteration formula involves a Jacobian matrix that changes as the iterations proceed [5].

For computations involving power systems under the usual operating conditions, some simplifications of the Newton-Raphson scheme are usually possible. One of these modifications is called decoupled power flow. It still requires the updating of Jacobian matrices for each iteration, but the dimensionality of the computation is reduced. Another modification is called fast-decoupled power flow. In this case, the updating of matrices is no longer required and the computational burden is greatly reduced.

AP-PROC-CS-10CN007

Actually the above methods can be described solving large-scale sparse linear equations, and expressed as following:

$$Jx=b \tag{2}$$

In (2), J is the nodal admittance matrix, and it is called Jacobian matrix which is multi-dimensional, nonsingular, symmetrical and sparse. x is a unknown vector, b is a known vector [6].

C. Distributed computing for power flow

In order to perform distributed computing in a power system, the Jacobian matrix is transformed into a BBDF matrix [7]. Each block is sent to corresponding client to calculate. Therefore, the (2) can be formulated as follows:

$$\begin{bmatrix} J_{11} & . & . & J_{1k} & J_{1n} \\ . & . & & & . \\ . & & . & & . \\ J_{k1} & & & J_{kk} & J_{kn} \\ J_{n1} & . & . & J_{nk} & J_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ . \\ . \\ x_k \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ . \\ . \\ b_k \\ b_n \end{bmatrix} \tag{3}$$

Where subscripts 1,…, k, represent k client, and n represents the cutset block, respectively. Form (3), the unknown vector in the cutset block can be solved by the following equation:

$$x_n = (J_{nn} - \sum_{i=1}^{k} J_{ni} J_{ii}^{-1} J_{in})^{-1} (b_n - \sum_{i=1}^{k} J_{ni} J_{ii}^{-1} b_i) \tag{4}$$

The unknown vector in each client can be solved by the following equation:

$$x_i = J_{ii}^{-1}(b_i - J_{in} b_n) \tag{5}$$

The Jacobian matrix is divided with LU decomposition algorithm as following:

$$\begin{bmatrix} J_{11} & . & . & J_{1k} & J_{1n} \\ . & . & & & . \\ . & & . & & . \\ J_{k1} & & & J_{kk} & J_{kn} \\ J_{n1} & . & . & J_{nk} & J_{nn} \end{bmatrix} = \begin{bmatrix} L_{11} & & & & \\ . & . & & & \\ . & & . & & \\ L_{k1} & & & L_{kk} & \\ L_{n1} & . & . & L_{nk} & L_{nn} \end{bmatrix} \begin{bmatrix} U_{11} & . & . & U_{1k} & U_{1n} \\ & . & & & . \\ & & . & & . \\ & & & U_{kk} & U_{kn} \\ & & & & U_{nn} \end{bmatrix} \tag{6}$$

Substituting (6) into (4), the unknown vector in the cutset block can be rewritten as (7):

$$\begin{aligned} x_n &= (J_{nn} - \sum_{i=1}^{k} L_{ni} U_{ii} U_{ii}^{-1} L_{ii}^{-1} L_{ii} U_{in})^{-1} \times \\ &(b_n - \sum_{i=1}^{k} L_{ni} U_{ii} U_{ii}^{-1} L_{ii}^{-1} b_i) \\ &= (J_{nn} - \sum_{i=1}^{k} L_{ni} U_{in})^{-1} (b_n - \sum_{i=1}^{k} L_{ni} L_{ii}^{-1} b_i) \end{aligned} \tag{7}$$

Substituting (6) into (4), the unknown vector in each client can be rewritten as (8):

$$\begin{aligned} x_i &= U_{ii}^{-1} L_{ii}^{-1} (b_i - L_{ii} U_{in} b_n) \\ &= U_{ii}^{-1} (L_{ii}^{-1} b_i - U_{in} b_n) \end{aligned} \tag{8}$$

III. DISTRIBUTED COMPUTING FOR POWER FLOW

A. Structure of distributed computing in power system

The structure of distributed computing in power system is illustrated in Fig. 1.

In Fig.1, the computer in dispatching center is host, and the computers in substation are clients. The host and clients are connected with ethernet network. The host builds the nodal admittance matrix according to power grid, and the nodal admittance matrix is separated into several matrix partitions, and these matrix partitions are sent the client to calculate the power flow.

In separating of the nodal admittance matrix, not the host but the connection mode of the power grid decides the number of matrix partitions.

B. Flow chart of distributed computing in power system

In distributed computing, the flow chart is shown by

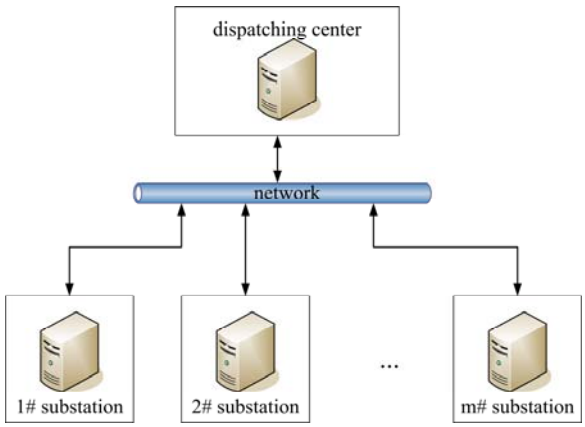


Figure 1. Structure of distributed computing in power system

Fig.2. First, the host forms the nodal admittance matrix, and then divides the matrix into several computers in substations, last, the client computers calculate and the result. The flow chart of distributed computing is shown by Fig.2.

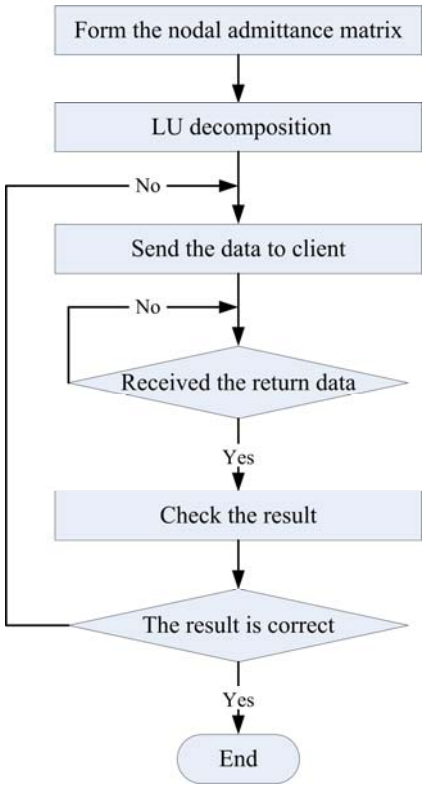


Figure 2. Flow chart of distributed computing in power system

Actually, in above program, a timer is set in order to avoid endless loop. The program will be exit if the timer achieves the preset number.

C. Simulation result

For the limit condition, this algorithm is simulated in Pentinum (R) personal computer. In IEEE-30 bus and IEEE-39 bus standard power systems, Newton-Raphson iteration, decoupled and distributed computing is used to calculate the power flow [8].

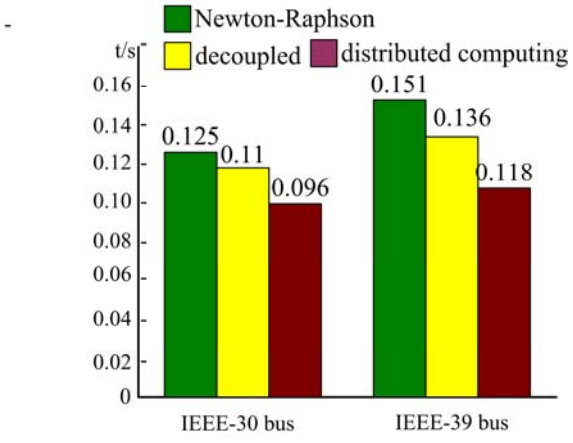


Figure 3. calculation time of three algorithm

Fig. 3 presented three algorithms cost time in power flow analysis in two power system. In IEEE-30 bus , the time of Newton-Raphson is 0.125 second; the decoupled is 0.11 second, and the distributed computing is 0.96 second. And in IEEE-39 bus, Newton-Raphson is 0.151 second; the decoupled is 0.136 second, and the distributed computing is 0.118 second.

The simulation result showed that the distributed computing algorithm is effective than Newton-Raphson iteration and decoupled algorithm.

IV. CONCLUSION

Distributed computing becomes more and more popular in power flow analysis, and the power system is a symmetrical and sparse, so the BBDF is fit for the distributed computing in power system. This study presents a distributed computing algorithm with BBDF method. In this algorithm, the nodal admittance matrix is divided into several blocks with LU decomposition method, and each block is sent to corresponding client to calculate the power flow. At last, this algorithm is simulated in IEEE-30 bus and IEEE-39 bus standard power systems, the simulation result proved availability of the distributed computing algorithm.

REFERENCES

[1] A. Vaccaro, D. Villacci, “Radial Power Flow Tolerance Analysis by Interval Constraint Propagation,” Power Systems, IEEE Transactions on, vol.24, pp.28-39, Feb. 2009 .

[2] K. Khan, R. Haines, J.M. Brooke, “A Distributed Computing Architecture to Support Field Engineering in Networked Systems,” Complex, Intelligent and Software Intensive Systems (CISIS), 2010 International Conference on, pp.54-61, 15-18, Feb. 2010.

[3] P. Gajalakshmi, S. Rajesh, “Fuzzy modeling of power flow solution,” Telecommunications Energy Conference, 2007. INTELEC 2007. 29th International, pp.923-927, Sept. 30 2007-Oct. 4 2007,.

[4] G. Granelli, M. Montagna, G. Pasini, P. Marannino, “A W-matrix based fast decoupled load flow for contingency studies on vector computers,” Power Systems, IEEE Transactions on, vol.8, no.3, pp.946-953, Aug 1993.

[5] W.F. Tinney, C.E. Hart, “Power Flow Solution by Newton's Method,” Power Apparatus and Systems, IEEE Transactions on, vol.PAS-86, no.11, , pp.1449-1460 Nov. 1967.

[6] Zheng Wei-Shi, S.Z Li, J.H. Lai, Liao Shengcai, “On Constrained Sparse Matrix Factorization,” Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on, pp.1-8, 14-21 Oct. 2007.

[7] A.I. Zecevic, D.D. Siljak, “Parallel solutions of very large sparse Lyapunov equations by balanced BBD decompositions,” Automatic Control, IEEE Transactions on, vol.44, no.3, pp.612-618, Mar 1999.

[8] P. Kachore, M.V. Palandurkar, “TTC and CBM Calculation of IEEE-30 Bus System,” Emerging Trends in Engineering and Technology (ICETET), 2009 2nd International Conference on, pp.539-542 , 16-18 Dec. 2009.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 196-199

Research on Dynamic Simulation of Indoor Scenes

Wang Yu-kun[1], Wang Xing[2]
[1] Institute of Computer Science & Technology, Henan Polytechnic University, Jiaozuo, China
Email: wyk@hpu.edu.cn
[2] Institute of Computer Science & Technology, Henan Polytechnic University, Jiaozuo, China
Email: wangxing_job@126.com

***Abstract*—According to design three-dimensional digital modeling based on two-dimensional drawings data and its annotation information and build up the original three-dimensional model in AutoCAD, and then the paper figures it out in 3DMAX further. And export to a file with the cpp extension name by explor3d tool. Developing a system which can carry out indoor dynamic scene rendering and interactive roaming by using VC++6.0 and OpenGL development kit. The main content of the paper is research the way of the data store of three-dimensional model and processing dynamic scene rendering.**

***Index Terms*—real-time scene rendering, OpenGL, Three-dimensional data store, VC++, Interactive Roaming**

I. INTRODUCTION

As the rapid development of computer hardware and software [1], previously time-consuming and expensive three-dimensional realistic image synthesis technology has been accepted increasingly and used widely. Fast rendering dynamic scenes has been the difficult and hot in computer graphics and virtual reality research. Dynamic simulation of indoor scenes is an important application in building of dynamic scenes. According to the interactive simulation of dynamic scenes indoors, various objects in the scene can be selected by the mouse, and then interact with the selected object. Objects in the scene can be added, deleted, modified and reproduced. The operations of object scaling, translation, rotation can be achieved through the graphical transformation. At the same time lighting and texture mapping can be adjusted so as to enhance the effect. Building Decoration Company can provide customers with the design of virtual building interior walkthrough in order to attract customers. Customers can have a more intuitive experience before the interior decoration, and the structures of space, form, sound and light and other effects can be appreciated in immersed sense. So the design can be changed timely and make it more perfect.

II. THE OVERALL ARCHITECTURE DESIGN

This section will introduce four-layer structure model design (Fig.1). Next, selection of tools and techniques will be analyzed (Fig.2).

A. Four-layer structure model design

The design of three-dimensional model is the core, and the overall architecture can be abstracted into four structural models shown in Fig.1 below.

In this framework, L1 is the bottom. Texture data and three-dimensional model data can be stored in this layer. And L1 is mainly responsible for provision of raw data to the upper so as to build and render three-dimensional model. At the same time, if the texture data or model data was revised through the top by the user, this module can also save the modified data. L2 is three-dimensional model layer. The layer is the core of the system. Whether three-dimensional model data, and texture data, or rendering, interactive modules, are based on the layer. OOA vision is abstracted as an independent entity with its own attributes, such as status, size, location, texture, etc. Also it has its own methods, such as show, hide, move, copy, delete, rotate, etc. This layer is based on the data of the L1 layer and it is the object of the upper handle individual. L3 level is rendering module, it is mainly responsible for realistic processing and calculate the intensity distribution and display of L2 layer of the target individual in accordance with the requirements of the upper. This layer is also the UI intuitive interface layer which the user can see. L4 is the top of this design and it is also known as the interaction layer. It contains the object picked up, moved, rotated and other operation control. Peripherals including keyboard, mouse, etc are used to achieve human-computer interaction (HMI) in this layer. The results processed and packed in this layer will be sent to the L3, and the display of L3 layer will

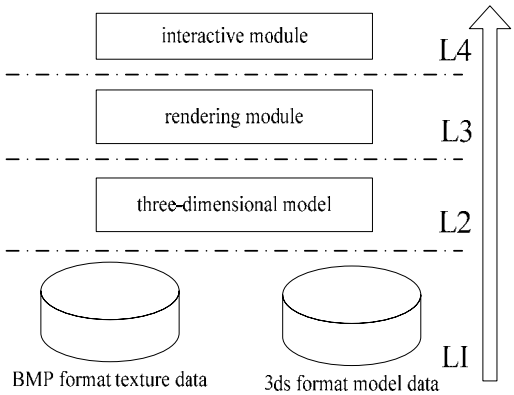


Figure 1. The overall architecture design

directly be affected. Each layer can communicate with each other by sending message.

This system architecture can reduce the development effort and shorten the development cycle and it also reduces maintenance costs and facilitates the promotion.

B. Selection of tools and techniques

OpenGL graphics library is developed from SGI's GL (Graphic library) graphics library. Currently, it is supported by several IT giants including Microsoft. Which provide a strong technical background for OpenGL, Almost every graphics card manufacturers have accelerated chips for OpenGL; therefore its use has a broad prospect. OpenGL provides a wide range of drawing method for three-dimensional objects [2], including the depth, anti-aliasing, flat shading, interaction, plus shadows and texture, and many other functions. There are several advantages of OpenGL: Firstly, universality. Most software and hardware vendors are supported. Secondly, stability. OpenGL has been stable for years and have a number of successful large-scale applications. Thirdly, cross-platform nature. OpenGL can be transplanted and run in the windows, UNIX, Linux, MAC. What's more, OpenGL has powerful functions.

VC++6.0 is an IDE development tool of Microsoft. The power of MFC class library provides the conditions for rapid development of the system. The object-oriented C++ language is used by the tool. The data structure can be easily built in OOP way so as to achieve the four structural model of the design.

III. GENERATION AND DATA STORAGE OF THREE DIMENSIONAL MODEL

The data source of three-dimensional model is mainly from images, good design drawings or the object itself. In regard to three-dimensional model for the needs of virtual simulation, engineering design drawings are generally well done in advance and then modeling by the use of digital drawings and data can be collected by drawing annotation. The process can be shown in Fig.2. After collecting data, there are generally two ways of processing the data collected: One approach is to use AutoCAD to establish the model directly, the model constructed in this way is high accuracy and virtually no errors generated.

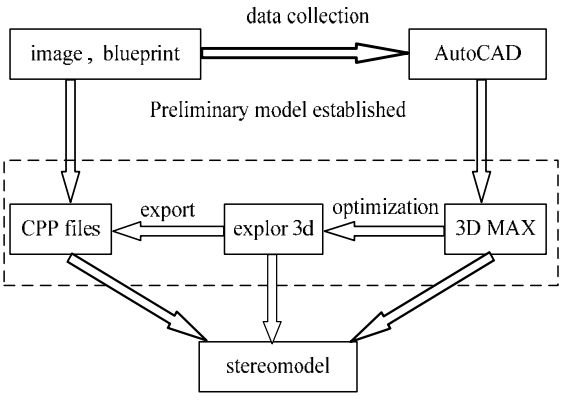


Figure 2. Generation of three-dimensional model

Although the rule of geometry is very convenient and fast, it is not good at dealing with irregular geometry. The other method is that 3DMAX is used to create more complex, irregular surface. Compared to the AutoCAD, 3DMAX is more adept at the complex and irregular surface processing. Two ways have their own benefit. The combination of AutoCAD and 3DMAX is used in this design and explor3d is used for post-processing. Firstly, a more precise three-dimensional model is obtained by using AutoCAD. Secondly, it should be imported into 3D MAX in further complex modeling. The model processed is stored temporarily in the format of 3ds file. And then the 3ds file is opened by using explor3d, Files are stored as the source file which can be directly applied to the VC++6.0 in the cpp format. The function of L1 and L2 layers of the four-layer structure can be completed in the system through this manner.

IV. THREE-DIMENSIONAL SCENE REAL-TIME RENDERING

Three-dimensional scene real-time rendering needs have the generated realistic three-dimensional model processing and achieve real-time rendering. Realistic handling means processing a series of model data so that make it looks closer to the real world. Real-time rendering is the graphics data in real-time calculation and output. The most typical graphical data source is the vertex. Vertex includes position, normal, color, texture coordinates, vertex weights and other information.

A. Realistic handing

Realistic treatment of three-dimensional model is mainly achieved by texture mapping technology in this design. Texture mapping technology should be used in order to make users could change the murals, wallpaper arbitrarily. An image is mapped into a polygon surface by texture mapping. And when the polygon be transformed or rendered. The correct behavior can be showed by the image mapped to the polygon surface. Implementation of texture mapping requires the following steps [3]:

1) Texture object is created and a texture is specified.
2) Determine how texture is applied to each pixel.
3) Texture mapping function is enabled.
4) Scene is drawn, texture coordinates and geometric coordinates are provided.

Fig.3 is a simple form for realistic handling process.

B. Realistic rendering

Indoor roaming must be real-time; otherwise, the results will be bad. There are two common methods of real-time rendering of realistic images [4]: One is the light tracing algorithm, and the other is the radiosity

Figure 3. Simple form of realistic process

algorithm.

The process of reflection and refraction of light between the smooth surfaces features have been successfully simulated by global illumination model of ray tracing algorithm. With the transfer point of view, the information of points of light texture model must be calculated and mapped out in real time. As the ray tracing technology is built on the basis of spatial sampling point, the number of sampling points is often very limited, which makes many of the performance of lighting effects are not correct. At the same time, the information of the light texture points must be calculated in real time and precisely. So the higher machine's hardware is required. OpenGL accelerator and faster light-processing algorithms are needed.

Radiosity algorithm is composed of the environment surface as a closed system, and assuming that the closed system of surfaces are diffuse surface, and then calculated the energy of each surface according to the energy balance in purpose to find the brightness of the observed points, and then quickly display the scene realistic view of different observation angles. Because of consideration of energy transfer between the surface of the closed system, and thus it is easy to calculate the correct light distribution from the overall environment and the hardware requirements are lower. However, the algorithm can not assess the visibility of drawing elements, which will consume some resources to calculate the distribution of invisible light elements.

Through the above analysis of two commonly algorithms, the design proposes a radiosity algorithm with visibility culling which meaning do visibility culling firstly before calculating radiosity to process real-time radiosity rendering. The process shown in Fig.4.

Visibility culling techniques [5] (Visibility Culling) is drawn through assess whether a drawing element or group of elements is visible or invisible in the list, then the invisible elements would be removed quickly from the list, the number of the drawing elements that sent into the rendering pipeline could be reduced, so that improve the rendering speed to achieve fast rendering of virtual world technology. Visibility culling techniques can draw all the elements contained in the virtual world refine a similar set.

According to the assessment based on visibility, visibility culling algorithm can be divided into three categories.

1) Horizon culling: If the sight of a drawing element out of the four-level pyramids, it will not be visible.

2) Occlusion culling: If a drawing element is blocked by other opaque elements (group), it will not be visible.

3) Back culling: If the drawing element's normal direction departure from the direction of observation viewpoint, it will not be visible.

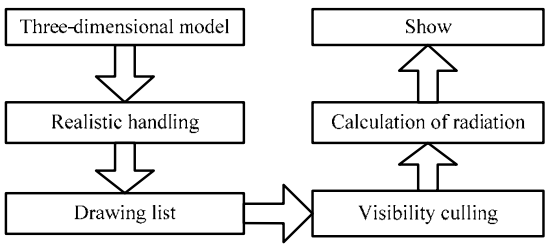


Figure 4. With visibility culling algorithm on the radiation treatment process

After culling the elements contained in the draw list, at this point, as only visible element or element group exists in the list, next using the Radiosity algorithm to calculate light distribution, which could reduce a lot of computation. Thus rendering speed could be speed up and will not affect the rendering quality. The following example shows a simple key code:

```
void RenderScene()
{
    glClear(GL_COLOR_BUFFER_BIT|
    GL_DEPTH_BUFFER_BIT);
    glPushMatrix();
    // Position / translation (mouse rotation)
    glMatrixMode(GL_PROJECTION);
    glLoadIdentity();
    glTranslated(0.0,0.0,-8.0);
    glRotated(m_xRotate, 1.0, 0.0, 0.0);
    glRotated(m_yRotate, 0.0, 1.0, 0.0);
    glScalef(m_ScaleX,m_ScaleY,m_ScaleZ);
    // give the List index
    ::glCallList(index);
    glPopMatrix();
}
```

Realization of the complex physical can use the identical manner, but the data of the model is large, L3 layer's rendering module could be achieved through the realistic handing and real-time rendering.

V. HUMAN-COMPUTER INTERACTION

As the design is a roaming system. Thus it needs to achieve human-computer interaction [6]. Speaking at the computer, the function should be achieved mainly through the keyboard and mouse. Use the up, down, left and right arrow keys of keyboard and the mouse up, down, left, right movement to control the viewpoint changes, thus changing the current window display content. The roaming function of module design is in the L4 level.

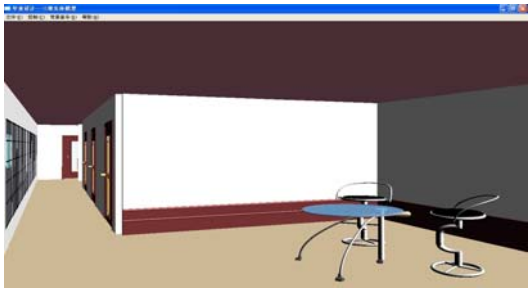
VI. CONCLUDING

This paper describes the creation of indoor dynamic scenes, as shown in Fig.5, the effect picture of indoor roaming shown in (a) and the all-round observation of digital building model shown in (b).

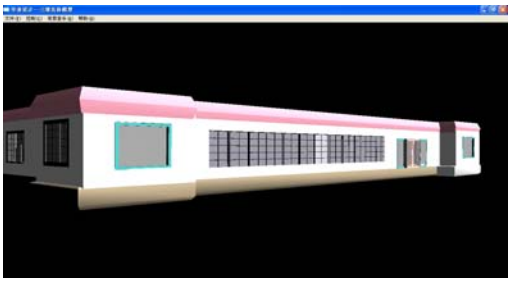
The paper mainly focused on the three-dimensional model data are generated and stored procedures, and real-time 3D scene rendering method. The paper improve the radiosity algorithm [7], which is used to render the Real-time 3D scene, Drawing list should be culled visibility before the data of the calculation of intensity distribution could be reduced and speed up the rendering speed, calculation of radiation intensity distribution, in this way, the data of the radiosity calculations could be reduced and speed up the rendering speed.

The human-computer interaction in the paper stated simply because of the limitation. In addition, the system

is still in the primary stage of development, it just makes more depth in solving the problem three-dimensional scene in real-time rendering, saving and modifying model data real-time have not be carried out in the aspect of the human-computer interaction. Further research is needed.



(a) The effect picture of indoor roaming



(b) The digital building model

Figure 5. Dynamic scene walkthrough system

ACKNOWLEDGMENT

The author received lots of well-advised guidance from tutor and amounts of benefits from his remarkable thinking, acute experience and bland working style when this paper was composed. Here, please allow me to thank him sincerely. Meanwhile, we want to express our appreciation towards Henan Polytechnic University and everyone who helped us before.

REFERENCE

[1] Zhou Xiaobo, Guo Shunsheng, Yang Mingzhong, “LED Lighting Simulation System Based on OpenGL,” Computer Simulation, Vol 21, No.1, January 2004, PP. 93-95.

[2] Yang Bo, “Interactive Stage Scenery Simulation Based on OpenGL,” Journal of Wuhan Metallurgical Manager’s Institute, Vol 12, No. 1, March. 2002, pp.71-73.

[3] Zhu Xiaoqiang, Xie Minghong, Ye Li, Yang Dianlong, “Texture Mapping Technology on VC,” Microcomputer Applications, Vol 29, No.4, April 2008, pp. 82-87.

[4] Hiroshi Sakurai, David C.Gossard. “Solid Model Input Through Orthographic Views.”Computer Graphics, Vol 17, No.3, May 1983, pp. 243-252.

[5] Liu Shixia, Hu Shimin, Wang Guoping, Sun Jiaguang, “Rec0nstructing of 3D Objects from Orthographic Views,” Chinese Journals Computers, vol.23, No.2. pp.141-146, Feb. 2000.

[6] Ding Mei, Hu Zhiqiu, “Virtual Architecture Model Space Roaming System,” Computer Technology and Application, vol.1, No.1. pp.46-48, Feb. 2005

[7] Gong Lin, Gu Daquan, “Simulation of 3D dynamic natural scenes based on fractal,” Science of Surveying and Mapping, vol33, No.4. pp.79-81, July. 2008

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 200-203

The Design and Development of Land Use Planning Management Information System at city level

He-Bing Zhang[1], Xiao-Hu Zhang[1], Gang Li[2]
[1].College of Surveying & Land Information Engineering, Henan Polytechnic University, Jiaozuo, China
Email:Jzitzhb@hpu.edu.cn, zhangxiaohu@hpu.edu.cn
[2].College of Resource and Environment, Northeast Agricultural University, Harbin, China
Email: shybabycn@yahoo.com.cn

***Abstract*— The design and development of land use planning management information system is the scientific guarantee of dynamic implement and real time management planning. Take Hebi city as example, we try to design the land use planning management information system basing on GIS. Using object-oriented modularization method and general-purpose programming language VB.NET 2005, we repeatedly develop the MapGIS and couple SQL Sever 2000 and Access 2003, in order to realize the functional module of land use planning management information system. The system provides a new model and measure for the information management. The results indicate that the systematic construction's design is rational, the system works steadily and can realize the dynamic, real time and informatization management of planning data.**

***Index Terms*—Land use planning, MIS, GIS, Repeatedly develop**

I. INTRODUCTION

The land is foundation of human survival and development. As the global land problems' increasing, how to use the land scientifically, rationally and effectively has become a question to the world[1]. Compiling and implementing the land use planning is the key to it. Currently, it has a tentative system adapt to our country. As the quickly development of economic society and the planning refer to mass land spatial and attribute information. The traditional model is not fit for the time require. In China, the planning at city level is a connecting link between province level and county level, which has the characteristics of micro-workable with macro-control. It not only relates to the practice of province planning but the city land use scale and reasonable arrangement. In this way, the city level planning results is very important. It's an availably form to construct land use planning management information system with the help of GIS. It can get and process data rapidly and exactly, realize the real time and informatization management. Therefore, it can realize the integration of "spatial and attribute information management – information procession and statistical analysis – teletext results export", which has become the hot spot in the land science research[2-5]. These years, the techniques of MapGIS repeatedly development is being perfect. Basing on this function , it has meaning to construct a land use planning management information system, which conclude information acquisition, procession and results management[6].

Taking Hebi as example, the paper uses object-oriented modularization method and general-purpose programming language VB.NET 2005, repeatedly develops the MapGIS and couple SQL Sever 2000 and Access 2003. We design and develop the land use planning management information system at city level in order to make the planning results electronize and informatization. It offers prompt and exact data, give technical measures and supports for the implement of the land use planning management information system at city level.

II. SYSTEM ANALYSIS

Considering the characteristics and work flow of the land use planning management information system at city level, it's usual work mainly concludes construction land pre-qualification, special construction land examination and approval, land use planning examination and approval of village (town), land use change and control examining, planning achievements management, planning implement management and so on. The system can be divided into several module and design separately. They must satisfy: (1) the construction land pre-qualification, examination and approval, planning examination and verify and land use change must has procedures; (2) the numerical statement and pictorial statement export function; (3) planning monitor; (4) achievements and implement management; (5) the planning notice and documents bring out on the net.

III. DESIGN AND DEVELOPMENT CLUE

A. Desing clue and development target

Basing on the characteristics and work flow of the land use planning management information system at city

"Evenleth Five-Year Plan" National Tecnology Supporting Plan Programe(2006BAG05A14); Natural Science Research Programe of Henan Province department of Education(2009A40002); Ph.D Fund of Henan Polytechnic University(B2010-87)

level and land use planning of Hebi, we repeatedly develops the MapGIS 6.7 using VB.NET 2005 and couple SQL Sever 2000 and Access 2003. A land use planning management information system at city level has functions, such as land use planning, geography data base query and statistical analysis.

We use system engineering principle making the planning management programmatic, which can be identified and processed by the computer. Try to achieve man-machine interaction by guide mode, create planning management database, and increase the automatic level of it. Through the friendly system interface, we realize the informatization of data management.

B. Development and working environment

The configuration of development environment is high required: CPU must be better than Intel 4-1.2GHz, internal memory must be better than 512M, video memory must be better than 128M, hardware is 40G, OS is Microsoft windows XP Professional PS2, programming software is VB.NET 2005; database software is Access 2003 and SQL Sever 2000, GIS software is MapGIS 6.7, MapGIS 6.7 SDK, development instrument is Microsoft Visual Studio 2005. The system operation environment: CPU is Intel 3-733MHz, internal memory is 128M, hardware is 20G, OS is Microsoft windows XP Home.

IV. SYSTEM DESIGN

A. System construction

According to the systematic target and requirement, the system is divided into five work platform, which are usually work platform, conference platform, monitor platform, website platform and system maintenance platform. In this way, the system database is felled into main database, geographic information database and website database.

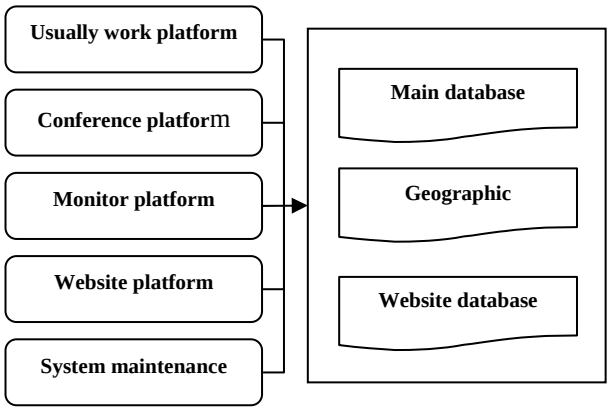


Figure1. System overall structure drawing

B. System function

- Usually work platform: concluding the construction land pre-qualification module, special construction land examination and approval module, land use planning examination and approval of village (town) module, land use change module, planning achievement management module and planning implement management module. It has functions: land use data statistical analysis, attribute resources querying, printing reports, and browsing geographic information and user personal data management.
- Conference platform: concluding staff module and expert leadership module. It has functions: conference preparing, conference introducing, program introducing, recommending, instructs feed backing and programs signing.
- Monitor platform: its main functions are browsing the planning implement by time sequence and item searching. Monitoring the planning targets through statistical analysis. At the mean time, it can explain and label questions, edit and print questionable material, send questions to professional stuff and wait for reply and so on.
- Website platform: news of land use planning, notices of land program, land use querying, information assorting. And it has functions: searching in the website, user login, introduce of website and entity.
- System maintenance platform: concluding mainstay system maintenance module and website backstage module. It has functions: user management, data back up and newly assort.

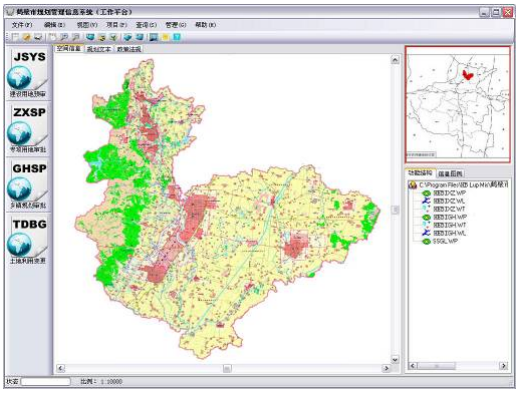


Figure2. Interface of usually work platform

Figure3. Interface of conference platform

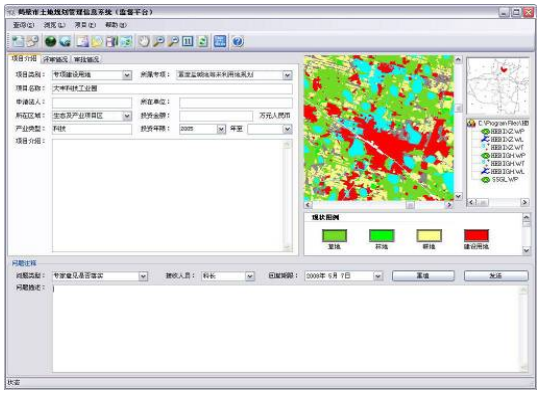


Figure4. Interface of monitor platform

Figure5. Interface of website platform

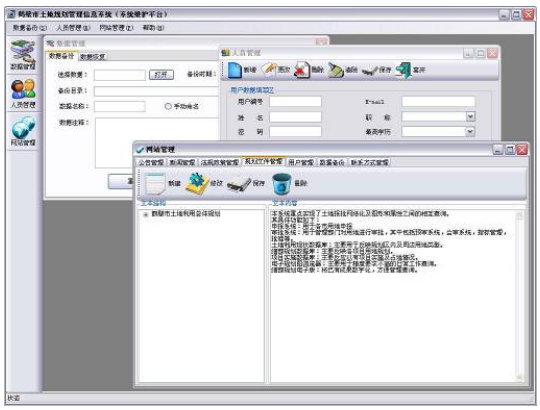


Figure6. Interface of system maintenance platform

C. *Construction of database*

It is consisted of main database, geographic information database and website database.

- The pattern of main database is SQL, concluding basic data tables, program qualification tables, implement management tables and interaction tables.
- Geographic information database is files with the pattern of point, line and area in MapGIS. It concludes points, lines and areas files of land use, land planning and land implement area.
- Website database concludes bulletin tables, news tables, planning files and invest reference tables, information back up tables, contacts tables and so on.

V. SYSTEM WORK PROCESS

The planning management can realize the functions through corresponding work platform and access corresponding database.

(1) The usually work platform, monitor platform and conference platform are the main planning management platform. They are connected with main database and geographic information database, such as program examination and approval, file management, monitor management, conference management, which realize date exchange between function platforms and main database and geographic information database.

(2) The website platform is connected with planning website database and geographic information database. The information browsing, suggestions feeding back and planning figures on it are directly exchanged through the website platform and website database and geographic information database. Furthermore, the website database connects the geographic information database unilaterally, which can get the update information.

(3) The system maintenance website platform is connected with main database and planning website database, which is used for data back-up, data renew, system user management. The management of website platform is mainly implemented by the system maintenance website platform operating the planning website database.

(4) All the platforms need vector data browsing and attribute data query function, except the system maintenance platform. Each platform accesses geographic information database, which realize the spatial entity browse function. Then, they access main database through relevant spatial entity ID, which realize the attribute data query function.

VI. CONCLUSION

The design basing on the usually management flow and require of land use planning at city level, referencing corresponding technology regulation and guide, using software engineering methods, combing GIS, database, net communication, design and develop the database. The system has functions of planning results management, statistic, query and export, which achieves the unified management of spatial and attribute information. The techniques that “program information unified management” and “flow processing batch by grade” has realized “obtaining and managing the spatial and attribute information – information processing and statistical analyzing – figures and texts exporting”, which has obviously increased the management level of the planning results’ scientific and automatic degree. The results indicate that the systematic construction’s design is rational, the system works steadily and can realize the dynamic, real time and informatization management of

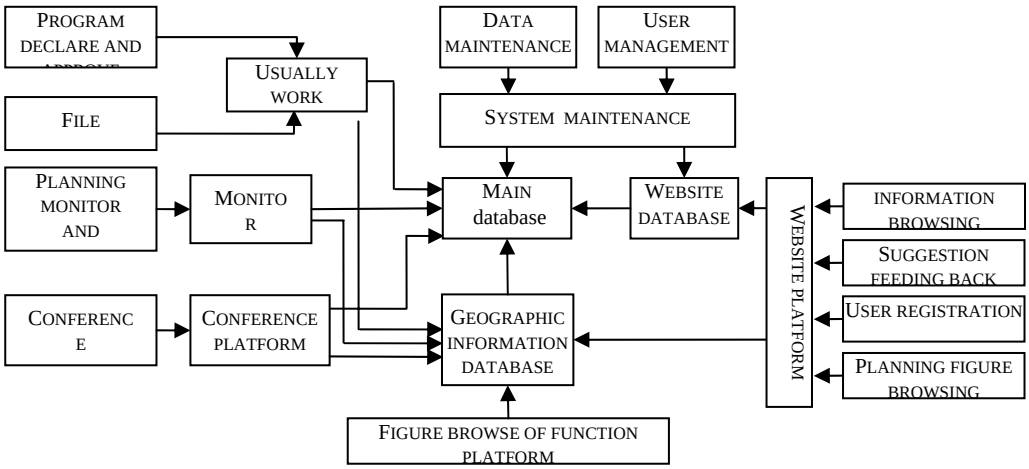


FIGure7. System workflow chart

planning data. It provides reference to the informatization of territorial resources management.

REFERENCES

[1] ZHAO Su-xia, The Design and Research of Land Use Planning Management Information System *2008 International Conference of Management Science and Engineering*, vol.16, pp. 1166-1170, January 2008.

[2] Sumbangan Baja， David M. Chapman， Deirdre Dragovich. Spatial based compromise programming for multiple criteria decision making in land use planning. *Environmental Modeling & Assessment*，vol.12, pp. 171-184, August 2007.

[3] YU Xiao，MIAO Fang. Information System for Land-Use Planning and Management. *JOURNAL OF SOUTHWEST JIAOTONG UNIVERSITY(ENGLISH EDITION)*, vol.16, pp. 372-379, October 2008.

[4] Frank Witlox. Expert systems in land-use planning: An overview. *Expert Systems with Applications*, vol.29, pp. 437–445, May 2005.

[5] Yuxiang Cao，Hongjiang Liu，Jianchun Xu. Design of Land-Use Planning Management Information System Base on ArcGIS. *2009 WRI World Congress on Computer Science and Information Engineering*, vol.2, pp. 359-363, February 2009.

[6] Uday Bhaskar Nidumolu, CAJM de Bie, Herman van Keulen, Andrew K Skidmore, Karl Harmsen. Review of a land use planning programme through the soft systems methodology. *Land Use Policy*, vol.23, pp. 187-203, April 2006.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 204-207

A Research on Micro Simulation of Signalized Intersection Based on Arena

Qinjun Zhang,Huiyuan Jiang
Wuhan University of Technology,Wuhan, China
sc20050206@126.com

***Abstract*—Comparing with the traditional traffic simulation software, Arena can define a discrete random function and can make simplification of the operation by using customizable interface. The discrete environment of modeling and simulation provided by Arena has obvious advantages during the research on the complex queuing system. In this paper, Arena is applied to resolve the problem of intersection. By using Arena's modeling and simulation method, this paper analyzes the flow of signalized intersection, and studies the relationship between signal control system and each subsystem. Special modules of signalized intersection are built in the discrete environment of modeling and simulation of Arena. It is concluded that these modules can be used expediently to study on simulation modeling, scheme adjustment and result analysis of a certain intersection.**

***Index Terms*—Signalized Intersection, Micro Simulation, Arena**

I. INTRODUCTION

With the rapid development of social economy, the traffic problem has been paid more and more attention. As the throat of urban traffic, the capacity of urban Intersection has restricted urban development. Intersection is an extremely complicated system, which is controlled by multiple factors and big randomness. It is far from satisfying by building an exact mathematic model or preestablishing a control schemes. In recent years, with the development of the computer technology, Traffic simulation had been one of the research highlights. As a system simulation software, the modeling environment of Arena is discrete. It is useful for studying complicated queuing system. Compared with traditional simulation software (such as VISSIM), Arena has the following advantages: (1) Make simplification operating by using customizable interface; (2) Arena can custom random distribution functions, which can be used to analyze some Intersections having special traffic flow. In this paper, it will try to take Arena to research the problem of Intersection.[1]

II. ANALYZE THE FACTORS OF INTERSECTION SYSTEM

A. Determine the Boundary of the Target System

Before starting to research the system, we must know the boundary of the target intersection system, and find out factors which must be included in the system while others can be ignored. Make it as simple as possible, and make sure that the primary problem we study can be answered, and can reflect the real system. The main objective of signal control is to reduce delay as possible. If a vehicle arrives at the intersection, we will think it has arrived at the system; if a vehicle drives away from the intersection, we will think it has leaved the system. Then, we define the exit and the entrance as the boundary of the system. In the successive research, we will adjust the boundary because of the change of the research scope.[2][3]

B. Process Analysis

When a vehicle arrives at the entrance of the intersection, it will choose different lanes based on different destinations and the using condition of driveway. If the signal lamp turns red or yellow, then stop; If it is green, then look around to make sure if there is a vehicle ahead, IF Yes Then drive follow it, IF No Then drive away from the intersection. See Figure2-1.

C. Object Aalysis

In order to build the intersection model, abstract each real entity as an object.[4] The objects have the same features with the real entity. By analyzing the flow of signalized intersection and all subsystem, and abstracting all kinds of the factors of the real system, we can obtain

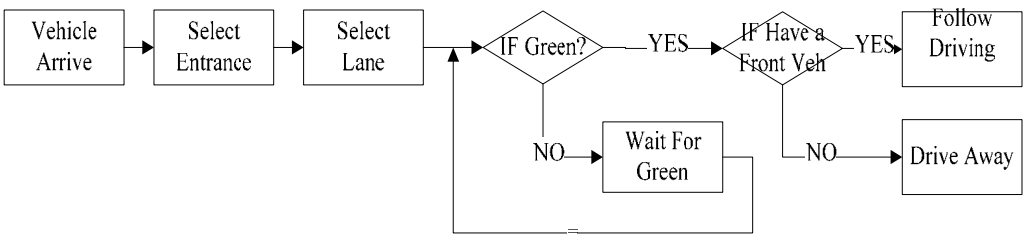


Figure2-1. Vehicle Process Analysis

the system objects: Vehicle Generator, Vehicle Controller, Lane Link, Signal Lamp, Signal Controller, Vehicle, Lane.

D. Relation of Objects Analysis

When the object analysis has completed, we start to analyze the relation of these objects. See Figure2-2, the number is the information code transferred between objects, when we encode in Arena.

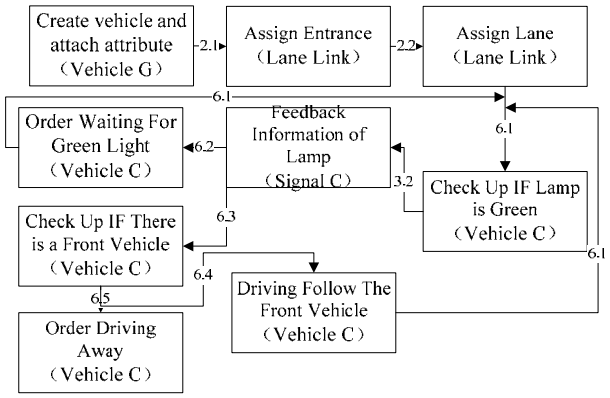


Figure2-2. Relation of Objects

E. Input and Output Parameters of the Model

Considering the requirement of the system, as well as the data collected, we confirmed the under parameters as the input parameters. See Table2-1.

TABLE2-1 INPUT PARAMETERS

Type	Input Parameters
Vehicle	*Arrival Distribution*、*Destination*
Lane Link	*Number of Entrances and Lane*
Signal Controller	*Signal Control Plans* *Number of Signal*
Lane	*Top Speed, Capacity of Lane* *Length of Lane*

Considering the aim of the research, we confirmed the under parameters as the output parameters. See Table2-2.

TABLE2-2 OUTPUT PARAMETERS

Type	Output Parameters
Vehicle	*Max/Min/Mean Travel time* *Max/Min/ Mean Delay* *Parking Rate*
System Structure	*Max/Min/ Mean Queue Length* *Traffic Capacity* *Number of Vehicles In Lane*

III. EXPLAINING OF MODULES IN ARENA

A. Module of Vehicle Generator

Module of Vehicle Generator is used to create vehicle and attach attribute, such as arrival time, type, destination and color. This module is the basic module of the simulation model. The arrival time of vehicles is a random event, so the time interval between two adjacent vehicles (time headway) is also a random variable. When time headway accords with Poisson distribution, we usual use negative exponential distribution to describe the time headway. The specific form is as follows:

$$F(x)=\begin{cases}1-e^{-\lambda x}(x\geq 0)\\ 0(\lambda \prec 0)\end{cases}$$

λ —Traffic Flow Rate.

B. Module of Vehicle Controller

Module of Vehicle Controller is used to control the driving of the vehicle, and this module is the heart module of the simulation model. It decides the overall process of vehicle's parking, freed driving and follow driving.[5][6]

- Parking

If the signal lamp turns red, yellow or there is a vehicle ahead, then stop.

- Free driving

If the distance between a vehicle and the front vehicle surpasses some dmax, we will consider the front vehicle do not impact on the following vehicle, and the vehicle drive away from the intersection as normal speed.

- Follow driving

If the distance is less than dmax, then the front vehicle will impact the speed of the follow vehicle. The following vehicle module is the most important dynamic model of a Traffic simulation model. This text uses the new line following model which is brought forward by Hglly. It considers the influence of the speed of the two front vehicle. The model as follow:

$$Xn(t)=C_1\Delta x(t-T)+C_2(\Delta x(t-T)-Dn(t)$$

$$Dn(t)=\alpha+\beta x(t-T)+\gamma Xn(t-T)$$

D(t)—expected speed of following drive

C1—Based on the survey of action of 14 drivers, when correlation coefficient greater then 0.8, the numeric area of T is 0.5～2.2s, then C1 is 0.17～1.3.

C2—To make the front and the following vehicle have the same acceleration by setting △v and △x.

At last, we will obtain the final formula:

$$x=0.5\Delta x(t-T)+0.125(\Delta x(-0.5)-Dn(t))$$

$$Dn(t)=20+x(t-0.5)$$

C. Module of Lane

This module is a carrier which connects the origin and destination. It is the important component of the simulation model. Its function is offering the space for the vehicle.

D. Module of Lane Link

In order to guarantee that we can build the model in Arena, we design the module of lane link. Its function is making vehicle arrive at the right lane with different destination.

TABLE2-1 PHASE STEP

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1G		—	—	---														
1Y					—				—									
1R	—					—				—	—	—	—	—	—	—	—	—
1A							—	---										
1PG		—	---	---														
1PR	—				—	—	—	—	—	—	—	—	—	—	—	—	—	—
2G											—	—	---					
2Y														—				—
2R	—	—	—	—	—	—	—	—	—	—					—			
2A																—	---	
2PG											—	---	---					
2PR	—	—	—	—	—	—	—	—	—	—				—	—	—	—	—
Time	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Phase																		

— means switched on, --- means blink, blank means turn off, unit of time is s.

1G—green light of E-W
1Y—yellow light of E-W
1R—red light of E-W
1A—green light of left turn of E-W
1PG—green light of pedestrian of E-W
1PR—red light of pedestrian of E-W
2G—green light of N-S
2Y—yellow light of N-S
2R—red light of N-S
2A—green light of left turn of N-S
2PG—green light of pedestrian of N-S
2PR—red light of pedestrian of N-S

E. Module of Signal Controller

This module is used to describe dynamic entity which is related to the road. It can produce light signal and restrict the action of vehicle, which is driven by system clock. We transform Signal Control Plans to Phase Step Table at the progress of encoding this module in Arena. Consequently, when input the parameter, we should also do this thing. Take four-phase (Dual Left Turn) for example, Its Phase Step Table as follow, See Table3-1.

IV. CASE ANALYSIS

A. Basic Condition

Collect the basic data of a intersection from Wuhan, its Hourly traffic volume as follow, look at Table 4-1, unit is pcu/h. S- moving straight, L-turn left, R-turn right.

TABLE2-3 HOURLY TRAFFIC VOLUME

		FLOW	LANE			FLOW	LANE
W	**S**	545	2	**N**	**S**	476	2
	L	104	1		**L**	41	1
	R	54	1		**R**	53	1
		FLOW	**LANE**			**FLOW**	**LANE**
E	**S**	564	2	**S**	**S**	558	2
	L	177	1		**L**	54	1
	R	109	1		**R**	51	1

B. Parameter Input

- Input of Arriving Function

We can confirm the arriving function by using the function of Input Analyzer in the Arena. After analyzing, we can obtain the parameter of arriving function:

$\lambda d = 4.23$， $\lambda x = 6.38$， $\lambda n = 5.39$， $\lambda b = 6.31$.

- Input of Signal Control Plans

There are two programs, one's Cycle time is 60s, the other is 100s. See Table4-2.

TABLE2-4 INPUT OF PHASE STEP

	STEP	1	2	3	4	5	6	7	8	9
一	**TIME**	2	3	3	2	2	15	3	3	2
二	**TIME**	2	5	3	2	2	21	3	3	2

TABLE2-5 INPUT OF PHASE STEP

	STEP	10	11	12	13	14	15	16	17	18
一	**TIME**	2	15	3	2	2	13	3	3	2
二	**TIME**	2	21	3	2	2	19	3	3	2

C. Simulation Result Analysis

Working procedure after input parameter, by 10 times simulation, we can obtain the result as Figure4-1. From

this figure, it is obvious that the programI is better than programII.

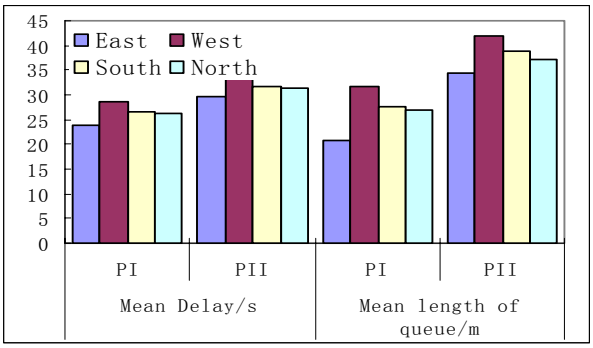


Figure4-1 Result Analysis

V. CONCLUSION

By using Arena's modeling and simulation method, special modules of signalized intersection are built in the discrete environment of modeling and simulation of Arena. It is concluded that these modules can be used expediently to study on simulation modeling, scheme adjustment and result analysis of a certain intersection. However, there still might be some other aspects that we hope to do more research on this topic. One thing is that the paper only takes one intersection into consideration, other than several adjacent intersections on main traffic flow which could possibly have more complicated scenarios. That is also one of our research focuses in next step.

REFERENCES

[1] W.David Kelton, "Simulation With Arena," M.The McGraw-Hill Companies, In USA,2002.

[2] Zaibao Guan, "The Research of Algorithm and Model of Traffic Control at Signaled Intersection," D.Chengdu, Sichuan Province,Southwest Jiaotong University,2007.

[3] Huapu Lu, "Modern Management of urban traffic,"China Communications Press, 1999.

[4] JixiuHao, and ZhaoruiZheng,"Traffic Simulation Based on Object-Oriented Method,"J.Journal of Taiyuan University Of Technology,2004,35(4).

[5] Yan Feng, Yulong Pei,and Chenghai Cao, "Application of VISSIM in the Public Transportation Priority Signal Timing," Journal of HARBN institute of Technology,2007.

[6] Jianhe sha, "A Microscopic Traffic Simulation System Based on MAS Theory,"D.Shandong University, Jinan,Shandong Province,2007.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 208-211

The Application of Data Mining Technology in the Intrusion Detection System

Zongpu Jia[1], Shichao Jin[2]
[1]Computer Science and Technology Department, He Nan Polytechnic University, JiaoZuo, China
E-mail: jiazp@hpu.edu.cn
[2]Computer Science and Technology Department, He Nan Polytechnic University, JiaoZuo, China
E-mail: swaybottle@sina.com

***Abstract*—This paper analyzes the current situation of the intrusion detection system, which is the basis to put forward that data mining technology is to be applied to the intrusion detection system in terms of the problems of the traditional intrusion detection system. Meanwhile, the paper designs the intrusion detection model of data mining. With the study on intrusion detection and data mining, the algorithm of classical relation with clustering in the characteristics on intrusion detection system is improved and optimized.**

***Index Terms*—Intrusion Detection; Data Mining; Error Detection (Abnormal Detection); Misuse Detection; Relation Rule; Sequence Rule; Clustering Algorithm**

I. FOREWORD

With the development of internet technology, more and more people have got the abundant network resources to learn the various patterns of network attack and may implement the seriously destructive attack only with the simple operation. So how to detect and prevent the invasive behaviors has become the highlight of computer field.

There are plenty of methods to strengthen the network security such as setting secret code, VPN and firewall. But most them are static and can't make the effective protection. However, the intrusion detection technology is a dynamic protective strategy, which can make monitor, attack and counterattack on network security to make up the weakness of the traditional static strategy.

II. THE INTRODUCTION OF INTRUSION DETECTION TECHNOLOGY

The intrusion detection technology monitors the operation state of network system and digs out each attacking attempt, attacking behavior and attacking result so as to ensure the confidentiality, integrity and usability of system resource. The intrusion detection system can be classified into the diverse patterns of being based on mainframe, being based on network, being based on kernel and being based on application. The paper mainly analyzes the structure of the intrusion detection system based on network. The intrusion detection system can be divided into two sorts according to the differences of the data analysis methods.

The first one is misuse detection, also named detection based on quality, which is to establish a quality-base in terms of a known invasive behavior to match the motion having occurred. When the consistent result is got, the invasive behavior is surely proved. The merit of misuse detection is that it has got the low misinformation rate. However, the quality-base will become larger and larger because of too many invasive behaviors and can only detect the known invasive behaviors.

The second one is anomaly detection, also named detection based on behavior, which is to establish a normal quality-base and judge the invasion according to the user's behavior or the consumption of resource. The merit of anomaly detection is that it has the strong currency and little relation to system and can detect the attacking methods that have never appeared before. However, it still has the high misinformation rate because of the impossibility of the produced outline to comprehensively describe all users' behaviors of the whole system as well as the variable behaviors of each user.

Therefore, the combination of the two methods can obtain the better performance. Anomaly detection can make the system detect the new, unknown and other situations; misuse detection can protect the integrity of anomaly detection by means of preventing the alteration of behavior patterns that some patient hackers may use to make anomaly detection consider it legal.

The data origin of intrusion detection can be obtained by some specialized wire-shark, for example, Winpcap is generally used to obtain data packet in the system of Windows, while Tcpdump and Arpwatch may be used in the system of Unix. Data mining technology is mainly introduced and will be used in the data analysis phase. The response consists of the active response and the passive response.

III. THE INDUCTION OF DATA MINING TECHNOLOGY

The mission of invasion analysis is to find the invasion trace among the numerous obtained data. The network invasion is judged when the obtained data is input into the detecting system as information and analyzed and disposed by the detecting system. It is a huge intellectual project to establish a rule-base (quality-base), because the current detecting rules are generally made by handcraft,

[1] Supported by the Doctor Fund of Henan Polytechnic University（NO.B2009-21）

especially the detecting knowledge of judging the invasion behavior, which is a series of deduction rules coming from the security experts' analysis experience on the suspicious behaviors to obtain the invasion quality so as to compile IDS. The IDS of this sort has got the obvious fault that the experts need to summarize and obtain the invasion quality continually so as to provide the comprehensive and abundant invasion detecting rule. It proves that the IDS can passively detect the known invasion or attack behaviors only with the outer aid but can't detect the variants or unknown invasion or attack behaviors. Besides, the reliability of IDS is not steady because of the man-made analysis and experience and the misinformation rate is high.

In order to overcome the current limitations of IDS, a new sort of technology---data mining technology is supposed to be applied. Data mining is the process of mining the unknown knowledge useful to decision from the numerable data sets. The clustering analysis algorithm of data mining is suitable for constructing the normal behavior model of network from the numerous data packets, while the relation analysis algorithm is suitable for describing the relation rules of invasion behavior pattern with which invasion detecting is carried out.

Ⅳ. The Realization of Invasion Detecting System Based on Data Mining

The invasion detecting system in the paper consists of the pre-processor of data, data partition, relation rule obtaining, rule base and data analysis, and so on. The flow chart is as the following.

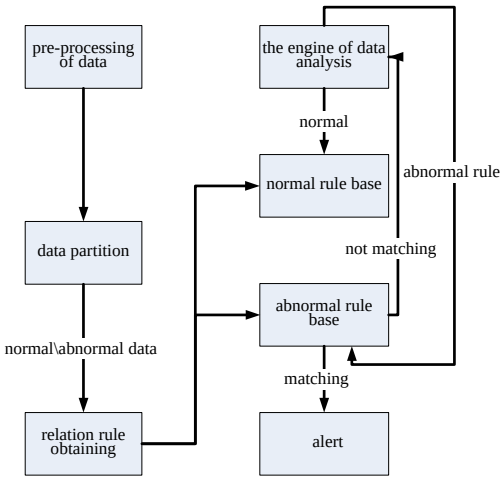


Figure 1. The Pattern of Invasion Detecting System Based on Data Mining

The pre-processor of data is in charge of selecting the reasonable property from the data stream such as the system log data and network data packets to create the data form fit for detecting pattern; data partition partitions the data into the abnormal or normal data set with clustering algorithm. The module of relation rule obtaining makes the quality obtaining of each data set, then creates the rule sets and the rule bases describing the current each data set to store the normal and abnormal rules respectively. Whether the invasion occurs is decided by calculating the matching degree between the rule of the current data set and the rule of the rule bases. After the rule of the unsuccessful matching is input into the data analysis engine which makes analysis and judgment, the correspondent normal and abnormal rules are formed.

Ⅴ. The Crucial Technology Analysis

A. Data Partition

The partition of data is to accomplish the following work. That is to form numerous data sets accurately, to make each data set comprise just the normal data or abnormal data, to judge whether each data set is the normal data set or the abnormal data set and utterly to favor the relation rule obtaining.

The clustering algorithm is a common sort of technology of data mining, the core idea of which is to gather the similar (or close) data into a cluster and make the same patterns of data close but the different patterns of data distant. Here the improved clustering algorithm of K-means and the data partition in advance is taken.

The disadvantages of the traditional algorithm of K-means include:

(1) The amount K of the ultimate clustering is supposed to be confirmed in advance and the record of the same amount should be designated as the initial clustering center. Then the whole record set is to be scanned by times, the clustering center and the clustering that the record belongs to are to be altered continually before the steadiness of the clustering center. The result of clustering has a direct and close relation to the amount K of clustering. The different clustering amounts will bring the different clustering results, so it is very hard to confirm the clustering amount bringing the best clustering result.

(2) Some of the clustering results are likely to be vacancy, that is, the vacancy clustering may be brought without any object similar to the clustering center of such clustering.

(3) The initial clustering center at random is likely to be not the best one at the beginning of the clustering and the clustering result is easily influenced by the initial clustering center. The traditional algorithm of K-means takes the arbitrary selection when the initial clustering center is obtained. However, obtaining a better initial clustering center can obtain a better clustering result.

The schedule to improve of the algorithm of K-means:

The improvement is made according to the disadvantages of the algorithm of K-means, that is, the two clustering parameters of the clustering semi-diameter and the nearest value H are to be added. The concrete method is the following. It is to calculate the smallest value of the distance of the current record from all the clustering centers; the record will be regarded as the new clustering center if the smallest value is more than the clustering semi-diameter. The final clustering amount K will be the best clustering amount of the data set instead of the beforehand clustering amount. It is not necessary for this method to confirm the clustering amount in

advance, which can sort out the records of the contiguous distances into the same clustering and isolate the abnormal records of great distances from the other clustering. The method can adjust the clustering amounts in a certain scope and sort out the abnormal records into an individual clustering, which is favorable for marking the abnormal records and cuts down the influences on calculating the clustering center. When there is vacancy clustering in the clustering results, the farthest object from the clustering center will be removed from the current clustering to bring a new clustering center so as to replace the vacancy clustering with the newborn clustering.

For the selection of the initial clustering center, the sample set of the amount T should be taken first and the initial center set of the amount T will be produced after the clustering of K-means on each sample set. Then all the elements of T×K are clustered with the algorithm of K-means with C1 being the initial clustering center and the clustering center sets of T will be obtained, then the best one will be selected as the ultimate initial clustering center.

The advantage of this method is putting forward a method to automatically select the initial clustering center, which reduces the complexity of algorithm time by means of selecting sample instead of the whole data set and can avoid the influence of "the isolated point" by means of the beforehand initial center clustering and multiple sample sets so as to improve the representation of the initial center.

Improved k-means algorithm is described as follows: Input: a database containing n items of data

Input parameters: the initial number of clusters M; cluster radius r; nearest neighbor threshold h

Output: k a cluster

(1) Select the M were the best initial cluster centers (w1, w2, …, wm) wj = xi, where, j∈{1..k},i∈{1…n};

(2) to correspond to each cluster cj and wi.

(3) Calculate the other records xi (i (1 ... n)) to the minimum distance from cluster center min.

(4) If min <r, will be assigned to the nearest wj * xi where cluster Cj *; that | xi-wj * | <= | xi-wj | mj = (1 .. k) Otherwise, create a new cluster, the xi as a new cluster center.

(5) Back (2), until all records are complete.

(6) to each cluster mean replace the original cluster center, namely:

(7) If there is an empty cluster, the furthest point away from the cluster center out to create a new cluster where the cluster center, the new cluster created to replace the empty cluster.

(8) until the same value until the cluster centers.

(9) Finally, calculate the value of all the cluster centers for any two centers The distance between the nearest neighbor with threshold h are compared, if the distance is less than h is to merge these two cluster.

(10) repeat (9), until the distance between any two cluster centers are greater than the value of h up

B. The Birth of Relation Rule

The normal and abnormal data are partitioned into different data sets to obtain after the clustering partition of the data, so the influence of the too high degree of minimum support is reduced so as to be favorable for the relation rule obtaining.

The mining process of relation rule consists of two steps.

(1) The first one is to find out all the frequent item sets, which are used to find out all the item sets whose support degree is not less than the received minimum support value.

(2) The second one is to bring the forceful relation rule by the frequent item sets, which must satisfy the minimum support and the minimum confidentiality. In the process of mining, the first step, which decides the general quality of mining relation rule, is the core of the relation rule finding algorithm.

By means of the mutuality between the appointed min_sup and min_conf and the hunting algorithm of the frequent item sets as well as the mutuality with the relation rule sets, the users make explanations and evaluations on the mining results. This experiment module realizes the relation analysis with the improved algorithm of Apriori.

Because the traditional algorithm of Apriori needs to scan the whole data base in the process of producing the frequent item sets, a back-up set needs to be formed before the formation of each item set K, which will be the bottleneck of algorithm efficiency when there is a big amount of data. The rule amount mined by the algorithm of Apriori is large. It is reasonable to get informed of the original IP address to have analysis and judgments for the record data of network link, so it is not favorable for the analysis of invasion behavior when the received rule sets include no rules of the original IP address and that should be excluded.

The improved algorithm of Apriori consists of two parts which are bringing the frequent item sets and bringing the relation rule. The process of bringing the frequent item sets comprises linking, which is used to bring the back-up item set, and pruning, which excludes part of the item set elements by means of minimum support. The process of bringing the relation rule is to obtain the relation rule sets with the following formula.

$$\frac{count(s)}{count(l)} \geq \min\_counf \tag{1}$$

With the above considerations, the improved project of the algorithm of Apriori is put forward.

(1) the simultaneous linking and pruning

(2) The core quality parameter P is supposed to be introduced that means the parameter of the original IP address. In the process of bringing the relation rule, it is necessary to check whether there is the original IP address in the rule. Otherwise, the rule bringing should be quit.

C. The General Quality Detecting

We downloaded 3 group tcpdamp form of network traffic data from the site, base on behalf of the normal state of network traffic, netl is the network traffic that includes analog IP Spoofing attacks, In the data file, an intruder is trying to guess the serial number to IP Gain access to the remote host, net3 is included simulated port scanning Describing the attack traffic .in the data file, the intruder tried to collect information about Web hosts and the services provided information. Experimental data can be easily obtained from the Internet, data is representative. Experiments show: Using the improved Apriori algorithm can improve the speed, the resulting rule set is also smaller.

KDD Cup 99 data set includes four major types of attacks, DOS attacks and port scanning attacks PROBE accounted for more than 80%, respectively, for detection of these two attacks. Select the data in the 15 key numeric attribute clustering, used in the clustering process does not record the type of identification, clustering the data set results can be clustered into different categories. Makes the exception classes and normal classes of data separately. Can be found through the experiment, take different number of clusters a great impact on the results, but can not predict the optimal number of clusters. The improved k-means clustering algorithm can solve the initial issue of the number of difficult choices. Clustering algorithm is introduced and the most close to the radius threshold, the number of clusters does not require pre-input data can be clustered into the best number. Table 1 shows the parameters of the DOS attacks on the introduction of test results. Experiments show that: the improved k-means algorithm DOS attack detection rate significantly increased, better clustering algorithm, and do not need to specify the final number of clusters in advance, you can change most near the threshold to control the particle clustering degree.

TABLE I. THE PARAMETERS OF THE DOS ATTACKS ON THE INTRODUCTION OF TEST RESULTS

M	h	r	The final number of clusters	Detection rate（%）	Error rate（%）
2	7	8	37	99.27	7.18
2	7	10	31	94.58	5.67
2	7	15	20	85.36	3.92
2	7	20	16	82.7	1.4

VI. CONCLUSION

Invasion detection has been developed fast in the recent years, which is the second security valve after the firewall as a kind of active detecting method of network security. The paper starts from the basic definitions, introduces invasion detection as well as several data mining methods commonly used for invasion detection and theoretically states the module of the invasion detection system based on data mining technology. The paper realizes a mining module based the mainframe log data. Besides, the abnormal visiting relation rule is brought by mining the log files of IIS so as to serve the invasion detection system.

There are various patterns of technology used for invasion detection. The technology of invasion detection based on data mining technology has become the highlight of the current development of invasion detection. However, data mining is still in the developing phase, so it is essential to make the deep research on data mining.

REFERENCES

[1] Zhang Yinkui, Liaoli, Songjun. “The Principle of Data Mining” [M] Beijing, Mechanism Industry Press. 2003 : 93-105.

[2] Dai Yingxia, Lian Yifeng, Wanghang. “System Security and Invasion Detection” [M] Beijing, Tsinghua University Press. 2002 : 99-137.

[3] Liu Guiqing, Zou Lidi, Likai. “The Analysis Methods of Data Mining in Invasion Detection” [J] Academic Journal of Hefei College (for Natural Science). 2004，14(13)：26—29.

[4] Liushen, Zhang Yongping, Wan Yanli. “The Application Analysis and Improvement of Decision Tree Algorithm in Invasion Detection” [J] Computer Engineering and Design. 2006

[5] Zhang Hanfan. “The Invasion Detection System Based on Data Mining”. Nanjing University of Technology. 2004

[6] Xiangji, Gaoneng, Jing Jiwu. “The Application of Clustering Algorithm to Network Invasion Detection”. [J] Computer Engineering. 2003 , 29(16): 1-3.

[7] Tanyong, Rong Qiusheng. “The Realization of Classification Algorithm Based on SLIQ”.[J] Computer Engineering. 2003 , 29(18):1-3.

[8] Moore D , Voelker G, Savage S. Inferring Internet denial of service activity [A] . In : Proceedings of the 10th USENIX Security Symposium [C] . Washington DC , 2001. 9 22.

[9] Lemon J . Resisting SYN flooding DoS attacks with a SYN cache [A] . In : Proceedings of USENIX BSDConp2002 [C] . San Francisco , 2002. 89 97.

[10] Bernstein D J . SYN cookies [EB/ OL] . http : / / cr. yp. to/ syncookies. html. 2000/ 2003207208.

[11] Check Point Software Technologies Ltd. SynDefender [EB/ OL] . http : / / www. checkpoint . com/ products/protect/ firewall21. html. 2002/ 2003207208.

[12] Netscreen Technologies Ltd. Firewall apliance[EB/ OL] . http : / / www. netscreen. com/ . 2002/ 2003207208.

[13] Schuba C L , Krsul I V , Kuhn M G, et al. Analysis of adenial of service attack on TCP [A] . In : Proceedings of IEEE Symposium on Security and Privacy [C] . Los Alamitos: IEEE Computer Society Press , 2007. 208-223.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 212-214

A Novel Anti-collision Backtracking Algorithm Based on Binary-tree Search in UHF

Wang Jianfang
School of Computer Science and Technology,Henan Polytechnic University,Jiaozuo, P. R. China
Email:wangjianfang2006@gmail.com

***Abstract*—To anti-collision problem of multiple tags of RFID in ETC (Electronic Toll Collection) system, a novel BBS (Backtracking Binary-tree Search) anti-collision algorithm is proposed in working in UHF (5.8G) based on the DBS (Dynamic Binary-tree Search) and RSBS (Random Split Binary-tree Search) to solve the only one bit collision in inquiry. The BBS algorithm can reduce twice inquiry process. The simulation in UHF (5.8G) show that the BBS algorithm can reduce the number of inquires and information sent to improve the ETC system efficiency.**

***Index Terms*—Tag, Reader, Binary-tree Search, RFID**

I. INTRODUCTION

RFID (Radio frequency identification) is the identification technology using radio wave which is the combination of wireless communication technique and semiconductor technique. The major problems of the RFID are included as follows: There are two or more tags to return information to the reader at the same time, collision will occur when multi-tags exist within the range of the reader signal scope, the phenomenon are known as tags collision. The anti-collision algorithm mainly include the time slot ALOHA anti-collision algorithm and binary tree anti-collision algorithm [1][2].

Using the vehicle RFID technology and the path-identifying stations method, the ETC (Electronic Toll Collection) system can be changed into the systems based on the mobile communications tolling systems' model for the need of the highway management. With applications in UHF (2.45G/5.8G/13.6G), and increasing number of the tags, the ALOHA algorithm cannot achieve a higher recognition rate and more exact efficiency. Especially the 5.8G in UHF is used for the frequency band of the ETC system by the ISO, so more and more BS (Binary-tree Search) algorithm is used to solving the tags collision.

The BS algorithm is a deterministic algorithm and has higher recognition rates. In theory, there is no recognition omission. In the process of anti-collision algorithm to BS, the tags always full character code as a response. But in practice, the character code of the tags may be very long, such as the code length of UID (Ubiquitous Identifications) is 128 bit, and can be expanded to the 256 bit, 384 bit or 512 bit according to requirement. Thus lead to a significant amount of data, the recognition speed is influenced by the code length.

The DBS (Dynamic Binary-tree Search) algorithm [3] reduces the time consuming of transferring the character code in ETC system to improve the ETC system efficiency. The DBS algorithm is proposed as the classical anti-collision algorithm owing to the Simple ideas, stable system performance and less system resource requirements in ISO/IEC14-3A. There are some problems in the DBS algorithm, such as all of the tags which are not to be activated compare their own to judge whether met to request command. The process is obviously a waste; meanwhile the tags which should be removing of the request from the scope also increase the interference within the system

The tags should have a counter and a 0 or 1 random number generator in RSBS (Random Split Binary-tree Search) algorithm [4]. The reader active the dormant tags, then select all or part of the tags to recognition flow by Select or Unselect command.

In order to take full advantage of the collision information has been obtained. Based on DBS and RSBS algorithm, a novel BBS (Backtracking Binary-tree Search) anti-collision algorithm is proposed to solve the tags collision problem in ETC system.

II. BBS ALGORITHM

A. Define tag states

The BBS idea is when there is some collision between tags, the reader firstly search the one path until a tag could be recognized; and then upward to the backtracking search from the lowest collision bit. If there are collisions; the collision bits are no less than one bit, and then downward search. If there are not collisions and then backtrack to the place of the second low collision, and then upward backtracking by bit-by-bit, until each tags are recognized. If only one bit collision has happen during inquiry show that only one tag need to be recognized, then the system no longer corresponds to the collision bit and directly select and read these two tags. So similar to each of these pairs of tags, the system can reduce twice inquiry process.

For implement the BBS algorithm, the tags need to add the dormant depth counter need, meanwhile the reader need to add the collision bit counter and jump flag bit. The dormant depth counter can implement the backtracking. The collision bit counter can judge whether only a collision bit occur. The jump flag completed secondary the process that read the two tags of only bit collision to reduce the number of inquires and information sent to improve the ETC system efficiency.

Define three states of the tags:

(1) Activation state;

(2) Dormant state;

(3) Quiet state.

The dormant state and the dormant depth counter work together to complete the collision record and reduce the impact of information search.

B. BBS algorithm steps

The anti-collision demand is as follows:

(1) Request(x,m): x is 0 or 1, m is the highest bit of collision detected, the reader sent the command to active the tags in the region. The tag of the activation state detect own m-bit tag number to compare the *x*, if the results is the same, then answer; if not, the tag enter the dormant state, and the corresponding dormant depth counter is set 1. If the tag is the dormant state, then the dormant depth counter adds 1. When m is all, no matter what the value of *x*, each bit is compulsory compared with 1.

(2) Active: Its role is to reactivate a dormant state tag. Only dormant tags can response to the command. The dormant depth counters minus 1 to be active in dormant state. If the dormant depth counter is 0, the tag became the activation state and can response the Request.

(3) Quiet: The flag bit of the read tags became 1 and the tags enter into the Quiet state no longer respond to Request command.

(4) Select: When the reader determines collision-bit counter is 0, the reading flag of tags is set 1. The tags of active state and only one collision is no more inquire when the collision counter is 1, and directly set the read flag which the collision bit is 0 to 1.

(5) Read-Data: When the read flag of tags is 1 and the reader counter is 0, the reader begins read the information and the read flag of read is reset. If the read flag is 1 and the collision counter is 1, then the reader begins read information and the read flag of read is reset and make the flag jump into the 1.

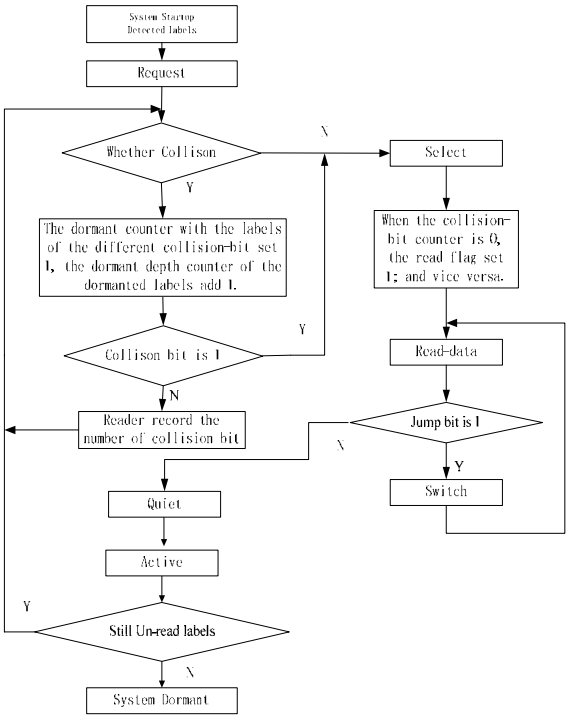


Fig.1 The flow of BBS algorithm

(6) Switch: When the jump flag is 1, the read flag which the collision-bit is 1 and is no Quiet tag is set 1, and the read collision is became 0 and the jump flag rest.

The BBS algorithm flow show in Fig.1.

C. Case Analysis

As shown in Fig.2, the BBS algorithm could be represented by the binary tree structure. We can see form Fig. 2 if the 5 tags are recognized; the 4 child nodes follow the root node. The parent-child nodes can bi-directional search, the tags of the only one bit collision read directly left and right *n* odes, Therefore the total number of searches is as follows:

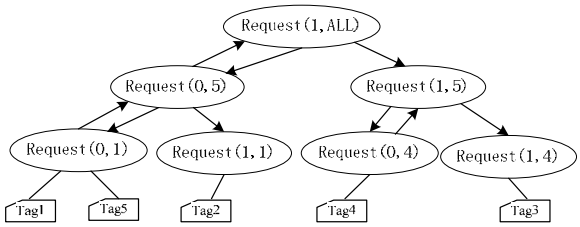


Fig. 2 BBS process diagram

S(m)：(m－1) ×2 + 1－2n = 2(m－n)－1

So based on the BBS algorithm, the reader identify m tags, the number of the search is as follow:

S(5)=(5－1)× 2 + 1－2 = 7

2n is that reader need not to inquire the tags can read directly when minimal un-continuous collision-bit is 1 bit.

The BBS algorithm character is as follows:

As long as within a range of values, the dormant depth counter can accurately identify each tag, the recognition can reach to 100% in theory.

The BBS algorithm draws the method of the RFBA (Random Fork Binary-tree Anti-collision) algorithm, and improved the RFBA algorithm and shortened the length of sending information and reduced the number of inquires and system overhead.

The pairs of the tags that minimal un-continuous collision-bit is 1 bit appear randomly, so system overhead depends on the number of a collisions and the tags themselves character code. The system efficiency with the increase in the number of tags is discrete, but the 2n-the pairs of the tags that minimal un-continuous collision-bit is 1 bi t- is always greater than zero. Therefore the system overhead of the BBS is certainly not greater than the DBT and BS.

III.SIMULATION

The anti-collision is to deal with the large number of a certain length of binary with serial number tags. If the number of tags with the same length is regarded as the matrix, the matrix row is the tags serial number; matrix column is the serial number tag values in a certain bit, the BBS simulation is processed by MATLAB [5].

For the tags number of choices we adopt the following principles: each additional 10 tags we select a observation points when the number of the tags below the 100, each

additional 20 tags we select a observation points when the number of the tags below the 400, each additional 50 tags we select a observation points when the number of the tags beyond the 400.

Through the BBS simulation, the ratio relations between the numbers of bits transmitted per tag and the number of the tags are shown in Fig.3. We can see as follows:

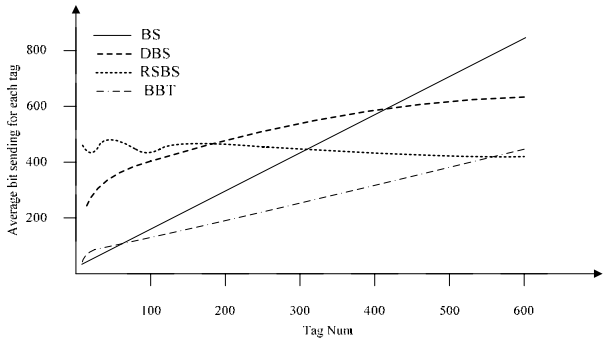


Fig.3 Relation between the ratio and tags number in kinds of algorithm

When tags number < 50, each tag of the average number of bits transmitted is the smallest number in BS.

When 50<tags number < 550, BBT is the relative slow linear increments, and its bit rate is the smallest change in all algorithm.

When tags number > 550, RSBS advantages in data transmission can be reflected in the Fig. 3.

By the relation between the tags number and the inquire number we can see as follows from the Fig.4:

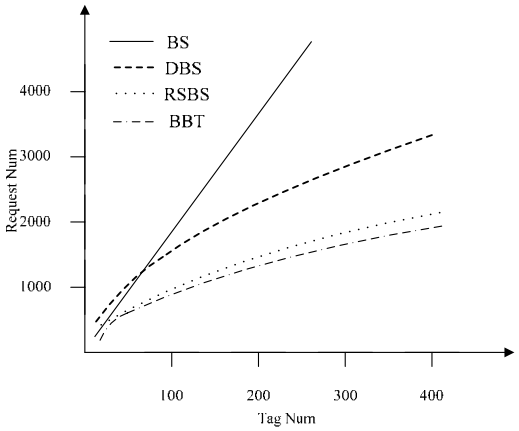


Fig. 4 Relation between the tags number and the inquire number

The number of inquire sharp increase and the process time rapid grow with the tags increase in BS. This is because BS is bit-by-bit algorithm. The branch that the tags are not assigned must to be searched. The invalid inquire can consume large amounts of system time, so reduce the system efficiency.

The number of inquire slowly increase in DBS, but the un-recognized tags need reply in every inquire process. After the two tags with only one bit collision still need send separately inquire command, the reader can send the select and read command.

The compare number of BBS is obviously less than the DBS. The backtracking can reduce the inquire number, if only one bit collision happen, the system no more send inquire command but read directly above two tags to the extent that reduce the inquire number.

IV. CONCLUSION

The BBS algorithm is proposed to solve the tags collision of RFID in ETC system. The BBS can reduce the number of inquires in UHF (5.8G). The simulation and compare results show the BBS can improve the system efficiency.

REFERENCES

[1] Liu L A, Lai S L. ALOHA-Based Anti-Collision Algorithms Used in RFID System[J]. Wireless Communication, 2006, 9:124-134.

[2] Lee S R, Joo S D, Lee C W. An Enhanced Dynamic Framed Slotted ALOHA Algorithm for RFID Tag Identification [J]. Mobile and Ubiquitous System, 2000, 7:166-172.

[3] Guo Z M, Hu B J. A Dynamic Bit Arbit ration Anti-Collision Algorithm for RFID System [J] . A nti-Counter feiting Security Identification, 2007, 9:457-2460.

[4] Myung J, LeeW, Srivastava J. Adaptive binary splitting for efficient RFID tag anti-collision [J]. IEEE Communications Letters, 2006, 10 (3):144 -146.

[5] Tsan P W. Enhanced binary search with cut through operation for anti-collision in RFID systems [J] .Communications Letters IEEE, 2009, 10 (4):236-238.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 215-218

The Implementation of Multi-Local LEACH Routing Algorithm Based on Wireless Sensor Networks

Qingpu Guo[1], Jun Li[2]
[1] Henan University of Economics and Law, Zhengzhou, China
Email: gqp@hnufe.edu.cn
[2] North China University of Water Resourcesand Electric Power
Email: lj@ncwu.edu.cn

***Abstract*—These LEACH is a widespread protocol in wireless senor networks to reduce the energy dissipation of wireless sensor system. However, through our analysis, we found there are still some limitations in this protocol. In this paper, we propose a revised LEACH algorithm called Multi-Local LEACH to address these problems, which incorporates multi-hop and internal rotation mechanism to the convention LEACH. We establish the simulation platform in a virtual wireless environment by matlab and simulate the operations of the two protocols. The result of experiments shows that Multi-Local LEACH can significantly extending the network lifetime and be superior in energy saving compared with conventional LEACH.**

***Index Terms*—LEACH; multi-hop; internal rotation; wireless sensor networks; simulation**

I. INTRODUCTION

The routing protocols of wireless sensor networks can be classified into two categories: flat routing protocol and tired routing protocol [1]. Nodes are peer to each other in flat routing protocol and they transfer data by multi-hop. However, as a majority of traffic will converge at a minority of nodes in flat routing protocol, this will lead to system performance degradation and drastic sensor lifetime reduction. Considering the energy consumption by data transmission is larger than that by data computation, wireless sensor networks are commonly using clustering-based tiered protocol to reduce data transmission. Compared with flat routing protocol, tiered protocol is more scalable, manageable and energy-efficient. LEACH [2](Low Energy Adaptive Clustering Hierarchy) protocol is a clustering-based tiered protocol for wireless sensor network which was firstly proposed by Hari Balakrishnan in 2000 [3]. For LEACH protocol, some of the sensor nodes are adaptively elected as cluster-heads to reduce the amount of information that must be transmitted. The other sensor nodes that are non-cluster-heads are in charge of collecting information and transferring them to the cluster-heads. The cluster-heads are responsible for performing data fusion in their own clusters. After data aggregation, they will transfer the data to the base station. Therefore, by using LEACH protocol, the energy consumption could be reduced compared with the way that all nodes directly transferring data to the based station. LEACH protocol works well at energy saving, however, it still has some drawbacks to overcome.

II. LEACH PROTOCOL OVERVIEW AND EXISTINGDRAWBACKS

In LEACH protocol, the cluster-heads are periodically and randomly elected. For each election, every node selects a random number between 0 and 1. If the number is less than a threshold T(n)[4], [5], [6], the corresponding node becomes a cluster-head at the current round. The threshold is set as:

$$T(n)=\begin{cases}\dfrac{P}{1-P\times(r \bmod \dfrac{1}{P})}, & n\in G\\ 0 & n\notin G\end{cases} \quad (1)$$

Where P represents the desired percentage of cluster-heads to all nodes, r represents the current round, and G is the set of nodes that has not been cluster-heads in last 1 P round. After a new cluster-head is determined randomly, it broadcasts this information to the rest of the nodes that notify them who is the new cluster-head. After this phase is complete, each non-cluster-head node selects a cluster joined for this round. This selection is based on the received signal strength of the advertisement. After each node has decided to which cluster it belongs, it must inform the cluster-head node that it will be a member of the cluster. In the set-up phase, cluster-header play a core role in the construction of a cluster[7]. Nodes of a cluster communicate with their cluster-head using different CDMA codes. In steady-state phase, nodes keep on collecting inspection data and transmitting them to their cluster-head. The cluster-head performs local data fusion to compress the amount of data being sent from the cluster to the base station. In order to minimize overhead of initiation, the steady-state phase is longer compared to the set-up phase[8].

III. MULTI-LOCAL LEACH ROUTING ALGORITHM DESCRIPTION

A. Overview of Multi-Local LEACH Routing Algorithm

Since LEACH employs direct communication between cluster-heads and base station, the nodes furthest from the

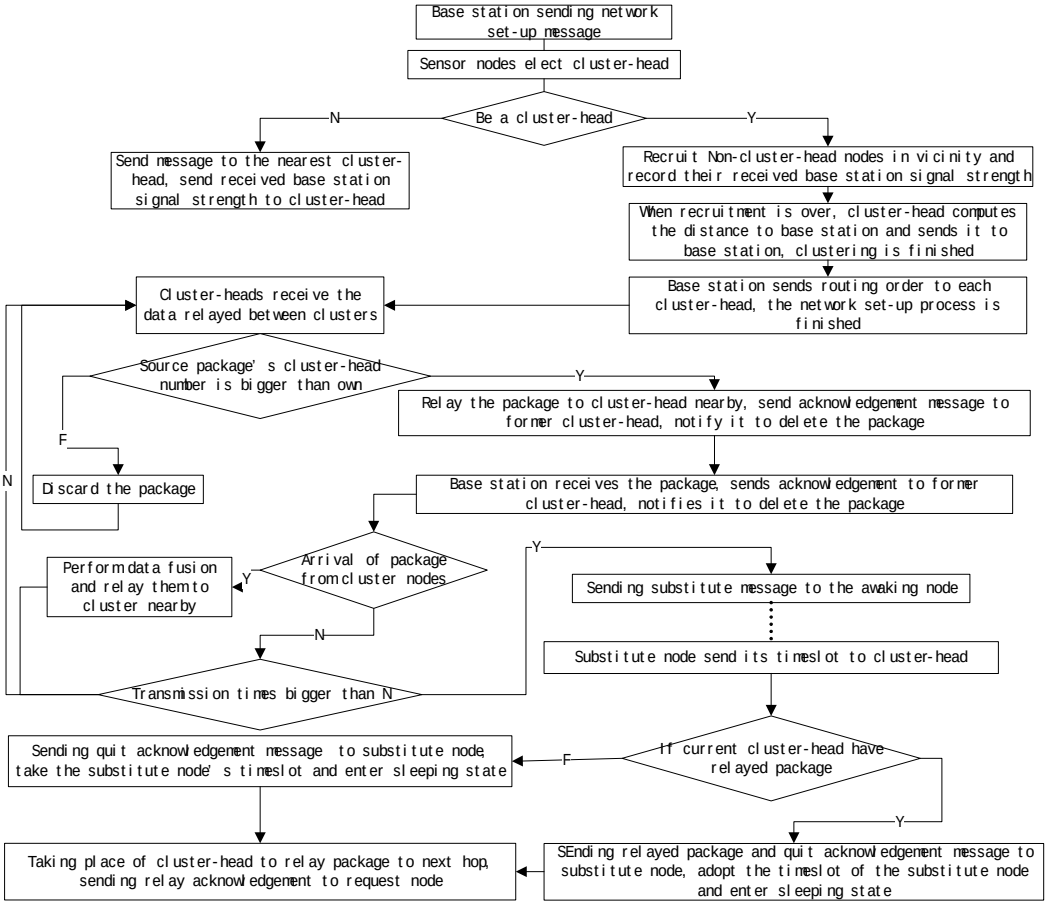


Fig. 1. Flowchart of Multi-Local LEACH algorithm

base station have the largest transmit energy consumption compared with the nodes close to the base station. Consequently, the nodes far form the base station will have less lifetime than that of the nodes close to the base station. In addition, the election of cluster-heads is performed in entire WSN, which leading to large amount of energy dissipation and network lifetime reduction. To address these problems of LEACH, we propose the Multi-Local LEACH routing algorithm. Compared with conventional LEACH algorithm, multi-local LEACH provides several advantages: 1, Multi-Local LEACH routing protocol only uses conventional LEACH algorithm once to organize clusters at the set-up phase so as to reduce the cluster organization times, by which will significantly alleviate the energy dissipation of network set-up. Consequent reorganization of network will be limited inside clusters; 2, In the process of network set-up, meanwhile, cluster-heads record each cluster nodes' signal strength that received from the base station. With the aid of mean signal strength, the distance between cluster and base station will be obtained. The base station relies on this value to determine the inter-cluster data forwarding mechanism. As a result of the first rotation mechanism within the cluster, flat routing protocol becomes an energy-efficient factor, eliminating the disadvantages of single hop data transmission in conventional LEACH algorithm; 3, Multi-Local LEACH provides a simple and practical internal rotation mechanism of cluster-head inside cluster nodes, which could basically address the drawback of excessive node energy dissipation caused by fixed cluster-heads.

as $RSSI_{average}$. By the formula $RSSI_{average} = -(10 \times n \times \log_{10} d + A)$,we can get d which is the distance to the base station and transmit the value to the base station. In this formula, n is the signal transmission exponent, d is the

B. Implementation Steps of Revised Algorithm

Base on the analysis of energy dissipation of network set-up, the complementation between flat routing protocol and internal rotation mechanism, and the constrains of communication between clusters and base station, we propose an optimized Multi-Local LEACH algorithm whose flowchart is shown in Figure 1.

1) The Set-up Phase of Network: Networking construction signal is broadcasted by the base station. When each node receives this signal, it firstly record the RSSI value of the signal's strength. After that, it enters the process of electing itself to be a cluster-head. When being elected to be a cluster-head, the node begins to recruit cluster member from other sensor nodes in the vicinity and calculate total signal strength of all the cluster members. The nodes that haven't been organized to a cluster need to repeat the reconstruction process using LEACH algorithm until everyone are added to a cluster. After that, cluster-heads record the sum of signal strength of base station as $RSSI_{total}$ and its average value

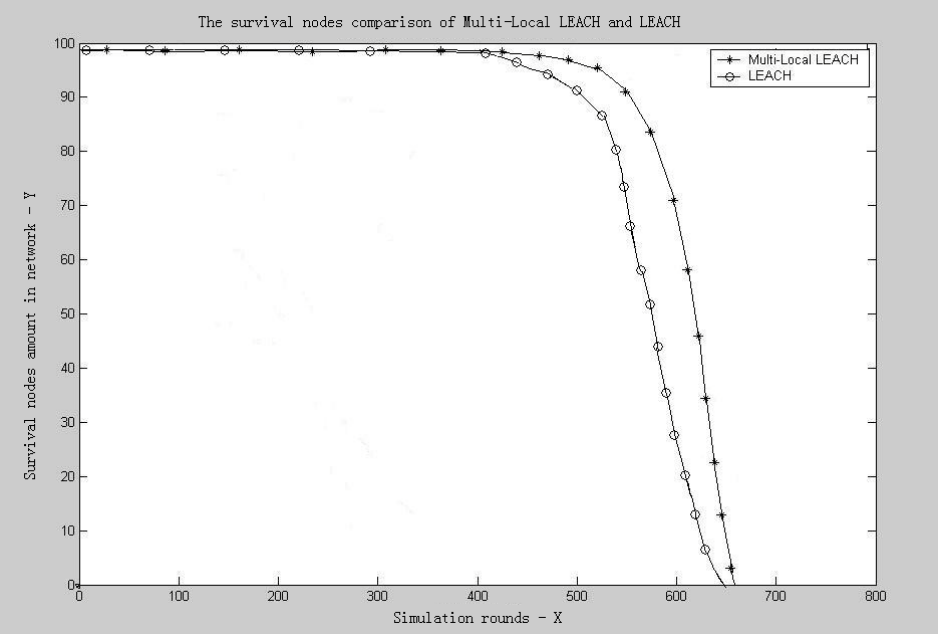


Fig. 2. The comparison of node lifetime.

distance between sensor node and the base station, A is the signal fading depth, normally taking the value between 45 and 49. The base station sorts the clusters in ascending order by the energy values it received from cluster-heads and then labels the clusters. The number of clusters will be subsequently sent to cluster-heads. Subsequently, the network moves into a relative steady-state phase where begins to performing data collection. According to the real situation of radio signal transmissions, we adopt multi-channel fading model which works well in practice to simulate the communication between cluster to cluster or cluster to base station.

2) Communication Mechanism Between Clusters: Since the output power of wireless sensor nodes is limited, we exploit multi-hop pattern of flat routing protocols to transmit data. In the network set-up phase, the base station numbers every clusters according to the distance values from the cluster heads. The larger the number is, the farther the cluster is to the base station. The policy of data transmission is : cluster-heads perform data fusion of local cluster data before sending out them, the cluster-heads in vicinity who receive the data firstly judge if the cluster number is larger than their cluster number, if it does they will forward the data and send confirming message, otherwise they will discard the data without forwarding them.

3) Internal Rotation Mechanisms: When a cluster-head receives a participation message of a cluster member node, it automatically increases the counter which records the amount of nodes. The cluster-head assigns a serial number and a special timeslot to each cluster node after a cluster is set up. The cluster nodes are only permitted to communicate with cluster-head in special timeslot, and they have to keep sleeping in other time. Once a cluster-head's transmission time reaches N, current cluster-head abdicates its role to a substitute node which is just awaking now and takes the place of the substitute node by using its timeslot. After that, the cluster head gets into the sleep state. To preserve the reliability of the replacing process, cluster-head firstly sends replacing message to the substitute node. When the awaking node receives the message, it sends its timeslot to cluster-head as an echo. After cluster-head receives the timeslot message, it transmits the data information and acknowledges information to the substitute node. Then, the former cluster-head node gets into sleeping state and waits for the arrival of its timeslot as a normal cluster node.

Ⅳ. SIMULATION AND EVALUATION

In wireless sensor network, the primary metric to evaluate an algorithm is network lifetime which means the network energy dissipation. This metric is usually measured by the number of survival nodes in the network. Without taking into account other external factors that may undermine the premise, we consider a node is dead when its energy is less than 0, and a network is beginning to dead when the first node died. The time consumed by the first dead node is defined as network lifetime. Excepting the energy dissipation of cluster set-up phrase, other energy are mainly consumed by the communication and signal amplification between cluster-heads and base station in LEACH algorithm. However, energy is mainly consumed by data fusion and relay in Multi-Local LEACH protocol. In this article, we performed multiple simulations of Multi-Local LEACH to get the mean network lifetime and compared it with LEACH.

A. Evaluation Metric

Our experiments use Matlab to generate 100 nodes which are randomly distributed in a region from (x=0,y=0) to (x=100,y=100). The base station whose energy is sustainable was located at the coordinate address of (x=25, y=150). Compared with the energy consumed by data transmission and reception in wireless sensor

network, the energy consumed by computation and storage is basically negligible. Hence, network lifetime is chiefly depending on data transmission. Assuming that the initial energy amount of each node is 0:5J and the energy consumption of transmission and reception is Eelec = 50nJ/bit. Amplification consumption ε_{amp} = 100pJ/bit/m^2; k which is the size of package is set to be 1000b; therefore the energy consumption of one transmission of a package is E = Eelec$\times$K + $\varepsilon_{amp}\times$ k $\times d^2$, the energy consumption on reception of package is E = Eelec $\times$ K, d is the distance between nodes. In LEACH, d represents the mean distance between inter-cluster nodes and cluster-head or the max distance between cluster nodes and base station. In Multi-Local LEACH, d is the mean distance between cluster nodes and cluster-head or the mean distance between adjacent cluster.

B. Simulation Experiment

TABLE I
THE COMPARISON OF LIFETIME OF WSN NODES.

	Round number of LEACH	Round number of Multi-Local LEACH
Round first node dies	415	465
Round 50% nodes die	580	624
Round last node dies	650	660

Figure 2 shows the Matlab simulation results of Multi-Local LEACH and LEACH. The x axis represents the round number of network simulation and the y axis represents the survival nodes amount after each round. Table 1 shows the statistic results of network lifetime of the two algorithms. According to the simulation results, the first node death in LEACH algorithm occurred in round 415, while the first node death in improved Multi-Local LEACH occurred in round 465 which is 1.1 times as much as the former. That is, the network lifetime is increased by 10%. All nodes of LEACH are dead in round 650 and that is round 660 for Multi-Local LEACH. There is no significant difference between these two algorithms apparently, but take a close look we could find that after the first node died, other nodes died more quickly in LEACH. In contrast, the trend of node death of Multi-Local LEACH before round 580 is much slower than LEACH. Obviously, Multi-Local LEACH is superior to LEACH concerning improving the lifetime of wireless sensor network. In figure 2, We could also find that the slope of curve of increase significantly after round 630, which is due to LEACH that is used at set-up phase of Multi-Local LEACH. Since LEACH algorithm leading to uneven distribution of cluster-heads, it is liable to increase the cluster-heads' energy consumption when increasing multiple hops. Because the internal rotation mechanism is adopted, the energy consumption of Multi-Local LEACH is still lower than conventional flat routing protocols. The curves show that, in improved algorithm, the energy of nodes can be more efficiently used in network, thus to guarantee the load balance between nodes in network and extend the network lifetime. If the drawback of uneven distribution of cluster could be overcome, energy dissipation will be further reduced and a longer network lifetime is attainable.

V. CONCLUSION

In this paper, we improve the drawbacks of LEACH protocol and perform simulation experiments for the revised algorithm. In addition, we analysis the performance of Multi-Local LEACH, which is the revised LEACH algorithm, in terms of network lifetime. The results of simulation show that the revised algorithm can obviously extend the network lifetime compared with conventional LEACH, where the first node death time is 1.1 times of LEACH and the network lifetime increase 10%, significantly improving the network performance. As a single-hop protocol, LEACH has the drawback of huge energy dissipation in long-distance transmission. To address this problem, Multi-Local LEACH evenly distribute the extra energy caused by relay data as a cluster-head in all nodes within a cluster through the internal rotation mechanism. If there are many nodes in a cluster, we find that the mean energy dissipation of each node is relatively huge, which is liable to result in quick death of local cluster nodes. How to implement efficient clustering and an even distribution of network nodes in clusters are the challenges need to be settle in our further work.

REFERENCES

[1] Z. Jiajia and H. Chen, "Clustered routing algorithm for wireless sensor networks," Chinese Journal of Sensors and Actuators, vol. 21, no. 1, pp. 130–134, 2008.

[2] A. C. Wendi Rabiner Heinzelman and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in Proceedings of the 33rd Hawaii International Conference on System Sciences, Hawaii, USA, 2000, pp. 10–20.

[3] X. C. Jin Ji and G. Yingjun, "A cluster-based routing algorithmłleach in wireless sensor network," Computer Applications and Software, vol. 23, no. 11, pp. 137–138, 2006.

[4] Z. G. Tiezhu Li and R. Zhang, "An improved node degree based leach algorithm," Computer Era, vol. 15, no. 2, pp. 16–18, 2008.

[5] G. Lijie, "Realization of multi-top leach energy-aware routing arithmetic on wireless sensor networks," Journal of Tianjin University of Technology, vol. 24, no. 5, pp. 47–51, 2008.

[6] L. Y. Du Xiangdang and S. Xiuhua, "Improved arithmetic in choice of head-note based on clustering of wsn," Chinese Journal of Sensors and Actuators, vol. 21, no. 7, pp. 1202–1206, 2008.

[7] Z. Lei and C. Shu, "Novel network protocol for wsn based on energy and distance," Journal of Computer Applications, vol. 28, no. 5, pp. 1117– 1119, 2008.

[8] S. Y. Fang Xiaofei and Y. Junjie, "A new leach-based routing algorithm for wireless sensor networks," Mechanical and Electrical Engineering Magazine, vol. 25, no. 5, pp. 100–103, 2008.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 219-222

Investigation of the Image Quality Assessment using Neural Networks and Structure Similarty

Chih-hsien Kung[1], Wei-sheng Yang[2], Chun-yuan Huang[3], Chih-ming Kung[4]
[1]Dept. of Engineering & Management of Advanced Technology
Chang Jung Christian University,Tainan, Taiwan
kung@mail.cjcu.edu.tw
[2]Dept. of Information Management,Chang Jung Christian University, Tainan, Taiwan
[3.4]Dept. of Information Technology and Communication
Shih Chien University Kaohsiung Campus,Kaohsiung, Taiwan
[4]alex@mail.kh.usc.edu.tw

***Abstract*—Artificial Neural Network (ANN) can be used to simulate the human nervous cells in the processing system. The advantage of ANN is constantly training to gain the accurate results. Structure Similarity (SSIM) expresses the quality of the images comprehensively by the image brightness, contrast, and structure. This research combines the Artificial Neural Network perceptrons and Structure Similarity characteristics to create different types of images suitable for weight value, expect through the video image intensifier to improve the visual identification, and provides the automatic image processing procedure in the future (e.g. analyze, detection, division, and identify).**

***Index Terms*—Artificial Neural Network, Structure Similarity, Perceptrons**

I. INTRODUCTION

As technology advances, the image display technology and industry have been evolved as well. In the research field of color images, image quality analysis is increasingly important. We establish an image color quality index, using the weight of brightness, color, and contrast. The index can enhance image quality assessment.

Digital image systems, such as digital cameras, printers, monitors, etc., the image quality can be presented by the system image quality assessment. The trainditional image quality assessment methods can be divided into Physical measurement and Psychophysics assessment categories. Physical measurement has been referred as an objective quality assessment which is the general formula property assessment. Psychological measurement emphasis on the assessment of subjective or perceived manners which is mainly derived from the feeling of image by the observer, that is the observing experiment by using human eyes.

We use image processing algorithms to make different quality of standard images by three important indicators (specify level, color level, and sharpness), perceived brightness, color, contrast, and image quality transforming by the human eyes. Taking some factors into consideration giving the weight values to produce an amount of formula assessment, and then the quality of image is evaluated according to this formula.

Color image quality assessment of three key indicators for the tone (which specify the degree of lightness), color and sharpness, the meaning of the metropolis lies generally against the three indicators to judge the quality of image. Using neural network to find the level of image quality of the three indicators, and give the weight value to establish a formula for image quality assessment. It can be successly improved and objective in image quality assessment methods.

II. ARTIFICIAL NEURAL NETWORK

Artificial Neural Networks is a widely discussed and re-studied topics in recent years. It refers to an imitation of biological neural network information processing system. To the biological point of view, the artificial neural network is a simple model of human brain. The simple operation element which corresponds to the brain's neurons and many connections throughout the network of neurons is known as the neural network. In fact, the advantages of neural networks can learn non-linear system, excellent learning ability, good fault tolerance and the characteristics of highly parallel computing power. Since the seventeenth century, doctors and anatomists lay the foundation of neural science.

In 1943, psychologist McCulloch cooperated with mathematician Pitts in the MP model [1], the mathematical model is called "form neurons". In 1949, Hebb proposed to change the connection strength of the Hebb neuron rule [2]. Rosenblatt introduced the concept of the perceptron [3] model in 1957, which is also the earliest and the simplest neural model. However, scholars believe that this model was the lack of hidden layer learning algorithm and the learning will be impeded. After that, the theory of neural networks was taken seriously again is because the development of artificial intelligence. Until the 1980s, Hopfiel neural network (1982) has been proposed. At this time, since the expert system encounters difficulties such that neural network theory has become important. Until now, neural networks have been widely used in various scientific aspects.

The American scholar F. Rosenblatt proposed the most original sensor model since 1957. As shown in Figure 1, the basic perceptron composed of components as a linear combination of function with accumulator and a hard limiter. It is also known as single-layer perception.While its input is greater than or equal to the weighted threshold value, the output is 1, otherwise is 0, the output as Equation (1). Basically, the perceptron is comprised of an adjustable value of synaptic weights, and the threshold of a single neuron.

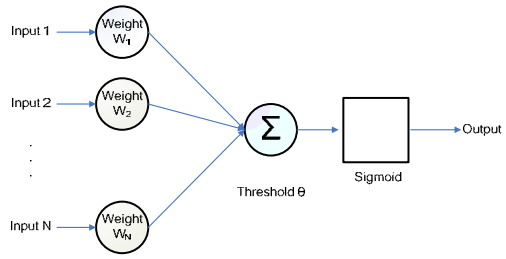


Figure 1. Perceptron

$$Y = \begin{cases} 0, when \sum_{i=1}^{N} \omega_i x_i < \theta \\ 1, when \sum_{i=1}^{N} \omega_i x_i \geq \theta \end{cases} \quad (1)$$

Generally, while the input of hard-limiter is positive, then the output of neuron is 1. Conversely, if the hard-limiter of input is negative, then the output of neuron is -1.

Perceptron has a good ability of identification for the linear segmentation data. The action function use the number of the sign function. There are two input features denoted as x1 and x2, and the output is obtained as (2).

$$y = \mathrm{sgn}(\omega x_1 + \omega x_2) = \begin{cases} 1, s > 0 \\ -1, s < 0 \end{cases} \quad (2)$$

Perceptron's output has two modes as the proof by the F. Rosenblatt's demonstration. It can be find a straight line s which separates the two modes and the ω will be constringed. If the problem can be divided into the two modes, then it is called a linearly separable problems. While the problem can not be separated, then s will does not exist, as ω1and ω2 will not exist.

III. IMAGE QUALITY ASSESSMENT

Image Quality Assessment in Image Processing plays an important role, as image processing algorithms and systems design benchmarks to help assess the best or the quality of the results. At present more commonly used by the image quality index for the assessment are the Mean Square Error (MSE) and the Peak Signal to Noise Ratio (PSNR), respectively, are defined as follows:

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (x_i - y_i)^2 \quad (3)$$

$$PSNR = 10 log \frac{255^2}{MSE} \quad (4)$$

where N is the size of image, xi and yi are the gray level of pixel of orignal image and test image However, these common approach, focused on the image gray value of the mathematical model to quantify the numerical standards, although with an objective assessment, but not all of the assessment results can meet the human visual judgement. By Fig. 6 can be found in the Test Signal 1, Test Signal 2 and Original Signal, Error Signal of the MSE results are the same, but the human visual judgement can only discover that the Test Signal 1 is closer to the Original Signal [4].

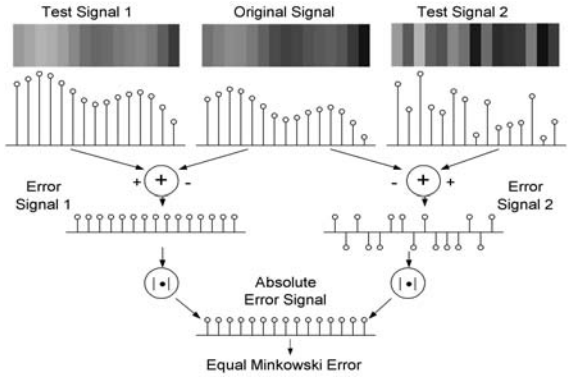


Figure 2. MSE distortion of the signal difference calculation [5]

A. Structural Similarity Index (SSIM)

In 2002, Wang is the first scholar to propose new image quality evaluation index [5-10]: Universal Quality Index and Structural Similarity index applied to video image evaluation criteria, The results showed that the two kinds of gray scale images were superior to focus on the mathematical degree of statistical indicators. Structural Similarity Index, SSIM, taking into account the image of the brightness, contrast and structural and comprehensive representation of the overall image quality. Fig. 3 is the diagram of the structural similarity (SSIM) measurement system [4].

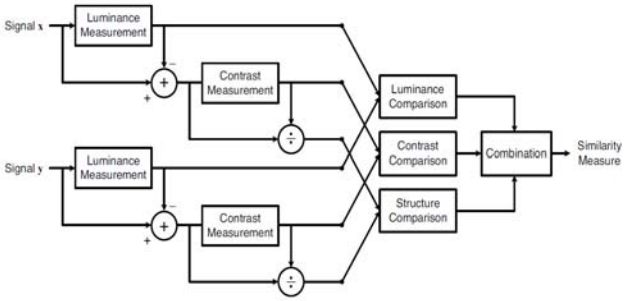


Figure 3. Diagram of the structural similarity (SSIM) measurement system [5]

The SSIM index is a full reference metric, in other words, the measuring of image quality is based on an initial uncompressed or distortion-free image as reference. SSIM is designed to improve the traditional methods like PSNR and MSE, which have proved to be inconsistent with human visual system. SSIM is also commonly used as a method of testing the quality of various lossy video compression methods. Using SSIM index, image and video can be effectively compared.

SSIM comprehensively indicates the structural similarity of the overall quality of the images which include the luminance, contrast and structure of images. The SSIM is defined as:

$$SSIM(x,y) = l(x,y)^{\alpha} \times c(x,y)^{\beta} \times s(x, \quad (5)$$

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \quad (6)$$

where $\mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i$, $\sigma_x = \left(\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu_x)^2\right)^{1/2}$

$$c(x,y)=\frac{2\sigma_x\sigma_y+C_2}{\sigma_x^2+\sigma_y^2+C_2} \tag{7}$$

where $\mu_y=\frac{1}{N}\sum_{i=1}^{N}y_i$, $\sigma_y=\left(\frac{1}{N-1}\sum_{i=1}^{N}(y_i-\mu_y)^2\right)^{1/2}$

$$s(x,y)=\frac{\sigma_{xy}+C_3}{\sigma_x\sigma_y+C_3} \tag{8}$$

where $\sigma_{xy}=\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu_x)(y_i-\mu_y)$

The SSIM index is a decimal value between 0 and 1. A value of 0 would mean zero correlation with the original image, and 1 means the exact same image. A motivating example is shown in Fig. 4, where the original "Lina" image is altered with different distortions. The "Lina'' image is dealed with various types of distortion, and the MSE is 225. However, the values of SSIM are significantly different. As shown in Fig. 4 (b), it is the image of best image quality, and the SSIM is 0.9327.

Figure 4. (a) Original Image "Lena" 512 x 512.
(b) Contrast-stretched Image, MSE = 225, SIM = 0.9327.
(c) Gaussian Noise Image, MSE = 225, SSIM = 0.3891
(d) Implusive Noise Image, MSE = 225, SSIM = 0.6494
(e) Blurred Image, MSE = 225, SSIM = 0.3461
(f) JPEG Image, MSE = 225, SSIM = 0.2871

IV. EXPERIMENTAL METHODS

This paper is focused on two issues, first is the output of the traditional PSNR and subjective assessment which are often contrary. Second, the SSIM is difficult for the serious blurred images to have an accurate assessment. This paper proposed a scheme which combines the Artificial Neural Network and technique of SSIM to improve the issues. In this study, the features of SSIM, including the overall image brightness, contrast ratio and image structural comparison are utilized. Since the image processing technique of SSIM is closer to the human eye, SSIM and the neural network are combined to perform the adaptive image quality assessment. Therefore, we combined single-layer percetpron and SSIM to establish the new single-layer percetpron. By the definition of SSIM, (10) is used and extend as (11).

$$\begin{aligned} SSIM(x,y) &= f[l(x,y),c(x,y),s(x,y)] \\ &= l(x,y)^{\alpha}\times c(x,y)^{\beta}\times s(x,y)^{\gamma} \end{aligned} \tag{10}$$

$$\begin{aligned} log\ (SSIM\ (x,y)) &= log(\ l(\ x,y\ )^{\alpha}\times c\left(x,y\right)^{\beta}\times s\left(x,y\right. \\ &= \alpha\times log\ l(\ x,y\ )+\beta\times log\ c(\ x,y \\ &\quad +\gamma\times log\ s(\ x,y\ ) \\ &= \sum_{i=1}^{3}w_i\times x_i \end{aligned} \tag{11}$$

where $w_1=\alpha$, $w_2=\beta$, $w_3=\gamma$, $x_1=\log(l(x,y))$, $x_2=\log(c(x,y))$, $x_3=\log(s(x,y))$

As shown in Fig. 5, is the single-layer model according to the established formula.

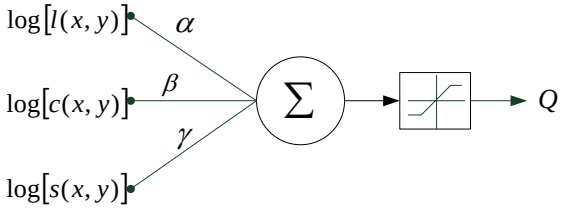


Figure 5. Combination of SSIM and Single-layer Perceptron

V. EXPERIMENTAL RESULTS

The experiments of picture image analysis have been performed on the portraits, buildings, animals and plants images. First, each image was compressed in different intensity of the JPEG compression, Gaussian blur, sharpening, noise processing, and contrast adjustments. As illustrated in Fig. 6, is one of the results of the image strengthened. The l(x,y),c(x,y),s(x,y) and SSIM values for each image is calculated. The training software used is the SuperPCNeuron which operated on a Pentium IV 3.0 GHz, personal computer with Windows XP operating system.

The learning rate is 0.01, weight range is from 0.1 to 0.5, the number of input variables is 3, and the number of output variables is 1. We define log[l(x,y)], log[c(x,y)], and log[s(x,y)] as input value of X1, X2, X3. Visually set a good image as 1, poor image is 0 for a training set of the output. The training mode employed are BPN network model and K-fold cross-validation method, and allow the system select images randomly as training data and test data.

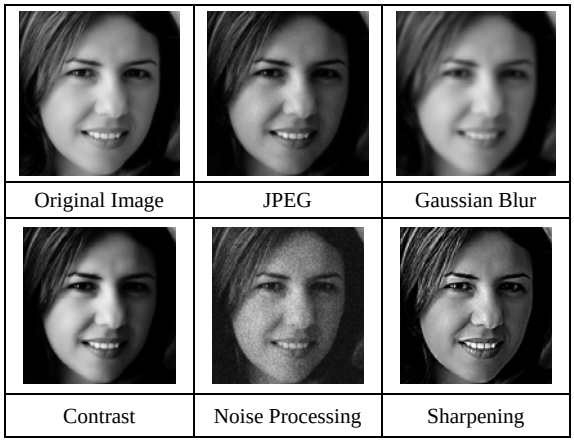


Figure 6. Sample Image

We use (11) to find the weight value of each picture of portraits, animals, buildings and plant which are denoted as W1, W2 and W3. Table I is the RMSE and weights for each portrait picture and non-portrait picture. The experimental results have illustrated that applying different weights in each group into different image values, the output values are approximate to the results using the SuperPCNeuron. Although the results are suffered from the errors, however, it does not affect the overall image evaluation.

TABLE I. WEIGHTS OF IMAGE

	W1	***W2***	***W3***	***RMSE***
portrait	-0.1073	0.1962	-0.0332	0.24349
animals	-0.2470	0.2659	0.0269	0.25951
buildings	-0.1287	-0.0143	0.2183	0.25612
plant	-0.2048	0.2146	0.0126	0.38116

F-Measure

To prove the proposed scheme can identify different types of images and suitable for image quality evaluation criteria, the F-Measure method is used to test and verify. The F-Measure contains order Recall and Precision are shown in Table II.

TABLE II. TRUE POSITIVES,TRUE NEGATIVES, FALSE POSITIVES AND FALSE NEGATIVES PARAMETERS

		correct result / classification	
		E1	***E2***
obtained result / classification	***E1***	tp (True Positive)	fp (False Positive)
	E2	fn (False Negative)	tn (True Negative)

Recall and Precision formula are defined as follows:

$$Precision = \frac{tp}{tp + fp} \tag{12}$$

$$Recall = \frac{tp}{tp + fn} \tag{13}$$

If R：mean to Recall, P：mean to Precision，then the F-Measure is defined as follows:

$$F = 2 \times \frac{Precision \times recall}{Precision + recall} \tag{14}$$

When the higher the value of Recall and Precision, the higher the F-Measure value which means that its quality is better. In Table III, we found that Precision values were above 0.7, F is also above average of 0.6. Experimental results have demonstrated that this approach can make the image quality of different types to achieve adaptability.

TABLE III. F-MEASURE DATA

	portrait	***animal***	***building***	***plant***
Precision	0.8730	0.8095	0.7619	0.8095
Recall	0.7432	0.5667	0.5517	0.6071
F	0.8029	0.6667	0.6400	0.6938

VI. CONCLUSION

In this paper, the Structure Similarity and Artificial Neural Network for image quality assessment are investigated. The SSIM can retain structural characteristics of the image. By using the ANN properties to find the coefficient of SSIM, all kinds of images of image quality assessment index can be establish. It can achieve adaptability for the image quality of different types, become the optimization of image processing parameters, and obtain both image structure and image quality of high-quality images. Experimental results have demonstrated that the proposed approach can make the image quality of different types to achieve adaptability.

ACKNOWLEDGMENT

This work is supported by National Science Council of Taiwan grants: NSC 98-2815-C-158-003-E, NSC 98-2221-E-158-005

REFERENCES

[1] W.S.McCulloch,and W.Pitts,"A logical calculus of ideas immanent in nervo-Us activity,"Bulletin of Mayhematical Biophysics, vol. 5,pp.115-133 1943.

[2] D. O. Hebb, The Organization of Behavior:A Neuropsychological Theory, Wiley, New York, 1949.s

[3] F.Rosenblatt,"The perceptron : A probabilistic model for information Storage and organization in the brain,"Psychological Review,vol.65,pp.386-408,1958.

[4] C. M. Kung , C. C. Ku, C. Y. Wang," Fast Fractal Image Compression Base on Block Property", *Advanced Computer Theory and Engineering, 2008. ICACTE '08. International Conference on , pp. 477-481*

[5] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, Apr. 2004.

[6] Z. Wang and Q. Li, "Video quality assessment using a statistical model of human visual speed perception," *Journal of the Optical Society of America A*, Dec. 2007.

[7] Z. Wang and X. Shang, "Spatial pooling strategies for perceptual image quality assessment," *IEEE International Conference on Image Processing*, Atlanta, GA, Oct. 8-11, 2006.

[8] Z. Wang and E. P. Simoncelli, "Translation insensitive image similarity in complex wavelet domain," *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. II, pp. 573-576, Philadelphia, PA, Mar. 2005.

[9] Z. Wang, L. Lu, and A. C. Bovik, "Video quality assessment based on structural distortion measurement," Signal Processing: Image Communication, special issue on "Objective video quality metrics", vol. 19, no. 2, pp. 121-132, Feb. 2004.

[10] Z. Wang, E. P. Simoncelli and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," Invited Paper, IEEE Asilomar Conference on Signals, Systems and Computers, Nov. 2003.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 223-225

Research on OGSA-based Distributed Computing Model

Wang Guowei[1], Chen Lin[2]
[1] School of Computer Science and Technology Henan Polytechnic University, Jiaozuo, China
Email: wangguowei@hpu.edu.cn
[2] School of Computer Science and Technology Henan Polytechnic University, Jiaozuo, China
Email: chenlin@hpu.edu.cn

***Abstract*—Grid computing is applying the resources of many computers in a network to single problem at same time, usually to a scientific or technical problem that requires a great number of computer processing cycles or access to large amounts of data. Many academic institutions devote themselves to define standard of Grid computing and hope to do some application engineering work. How to design the general architecture is the most important factor of Grid computing application. This article given a Grid computing model contains general architecture and interfaces. The model is a simple distributed computing application based on Open Grid Services Architecture, which given a detailed description on GridService architecture, interfaces, creation of transient services, factories, services lifetime management, notification by analyzing the architecture and interfaces function of Open Grid Services Architecture.**

***Index Terms*—Grid computing, distributed computing, OGSA, Lifetime management**

I. INTRODUCTION

With the development of computer network technology and internet, more and more scholars show interest in research of grid computing. Grid computing is a sort of distributed computing which can connect many computers in the world and make them a fully shared integrate resource by high speed internet network. Although grid computing is mainly researched in academe and scientific research at present, but more and more big IT academic institutions hope by using it to do some application engineering work.

Just like TCP/IP protocol is the core of Internet and each entity must use IP protocol, grid computing define standard protocol and standard service. After a long period of practice and improvement an approbatory standard – open resource Globus Toolkit was defined. In 2002, Open Grid Services Architecture (OGSA) was defined as the base services of the open grid services in global grid forum. OGSA is a sort of system architecture of grid computing, the center is Services based on Globus technology and Web Services, by combining Globus with Web Services OGSA expand Globus Tooklit protocol and define convention and Web Services Description Language (WSDL) interface[1]. Grid Services is a potential transient and stateful services[2] instance supporting reliable and secure invocation, lifetime management, notification, policy management, credential management, virtualization[1]. OGSA also defines interfaces for the discovery of Grid services instances and for the creation of transient Grid service instances. The result is a standards-based distributed service system that supports the creation of the sophisticated distributed services required in modern enterprise and inter-organizational computing environments.

II. A MODEL OF OGSA-BASED APPLICATION

Wherever Times is specified, Times Roman or Times New Roman may be used. If neither is available on your word processor, please use the font closest in appearance to Times. Avoid using bit-mapped fonts if possible. True-Type 1 or Open Type fonts are preferred. Please embed symbol fonts, as well, for math, etc.

A. General architecture

The model comprises four components: user application; data storage services; data operation services; database services. Each component initially encapsulated in a running environment. The general architecture is illustrated in Figure 1, the "R" represent local registry services of each service.

The working of the model illustrates as following:

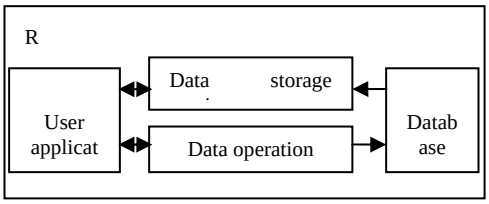


Figure 1. General Architecture

According to the data operation requirement user application request the creation of Grid services, the request should contains a creation of data storage services and a data operation services. Each request involves mutual authentication of the user and the relevant service followed by authorization of the request. Each request is successful and results in the creation of a new Grid service instance with some initial lifetime. The newly created data operation services request data operation from the database services and place the intermediate result in local storage. Meanwhile, the user application generates periodic "keepalive" requests to the Grid service instances that it has created. If the data operation was successfully finished in the initial lifetime the lifetime keeps constant, otherwise, the keepalive messages should send a request to prolong the initial lifetime to perform the data operation. If any other part

has an interest in the results, further keepalive messages should generate. If the application fails for some reason, the data operation computation continues for now, but user application can't receive the keepalive messages, due to the application failure, keepalive messages cease, and so the Grid service instances eventually time out and terminated, then free the storage and computing resources that they were consuming[3].

B. Function of components

A Grid service implements one or more interfaces. OGSA provide necessary Grid services interfaces and other optional interfaces to create Grid services, manage lifetime. OGSA defines a variety of behaviors and associated interfaces, all but one of these interfaces (GridService) are optional.

- Factory interfaces

Factory interfaces support the creation of transient Grid services instances.

User application request access transient Grid services instances by factory interfaces, and instances creates data operations Grid services by responding the factory interface's request or other proxy request, at the same time, the factory interface's creation service operation creates a requested grid service and returns the Grid Service Handle (GSH, global unique name of Grid service instance) and initial Grid Service Reference (GSR, abstraction of instance-specific information) for the new service instance[4]. The whole procession is illustrated in Figure 2.

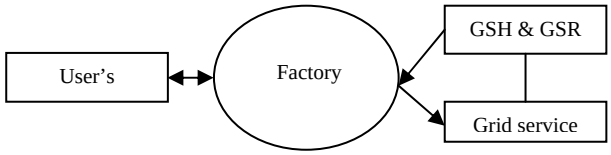


Figure 2. Creation of Grid service

Factory interfaces can control services, if a service fails, it can be restarted by higher level controlling services by calling on the factory interface.

- NotificationSource and NotificationSink interfaces

Services can deal with notifications in standard ways. OGSA defines common abstractions and service interfaces for subscription to NotificationSource and delivery of NotificationSink notifications.

- GridService interfaces

GridService interfaces are essential interfaces of OGSA, one of its important functions is lifetime management.

When requesting the creation of a new Grid service instance through a factory, client indicates minimum and maximum acceptable initial lifetimes. The factory selects an initial lifetime and returns this to the client.

Most of services in OGSA are transient instances, so determining when a service can or should be terminated is very important. In distributed system, messages may be lost, one result is that a service may never see the termination request, thus causing it to consume resources indefinitely. The other is that a service is terminated in initial lifetime but does not successfully perform the specific task.

OGSA resolve this problem through a soft state approach in which Grid service instances are created with a specified lifetime. The approach to Grid service lifetime management has two operations: Destroy and SetTerminationTime.

If the initial lifetime period expires without having received a message, the data operation services should be terminated (Destroy) and release any associated resources, even failures of servers, networks or clients.

If client wish the Grid services exist or perform another new task, he can send a keepalive message or requests a lifetime extension via a SetTerminationTime message to data operation Grid services, which specifies a minimum and maximum acceptable new lifetime. The factory selects a new lifetime and returns this to the client.

The periodicity of keepalive messages can be determined by the client based on the initial lifetime negotiated with the service instance and knowledge about network reliability. the interval size allows tradeoffs between currency of information and overhead.

Lifetime extension requests from clients are not mandatory. A service can decide at any time to extend its lifetime, either in response to lifetime extension request by a client or any other reason. If resource constraints and priorities dictate that it relinquishes its resources, the service instance can cancel itself at any time.

The message delivery procession is illustrated in Figure 3.

- Registry and HandleMap interfaces

Registry interfaces provides operations by which GSHs can be registered with the registry service, and an associated service data element user to contain information about registered GSHs. The Registry interface is used to register a GSH and the GridService interface's FindServiceData operation to retrieve information about Registered GSHs. By implementing

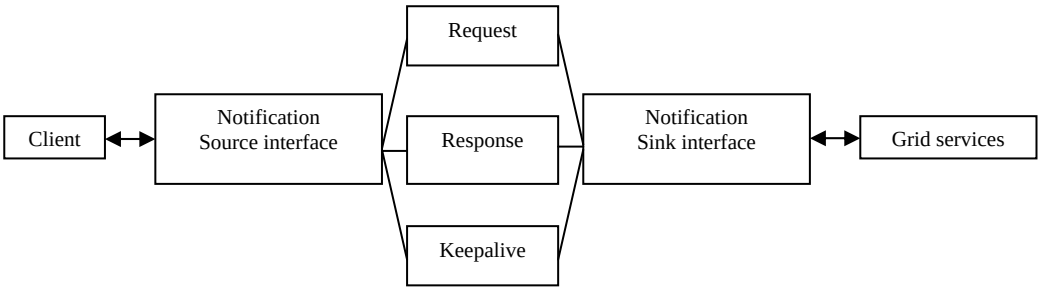


Figure 3. Message Delivery Procession

registry interfaces in all services so that they always have access to a registry to discover other services, registry interfaces can make service discovery possible. Registry interfaces should be run at a large computing machine if all services need to be notified of a service joining or leaving , because the number of message will increase with the number of services.

HandleMap interface is a handle-to-reference mapper that take a GSH and return a valid GSR.

Registry and HandleMap service are created automatically when the service starts up as a function of the hosting environment.

By using Registry and HandleMap users of many services and services get hold of the service descriptions to discover which services meet their needs. The registry needs to be searched to find the GSH of the services that fulfill the user requirements, and the HandleMap can be contacted to retrieve the detailed description of the services in question.

III. CONCLUSIONS

Grid computing has emerged as an important new field, distinguished from conventional distributed computing by its focus on large-scale resource sharing and high-performance orientation. Until recently, Grid computing technology needs to improve. Many academic institutions devote themselves to define standard and software application of Grid computing. The Globus Tookit that developed by Globus project made a great impact on Grid computing, and the OGSA perfect the Globus Tookit. Grid computing technologies are evolving toward an OGSA. OGSA defines a Grid service based on concepts and technologies from both the Grid and Web services communities. OGSA defines standard mechanisms for creating, discovering transient Grid service instances. The model this paper presented is a simple distributed computing application based on OGSA, which is an abstract representation of both real resources, such as nodes, file systems, and logical resources. It provides some common operations and supports multiple underlying resource models representing resources as service instances.

REFERENCES

[1] Ian Foster, Carl Kesselman, Jeffrey M Nick.steven tuecke,“the phvsiology of the grid”, http://www.globus.org/research/papers/ogsa.pdf, 2002

[2] Clark, D.D, The Design Philosophy of the DARPA Internet Protocols. SINCOMM Symposium on Communications architectures and protocols, (1988), ACM Press,106-114

[3] Ian Foster. What is the grid? A Three Point Checklist, http://www.fp.mcs.anl.gov/~foster/articles/whatisthegrid.pdf, 2002

[4] Du Zhihui, Chen Yu, Liu Peng , Grid computing, Tsinghua University Press, 2002, pp 56-58

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 226-229

Research for Constructing the phrases Database of the Shang Oracle-Bone Inscriptions Based on N-Gram Model

Kai Jin-Yu[1], Li Na[2], and Liu Yong-ge[3]
1. School of Computer and Information Engineering, Anyang Normal University
Oracle Information Processing Key Laboratory
Anyang, China
aykjy@qq.com
2. School of Computer and Information Engineering, Anyang Normal University
Oracle Information Processing Key Laboratory
Anyang, China
Lina_youxiang@yahoo.com.cn
3. School of Computer and Information Engineering, Anyang Normal University
Oracle Information Processing Key Laboratory
Anyang, China
ay_liuyongge@163.com

***Abstract*—The key to using computer technology to correct segmentation and processing the information of the Shang Oracle-Bone Inscriptions is accurate identification of the words and phrases. Currently, there is not the special words and phrases database of the Shang Oracle-Bone Inscriptions in this field, it is very important to construct the special words and phrases database automatic, efficient, scientific and dynamically, depending on the characteristics of the Shang Oracle-Bone Inscriptions and the own of the large resources of the Shang Oracle-Bone Inscriptions, in this paper, we Construct the words and phrases Database of the Shang Oracle-Bone Inscriptions Based on N-Gram Model, experimental results show that using the statistical computing language model for constructing the words and phrases Database of the Shang Oracle-Bone Inscriptions is a strong practical and feasible.**

***Index Terms*—N-gram model, Bigram model, High-frequency characters, the special dictionary of the Shang Oracle-Bone Inscriptions**

I. INTRODUCTION

The Shang Oracle-Bone Inscriptions is the China's earliest systemized character. The research on the Shang Oracle-Bone Inscriptions is significant for studying the development of China's history and society, for understanding Chinese language, Chinese characters' development, evolution, and voice structure, vocabulary, the transformation of meaning.

With computer technology penetrating into all research areas, the computer-aided the Shang Oracle-Bone Inscriptions textual research and explication is under development. The use of computer technology converts the Shang Oracle-Bone Inscriptions research form manual to intelligence, and liberates researchers from the tedious work, thus makes the study of the Shang Oracle-Bone Inscriptions advance into a new era of information technology.

In order to enable computer to understand and process the information of the Shang Oracle-Bone Inscriptions, the key problem to be solved initially is to split the Shang Oracle-Bone Inscriptions correctly. At present, the commonly used automatic word segmentation technology has three categories: the technology of matching based on string, based on understanding, and based on statistics. Behind the two technologies are based on the first technology, and has improved the first technology, these two technologies are involved to the technology of matching string based on word.

The technology based on matching characters is also named machinery word segmentation; it is to be used with the dictionary. At present, there is no special dictionary in the field of the Shang Oracle-Bone Inscriptions. So, if we want to use the technology of machinery word segmentation, we must construct the dictionary of the Shang Oracle-Bone Inscriptions firstly. Today, the computer participate in the research work in every fields, artificially, to establish and maintain the dictionary of the Shang Oracle-Bone Inscriptions is a very stupid and clumsy thing, it will be a very meaningful work to automatic, efficient, scientific, and dynamically build the dictionary of the Shang Oracle-Bone Inscriptions. It is good, as following:(1)To extract words and phrases from the corpus linguistics of the Shang Oracle-Bone Inscriptions, can reflect the special vocabulary in the field of the Shang Oracle-Bone Inscriptions.(2)The statistical data information produced, during the course of constructing the dictionary of the Shang Oracle-Bone Inscriptions, can be used for future research.

It is a good choice to use statistical language model to construct the dictionary of the Shang Oracle-Bone Inscriptions, based on the large number of machine-

This paper supported by NSFC (60875081).

readable corpus, in this paper, we will build the dictionary of the Shang Oracle-Bone Inscriptions using the N-gram model which statistic data depending on the frequency of the word combinations in the corpus.

II. THE IDEAS OF CONSTUCTING PHRASE BASED ON STATISTICAL

From the formal point of view, phrase is a stable combination of characters, In context, the more time the adjacent characters appear, the more likely to form a phrase. Therefore, the frequency or probability of the adjacent characters appearance can reflect the credibility of the phrase better; we can calculate their mutual information through the frequency or probability of the adjacent characters appearance in the corpus.

Mutual information reflects the current relationship between the combination of Chinese characters tightness. When the close is higher than a certain threshold, the word can be considered that this group may constitute a phrase. Definition of the word of the mutual information is calculated two characters X, Y are adjacent total probability. Using this method would extract some of the current vocabulary of high frequency group; therefore, thinking based on statistics of word formation is based on statistical methods of high-frequency word string.

III. N-GRAM MODEL INTRODUCE

A. N-gram model

High-frequency characters theory is based on N-gram statistical model. An N-gram model can be utilized to find the most probable segmentation of a sentence.

Given a character sequence like C= C_1 C_2 C_3⋯C_1, the length of it is l, in the context, if only the first n-1 characters have impacted to the probability of the next character that the first n characters, Probability that:

$$P(C_l \mid C_1\ C_2\ C_3 \ldots C_{l-1}) \approx P(C_1 \mid C_{l-n+1} \ldots C_{l-1}) \quad (1)$$

According to the probability multiplication theorem and N-gram model, the probability of the characters sequence C= C_1 C_2 C_3⋯C_1 can be expressed as the product of the probability every characters comprising, as following:

$$P(C)=\prod_{i=1}^{l} p(C_i \mid C_1 \ldots C_{(i-1)}) \approx \prod_{i=1}^{l} p(C_{l-(n-1)} \ldots C_{i-1}) \quad (2)$$

When P (C) over a certain threshold, indicating the n-characters binding strong, then these n-characters can be viewed as a phrase.

B. The value N

As for the value N, ZHANG Shuwu in the paper of analysis of the value N in the Chinese statistical language model finds that from the view of detecting no-logged word (Chinese rare words) and the reconfiguration the values of N 4 is better.

But, in the field of the Shang Oracle-Bone Inscriptions, from the results of the research, we can find the Shang Oracle-Bone Inscriptions have the following characteristics:

1) At present, from the 150,000 or so discoverd Oracle Bones, there are more than 5,000 have been released about 1,500 words. Using 5000 characters to build 150,000 pieces of Oracle Bones, so the special dictionary in the field of the Shang Oracle-Bone Inscriptions that we want to build is not make of the uncommon characters.

2) In the Shang Oracle-Bone Inscriptions, as there are not many inscriptions, phrases are not rich, the phrase is offen expressed by a single character through the method of Multi-word Or Utilizing part of speech Or Multi-word word, expression of the different meaning, so there are a large proportion of the sing-character phrases in the Shang Oracle-Bone Inscriptions.

3) As for the two-characters phrases in the Shang Oracle-Bone Inscriptions, mostly including of the Ganzhi time terms or names, place names, titles, etc., such as: Xin-Chou, Yi-Hai, Shang-Jia, Zu-Yi etc., those two-characters phrases also hold a certain proportion in the Shang Oracle-Bone Inscriptions, and the probability of appearing in the Inscriptions is large.

4) As for the three-characters phrases or more-than-three-characters phrases, there are only a little.

During the course of constructing the special dictionary in the field of the Shang Oracle-Bone Inscriptions through the method of the N-gram statistical model, the dictionary is mainly built using a high frequency of common words, not rare words. According to the characteristic of the Shang Oracle-Bone Inscriptions, there are mainly single-character and two-character phrases, so, in the course of constructing the dictionary through processing the Shang Oracle-Bone Inscriptions, we find that the values of N-2 is better, therefore, Bigram model. Bigram Model is used by statistical two adjacent characters co-occurrence frequency; the system self-organization generates the corresponding phrases, the formation the special dictionary.

C. The values of statistical probability P

How much on earth are the values of statistical probability P for the two adjacent characters co-occurrence frequency, then, the two adjacent characters co-occurrence can be considered as a phrase? Popularly say, how high is the frequency of the two adjacent characters co-occurrence can be considered as a phrase? Usually define it in three ways.

Way 1: Calculate the probability of composition of the characters sequence C = C= C_1 C_2 C_3…C_{l-1} C_l:

$$P(C_1 \mid C_1\ C_2\ C_3 \ldots C_{l-1}) = \frac{Count(C_{(l-(n-1))} \ldots C_l)}{Count(C_{(l-(n-1))} \ldots C_{(l-1)})} \quad (3)$$

By the formula (2) and formula (3), we can estimate P(C_1 C_2 C_3…C_{l-1} C_1), the C_1 C_2 C_3 ⋯ …C_{l-1} C_1 probability. Function will statistic the number of the characters sequence C1 C 2 C 3⋯C l-1 Cl.

Way 2: Through the absolute frequency, that is, depending on the times that those two adjacent characters

co-occurrence in the literature determines the high-frequency phrases, also depending on the length of the literature. That is: we expect that they can be adjusted by the formula f(c)>f(l), f(c) means the sequence of the high-frequency strings, and f(l) is a statistical value decided by the length of literature. In this paper, when $l \le 500$, f(l)=2; when $500 < l \le 2000$, f(l)=3; and when $2000 < l \le 10000$, f(l)=4; when $l > 10000$, f(l)=4.

Way 3: By calculating the relative frequency. Suppose the string frequency is f (C) for the string C. If f (c) meets the formula (4), then string c is as high-frequency strings, where P is the user-specified threshold.

$$\frac{f(c)}{\max\{ f(C) \}} \ge P \qquad (4)$$

Those three ways can be used very simple. In this paper, we will make experiments to adjust which one is the best for constructing phrases.

IV. THE ALGORITHM AND EXPERIMENT

A. The selection of the Shang Oracle-Bone Inscription

In general, the total number of characters of a piece of the Shang Oracle-Bone Inscription are only 20 or 30, and we can not complete our work through one piece Inscription, we use together a considerable number of pieces of Inscriptions to form a training text database. Because the characters on the turtle shells and bones have existed a very long time, some of them are not clear, illegible, or because the turtle shells and bones are incomplete, resulting in the contents of the Inscriptions is incomplete. Therefore, before reading the training database of the Inscriptions into memory, they should be pretreated. In this experiment, we add the label people ";", where the turtle shells and bones are not complete, thus a piece of Inscription will be divided into a number of Inscriptions. The experiment data is from the graphics library of the Shang Oracle-Bone Inscription which finished by Anyang Teachers College Professor Liu Yongge including of 61256 turtle shells and bones. In the graphics library, there exist even if the number of the rubbing different but really a same piece of rubbing. Rubbing on the bones of these inscriptions experimental samples formed after the merger, using of ";" to separate the Inscriptions.

B. The algorithm

N-Gram model algorithm include of three parts: pretreated, substring extraction and post-processing. Pretreatment of the training text into memory, then deal with them to form the Inscription training database; the part of sub-string extraction is that according to Bigram model and the three probabilities adjust way to statistics and extract probably phrases; the third part will process the Incorrect "phrases". For example: turtle shells and bones have no punctuation, but, during forming the training database, we added in";" together the Inscriptions,";" and other characters may be composed of a high probability of word combinations, so these incorrect "phrases" should be removed form the dictionary.

C. The Data of the Experiment

Do the experiment on the basis of researching and analyzing above, we can drew the data shown in the following table 1.

TABLE I.

EXPERIMENT DATA

The training library of the experiment	61256 pieces of the Shang Oracle-Bone Inscription
Algorithm	N-Gram(Bi-gram)
The number of the single-word phrease abstracted（Nsingle）	835
In comparison with 1500 recognized, the coverage of the single-word phreases（Nsingle/1500）	55.66%
The number of the two-words time phrease abstracted（Ntwo）	68
The number of the correct time phrease（Ncorrect）	53
The correct rate of the time phreases（Ncorrect/Ntwo）	77.94%
In comparison with 60 recognized time phreases, the coverage of the two-words phreases（Ncorrect/60）	88.33%

V. CONCLUSION

Through this experiment we can draw the following conclusions:

(1) It is practical and feasible that constructing the special dictionary in the field of the Shang Oracle-Bone Inscriptions using the method of the N-gram (Bi-gram) statistical model.

(2) In the extraction process of the two-characters phrases, as for the higher frequency of Ganzhi time-expression person names, place names, those three kinds of probabilistic methods to judge the results are good. Thus, selection of the appropriate parameter, every one of the ways will be ok to adjust the result.

(3) For idioms extraction, according to different parameters, the extraction result is not the same, but because the length of idioms are often not only two-characters combinations, the idiom are often combined with three or three more characters, so as for the extraction of the idioms, the correct rates using N-2 model are not high, so do not recommend using this approach to extract idioms.

(4) As for three-characters phrases there are a little in this field, because there are very few, no need for 3 –

gram Model (Trigram Model) to statistics, we can manually add them to the dictionary.

(5) Because the training library is not complete, so, the capacity of the special dictionary in the field of the Shang Oracle-Bone Inscriptions is not large, we can expand the dictionary through expanding the training text database. We will continue to improve it in subsequent work.

REFERENCES

[1] HUANG Xuanjing, WU Lide, WANG Wenxin, and YE Danjin, “A Machine Learning Based Word Segmentation System Without Manual Dictionary”, Pattern Recognition and Artificial Intelligence, vol. 9, December 1996.

[2] Zhang Shu-wu,Huang Tai-yi. The analysis of N-Value in the chinese statistic language model.Journal of chinese information processing. No.1, 1998

[3] WU Yingliang, WEI Gang, and LI Haizhou, “A Word Segmentation Algorithm For Chinese Language Based On N-Gram Models And Machine Learning”, Journal of Electronics and Information Technology, vol. 23, November 2001.

[4] HAN Kesong, WANG Yongcheng, and CHEN Guilin, “Research on Fast High-frequency Strings Extracting and Statistics Algorithm with no Thesaurus”, Journal of Chinese Information Processing, February 2001.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 230-233

Equipment Status Management System of Coal Mine Base on Internet of Things

Zhao Wentao[1], Dong Jun[1]
[1]School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo City, China
zwt@clc.hpu.edu.cn
dj8519@163.com

***Abstract*—The concept and working principle of Internet of Things are presented. Then a simple equipment status management system of coal mine has been designed in VS2008 which effectively monitors the status information of variety of devices on the underground, such as work location, operating conditions, fault records and the main components parameters. Besides, it can timely exchange the information with the remote customer and manufacturers. what's more, the system has been designed for the transmission of the EPC code information of all kinds of underground equipments by the way of combining the industrial Ethernet and wireless sensor network, avoiding the problem of the privacy leak caused by the tradition of using RFID as tags.**

***Index Terms*—Internet of Things; coal mine; sensor network; Industrial Ethernet**

I. INTRODUCTION

Early in 1998, two professors at Massachusetts Institute of Technology presented that unique number should be given to all items to be identified based on radio frequency identification technology (RFID), which birthed the concept of Internet of things; 2005 International Telecommunication Union (ITU) released《ITU Internet Report 2005: Internet of Things》which pointed out the "Internet of things" communication age' coming; Internet of things new ideas and new technologies were discussed in the world's first international Internet of Things Conference "Internet of Things 2008" in Zurich in 2008; the speech of Premier Jia-bao Wen in Wuxi proposed the "Internet of things" concept in August 2009.

From "wisdom of the Earth" to "induction China", another wave of the information industry following the computer and the Internet is triggered with the concept of Internet of Things. Internet of Things is based on the extension and expansion of the Internet network, and the Internet is still the core and foundation of it. Extension and expansion of its client to any goods and goods between the various types of objects through the device of electronic tags, sensors, connects wireless networks through the interface realize communication and dialogue of man and object, the object and the object. Traditional thinking is broken down by Internet of Things, full sense, reliable transmission and intelligent networking processing are the basic characteristics of it. Internet of Things concepts and technologies are quietly going into our lives.

Internet of Things refers to the link between the various types of sensors and the existing Internet. Its meaning is derived from the concept of the object recognition of the RFID and the data exchange networks. Internet of Things can be divided into 3-tier system: perception layer, network layer and application layer. Readers, reader, sensor networks, M2M terminals are included in perceptual level, so as to implement the "objects" of the identification. Communications networks as well as network integration are included in Network layer. Network layer as universal service infrastructure make the Internet of Things become universal service. The role of the application layer is that Internet of Things combines with specific industry, which can realize the intelligent application control [1].

II. RELATED WORK

With the development of Internet of Things Technology, which will effectively reduce operating costs and increase efficiency with large-scale application in various industries. In such a network, goods can "exchange" for each other, without human intervention, and can identify any of the products to make the product with dynamic information of "smart products" in any place, any time.

Currently there are some applications about Internet of Things, RFID is a key technology in Internet of Things. Every RFID success stories can be regarded as the pro totype of Internet of Things application.

RFID technology originated abroad, as a replacement for bar codes, RFID is an advanced non-contact automatic identification technology; in the Internet of Things it is RFID make article "speak". Automatic identification of target signal and obtaining the relevant information both need radio frequency signals, and information and applications are processed by the back-office software, and then they achieve the "transparent" administration through open computer networks for information exchange and sharing of the goods. RFID devices are generally three parts: passive RFID and reader for the writing tag data, antenna for transmitting and receiving signals [2].

The role of Electronic Product Code (EPC) is to establish a global open identity standard for each product. EPC carrier is RFID electronic tag, which is equivalent to a pointer. The corresponding IP address can be found from the network by this pointer, relevant things

information that stored in that address. However, in order to make the technology and advantages of Internet of Things fully realized, a unified coding standard must be formulated. Already published or under development criteria are related with the data collection. And there are much standardization, such as the protocol of the reader and the computer data exchange, performance and conformance test specifications of RFID tag and reader, communication protocol between the RFID tag and reader and the label in the format ID and the structure of data retrieval aspects and so on [3,4,5].

However, it is easy to produce items of information disclosure if identify items with RFID tags, because every reader can read the EPC information in the tags; Besides, present RFID tag is made of metal and bonding agent, although it is very small, but it will be the issue of renewable materials, likely to cause environmental pollution. Equipment management system of coal mine is designed to transfer all kinds of EPC information of underground equipment which makes use of the industrial Ethernet and wireless sensor network combination, avoid the traditional use of RFID tags for the privacy leak and environmental pollution.

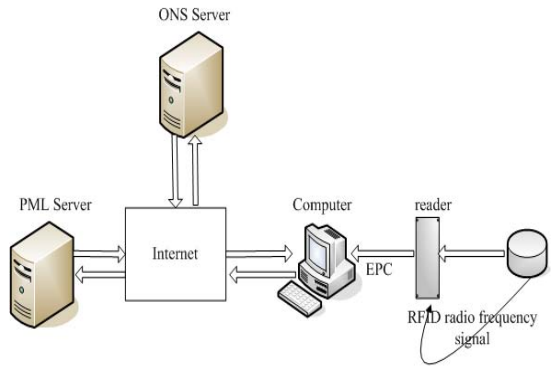


Fig1. Traditional way of reading RFID tags

Ⅲ. Internet of Things Technology in Coal Mine

The coal industry occupies a pivotal position in China's economic development. Following the time of Internet of Things, its related technologies can be applied to the coal industry, which can effectively manage the mine equipment, reducing accidents. In mines, there are a variety of related equipments, such as mining equipments: loader, belt conveyor; roadway transportation equipment: cranes, elevators and so on.

A. Equipment Management System of Coal Mine

Equipment management system of coal mine which has been developed makes the Internet of Things applied to mine, and mainly provides a more convenient and secure management platform. Much information are fed back to administrator, the Internet of Things that using in coal mines can be a continuous focus on real-time monitoring of all measuring points of the mechanical and electrical equipment on / off and other environmental parameters. The parameters of various information can be collected and data can be storied and processed. In addition, graphics and real-time data can be displayed. So as to make the equipment down hole alarm and power off based on various parameters by vinous' monitoring center. Then the equipment can be adjusted timely to adapt to humidity, temperature, wind, gas and other environmental factors to make information down hole input automatically and each device work link recorded accurately. Then the accidents can be reduced effectively. This work status of various types of underground equipment can not only be fed back to monitoring center on the ground in order to better control the operation of equipment, also can contact remote manufacturers for the product tracking, maintenance, troubleshooting, etc.

An ID tag for each equipment of the coal mine can be deployed, which access to the EPC product code down hole device or network to read and write. The code information is sent to the local control center through local data interface, so that the operator can easily control the various equipment of underground. What's more, use unique feature of EPC code for each device to find corresponding information, and access PML server through remote data interface to obtain the device information through a common format. Generate a list of coal products, then update the devices information to local monitoring center through local data interface[6].Generally speaking, this information includes the device manufacturer, product name, the effective service life, working hours, etc.

B. System Design

According to the above principles, a simple equipment management system of coal based on Internet of Things technology is designed in the windows platform using visual studio 2005. This system interface shown below, EPC codes of drill and pump automatic generate work status and parameter information list through the down hole wireless sensor network and industrial Ethernet transmission, in order to operate equipments conveniently for administrators in the monitoring center.

Fig2. System interface

C. Key Technologies

1) Sensor Node

In the coal mine requires a sensor network and three kinds of sensor nodes used to work with the local monitoring center are needed to install [7].

- Wireless sensor monitoring node: it is installed in the mine to test the operation of each device, wireless sensor nodes make use of the magnetic wave sensors to transmit and receive information, and it also can accept the orders from local monitoring center in this application.
- Relay node: the main task is to relay the data transmitted. Relay node can boot from the local monitoring center to receive information, and to issue timely signal to staff, which both still would be able to receive orders in time to solve the problem of underground equipment.
- Cluster nodes: this node controls the entire information of underground network; the main task is to serve as a gateway between the underground sensor networks and external network on the ground. All kinds of information through wireless means are delivered timely to local monitoring center.

2)Industrial Ethernet

Wireless sensor network is joined with main network Ethernet in underground mines, and here industrial Ethernet is adopted. Industrial Ethernet in mine employing ring network structure provides high reliable information transmission. When a breakdown occurs, time of network recovery is short, and monitoring data of reliable transmission is guaranteed.

Mine digital communication source mainly are data, image and sound. All kinds of information flow not uniform, and if there is requirement of real-time information transmission, and various equipment works synchronously, traditional TCP/IP and Ethernet cannot meet the requirements. But in the industrial Ethernet switching technology can be introduced to solve problems quickly network real-time, higher bandwidth can be enjoyed by each Ethernet devices. In this system, without RFID tag reader, make use of industrial Ethernet network card to read the related EPC Information of goods.

3)PML Server

PML Server is a product built and maintained by the manufacturers. The product code is determined according to the principles laid down in advance. Its role is to provide detailed information about equipment, that is based on XML standard and make the product searched by the ID. The advantage is that the heterogeneity of data storage is shielded and product information services are provided in a uniform format.

4)Savant Middleware System

The primary mission of middleware is to provide a series of calculations and data processing. And it is also used to capture、filter、calculate、proofreading、store data、manage the tag data read by the reader. So as to improve the efficiency and obtain the useful resources from the mass information [8]. Whichever level the Savant system in the hierarchy and all the Savant systems have a unique task management system. There are some user-defined modules in it, such as read-write interface, event management, application software interface. Then they can achieve user-defined tasks to carry out supervision and management. When the products are added to RFID tags, readers identify them in the process of EPC tags, and then EPC code is passed to the Savant system. This product information stored in the location can be found by re-using product name server ONS through Savant system .

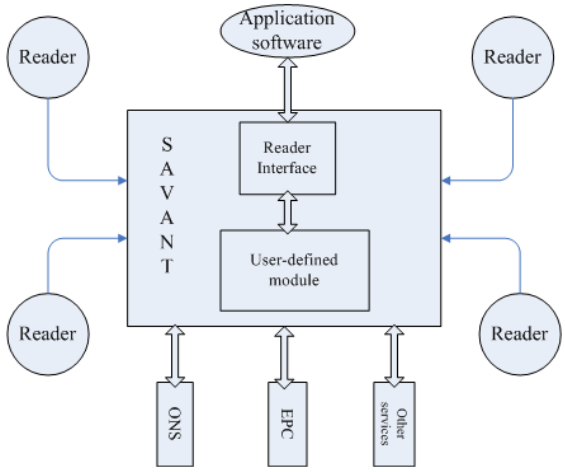


Fig3. Middleware system structure

5)Object Name Service (ONS)

EPC tags only store electronic product code, and these codes would also be matched to the corresponding product information through the middleware system. ONS provides two kinds of static and dynamic content. Items URL returned by the manufacturer is provided through static service, besides dynamic service can record the production process of specific equipments under the conditions of operation in sequence [9]. One of the basic roles of ONS is that an EPC is mapped to one or more URL; more information of items related to specific device can be searched in the URL. ONS is used to create map link between EPC code and the PML description. In

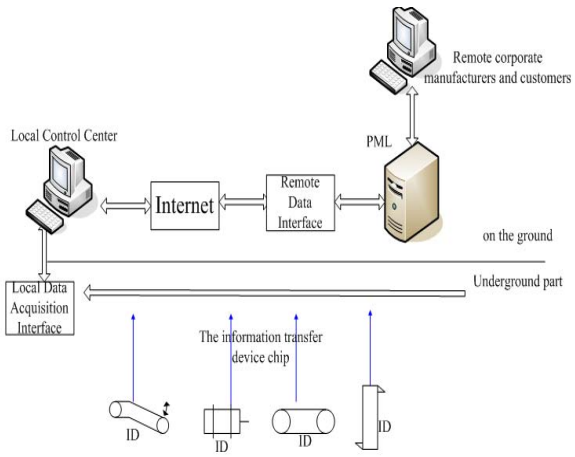


Fig4. Internet of Things applications in mine system structure

the end, the product EPC code information can be obtained from the management system. And PML information server of the product can be found from the ONS, so as to gain the detailed product information.

6)Local Control Center

Local monitoring center is used to manage the local database of equipment underground. All kinds of product information of production equipment will finally store in the database through the local interface in order to control equipment running. Local monitoring center uses the data acquisition interface to access underground equipment EPC code, and through remote data interface to retrieve detail information of the equipment from the PML server.

Ⅳ. CONCLUSION

Internet of Things technology for the coal industry has great development potential, and there are possibility and necessity of its implementation. Core technology of Internet of Things has been studied based on a simple coal mine equipment management system in this paper, the working principle of Internet of Things were elaborated. With its large-scale application in various industries, ruling out the traditional reader to identify information and using industrial Ethernet to identify EPC information will be the future trend. Then the next step, we will combine related technology of industrial Ethernet and wireless sensor network to have further improving.

REFERENCES

[1] Wei Liu, Hong-mei Wang, Qing Xiao and Jian Yang. “Internet of Things Concept [J],” Telecommunication Technology, 2010, vol.430, pp. 5-7.

[2] Bao-yun Wang, “Internet of Things Technology Research [J],” Journal Of Electronic Measurement And Instrument, 2009, 12(23).

[3] Zhi-feng Liu, Hong-hai Zhang and Jian-hua Wang, “The EPCglobal network construction based on RFID technology[J],” Computer Applications,2005,vol.25.

[4] Zhi-yu Ren, Pei-ran Ren, “Internet of Things and EPC / RFID technology [J],” Forest Engineering, 2006,22 (1).

[5] Daniel W. Engels. “A Comparison of the Electronic Product Code Identification Scheme& the Intemet Protocol Address Identification Scheme.”

[6] Yong Xie, Hong-wei Wang, “Automatic object-based storage networking management system and its application [J],” Logistics Technology, 2007, 26 (4).

[7] Li Ping, Tu kui, “M2M-based Intelligent Parking Information Service System Application [J],” Telecommunication Technology, 2010, vol.430, pp. 35-36.

[8] Qing Hu, Yi-ju Zhan and Xiao-hu Huang. “RFID-based enterprise networking and middleware technology materials [J],” Micro Computer, 2009, 25 (20).

[9] GSl EPCglobal. “Object Naming Service(ONS)Version 1.0[s],” EPCglobal Ratified Specification Version of October 4，2005.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 234-237

Research on the campus Emergency Command System Based on GIS

Yongqiang Ma, Jiyu An
School of Computer Science and Technology, Henan Polytechnic University, JiaoZuo, Henan, China
Email:{ysumayong,ajy770406}@163.com

***Abstract*—By analyzing of the demand for campus safety, combined with multi-data fusion and GIS, application of campus emergency command system is proposed. Designed the framework of emergency command system and the major subsystems and modules, combined with multi-level multi-source data fusion the framework of the system is improved. Finally, the developmental patterns of the emergency system functions, system data and management platform based on GIS are instruction.**

***Index Terms*—Geographic information system; emergency command system; campus safety**

I. INTRODUCTION

In recent years, the growing problem of school safety and emergencies occur. School need to strengthen education for prevention, but also need to use modern technology to ensure the performance of its functions [1].

Unexpected events with Time and Space, there is a variety of forms, especially, sudden vicious violence, generally have serious detrimental effects and large range of features. Therefore, when incidents occur, school departments need to immediately take quick and effective measures, rapidly and accurately understanding of the accident site, the scope, the possible proliferation of area, and other spatial information, the continuing impact of the event for the control, the development form of appropriate contingency measures. Combined with Internet and large databases, using spatial data management and spatial analysis capabilities of geographic information system (GIS)[2],Campus emergency command system achieve long-range real-time monitoring of events and can quickly derive the best line of emergency plans and rescue vehicles, as departments control the spread of the incident, winning the event's handling time, emergency decision to increase the level of event processing to provide a basis for accurate and timely information, and. In order to guarantee for peace and stability for school, computer technology and network technology will be the core of information technology which widely used in the school's security management .In this paper, combined with multi-data fusion and GIS, we proposed application of campus emergency command system.

II. ANALYSIS OF CAMPUS EMERGENCY COMMAND SYSTEM REQUIREMENTS

Emergency command system which is based on framework of the GIS is a spatial information and non-spatial information integration system, works with the process of crisis management in the emergency command, achieves a dynamic campus environment monitoring and real-time monitoring of emergencies, provides emergency response and other functions and statistics of emergencies environment and simulation analysis on the events, control and auxiliary decision-making ,and can conduct real-time monitoring of vehicles and scheduling.

To achieve real-time scheduling with the information resources, processing resources and communication resources of department , and more scientific and visualization for process of emergency command, Emergency command system which control connection between of the software and the hardware, exchange the content and format of information, interaction between subsystems control, standardization of the various subsystems ,which will bring each separate device, features and information integration to the interconnected system , unified and coordinated to achieve resource sharing and centralized management.

Emergency command system ,which help the security department of campus to better respond to criminals escape, hostage taking, suicides, the impact of school classrooms lawless, vicious violence, and other emergencies, can share resources for the analysis of decision-makers to provide timely information and emergency measures to help the work of a more scientific management and provide a theoretical basis.

III. STRUCTURE AND FUNCTION OF SYSTEM

A. Structure of Campus Emergency Command System

Campus safety precautions and emergency command system with three levels the main frame fusion center is considered[3-5], the first stage is the two types of data collection, to form the initial integration of data, the second level is constituted with map data and information of the campus, the third level is the emergency response decision which, in the case of an emergency, it can quickly activate contingency plans, in the electronic map display and plotting information, support decision-making for leadership of the command, to handle emergencies, as shown in Figure 1.

B. The major subsystems and functional modules

1) data acquisition

Emergency command system used to monitor the campus network, real-time monitoring the spatial distribution and flow of state school information, which

dynamic monitoring, to predict the warning of major incidents that may occur in particular.

a)Camera monitoring system: monitoring system is in the campus dormitory area, roads, education areas and other important parts of the school gates, and the students focus on venues to install monitoring equipment, through the computer network to the monitoring center to provide on-site real-time information, which improved processing ability to respond quickly to incidents.

b)Security Patrol: refers to the site by the security patrol, unusual events reported to the monitoring center, with the monitoring system. This can increase the warning range, which dynamically monitor anomaly events which happened outside cameras surveillance.

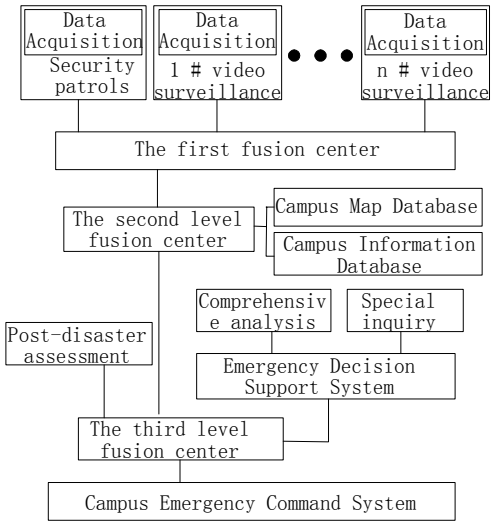


Figure 1. Campus Emergency Command System

2) Emergency Decision Support System

a)Emergency response: An emergency happens, according to Alarm and forecast information, according to event type, level and the corresponding approval process, activate the relevant plans, which pooled analysis of the security sector through the forecast results, the extent of the accident, methods, duration and comprehensive analysis of the degree of harm, the first time to respond, according to plans set processes and with the human and material resources, which through the communications center and emergency systems inter-connected, for task allocation, implementation and interconnection of information, automatic emergency response.

b)Decision-making chain of command: Responsible for the security department or level based on type of event, command centers set up, which pre-call plan in the resource information and configuration information, supplemented by modern technology: spatial information, on-site center video, expert resource library, meeting Health and derived from these projections, helping to protect, save manpower, and the appropriate resources to retrieve and deploy.

c) The completeness of Emergency plans and the rapid of check is very important, classification of plan and detail stored in the database which is the path of plan, so the query efficiency is increased accordingly.

d)Simulate the spread of unexpected events: Choose the map location of unexpected events, such as under the influence of information calculated using the Gauss algorithm for the proliferation of simulation events, and map out. The same time you can choose the extent of the information, including school hours, classrooms, dormitories, office space and other information. Select from the affected areas and the scope of these effects which can be drawn histogram. Finally it gives to calculate the extent of the spread of the optimal results which circuit block diagram, and marked with different colors.

e)Emergency Plan Management Subsystem: [6, 7]The corresponding experience in the prevention or emergency matters, which establish the criteria for plan services, grade, response departments ,assist the resources to conduct appropriate follow-up treatment processes and mechanisms facilities , provide appropriate emergency command for decision-making based on quick response.

f)Emergency training and drills subsystem: the corresponding departments, and emergency personnel, which training by the appropriate emergency services with the plan to drill. In accident and emergency preparedness exercise simulation, rational organization to deploy emergency resources, coordinate the emergency departments, agencies, staff relations, improve coordination between the various plans and overall emergency response capacity.

g)Special inquiry: [8]Special search system provides a variety of emergencies around the space and time information, dormitories, classrooms, office space, according to the class time. Administrators can further processing by topic.

h)Post-disaster assessment: [9]From the social, economic and psychological impact, emergency assessment system carried out research on effects of Post-disaster assessment with qualitative and quantitative. For uncertainty of unexpected events and the spatial extent affect largely, to carry out methods of direct economic losses of monetary theory and assessment , proposed principles for grading emergency and compensation mechanisms. using quantitative assessment of damage theories and methods of quantitative evaluation of monetization, to adopt a classification and compensation mechanisms for post-disaster assessment and develop classification of compensation mechanisms.

3) Map data, and campus information

A variety of geographic information and graphics data which are the basis of the system module, stores school information of all students and staff and their families information, including address and school information. Map data management add and delete data on the graph, sets the display of different layers of data (set the color, symbols, etc.), makes the dynamic adjustment of each layer, defines the order in such layers show. Graphic data by the data query acquired the property and attribute in data center which showing its corresponding graphics data. Its feature include multi-point between two points

and the distance calculation, area calculation, comprehensive database query, map location query, and overlay analysis, buffer analysis, shortest path analysis, these analytical tools and related school information, can be related analysis of information, emergency routes, can statistical impact of unexpected events.

C. System Design

1) System Architecture

Campus Emergency Command System includes the geographic data access, comprehensive analysis, emergency response training, special inquiries, and other components, its structure, shown in Figure 2.

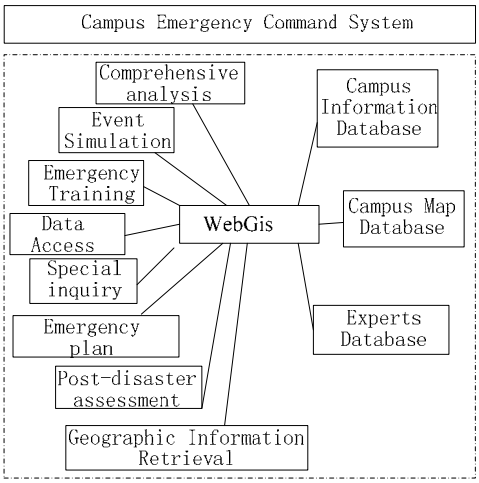


Figure 2. Architecture of Campus Emergency Command System

2) System Workflow [6]

In case of emergencies, required the security sector in the shortest possible time access to information, quickly response, effectively deal with emergencies, maximize the safety of students, reduce the scope of the event, through the monitoring system could be the first time access to information, and can quickly mobilize stakeholders, and the Advisory Expert decision-making, start the appropriate plan, shown in Figure 3.

3) System data design

Campus emergency command system, the data preparation is particularly important, which is to achieve its basic geographic information system function, logical structure shown in Figure 4.

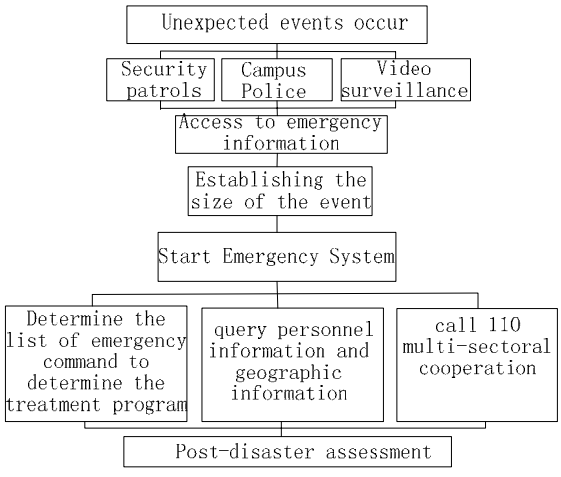


Figure 3. System Workflow

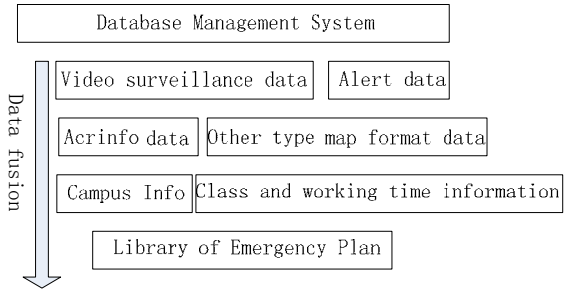


Figure 4. Data fusion process

a) Video surveillance data, alarm data, describing the incidents of primary data, a description of the emergency is unilateral and blurred.

b) Output of electronic maps is in the form of graphic image, and information analysis and queries, is another description of the incident, the description of space.

c) Staff working time information and class information is critical, which is the scope emergencies and post-impact analysis, the most important data.

d) According to the completeness of the information on the incidents described in the final database, according to experts, given the appropriate contingency plans.

Emergencies is a multi-level data fusion, multi-level data processing, mainly from multiple information sources to complete the data automatically detect, association, correlation, estimation and combination, the treatment.

IV. CONCLUSIONS

Campus emergency command system use geographic information system (GIS) spatial data management and spatial analysis functions and advanced data fusion concepts, which can quickly arrive at the data the best emergency plans and rescue vehicles, routes, events for the control of the spread of the various departments, won the event's handling time, emergency decision to increase the level of event processing to provide accurate and timely information and basis. The campus security department better respond to criminals escape, hostage taking, suicide, the impact of criminals outside the classroom, vicious violence, and other emergencies. Using the system for the analysis of decision-makers can provide timely information on emergency measures to help the work of a more scientific management and provide a theoretical basis.

REFERENCES

[1] He Xiaoxia, “multi-campus university crisis management mode of incident management response mechanisms,” Ideological Education, 2009 (S1), pp. 243–244.

[2] Jia Jianhua, etc., “GIS in the emergency plan in the application, ” Surveying and Mapping, 2009 (06), pp. 282–284.

[3] Niu Min and Jiang Jie, “China's disaster prevention and reduction of the Basic Law framework for the design,” Jiangsu Social Sciences, 2010 (01), pp. 155–160.
[4] Fu Chaoyang and Jin Qinxian, “environmental emergency management information system and constitute the overall framework of the study,” China Environmental Monitoring, 2007 (05) , pp. 83–86.
[5] Jun Wu, “public emergency response mechanism constitute the framework,” Statistical and Decision, 2006 (13) , pp. 54–57.
[6] Li Qi, etc. “Design of Integrated Emergency Response System,” Manufacturing Automation, 2010 (02) , pp. 1–5.
[7] Ding Jie, Ji national and Liu Faneng, “the city emergency command system based on the optimal path algorithm,” Journal of Xiamen University (Natural Science), 2009 (05) , pp. 462–468.
[8] Huang Shimin, etc., “geographic information system in urban Earthquake Disaster Mitigation Research,” science research, 2007 (S1) , pp. 1–7.
[9] Wang Shuzhen and Feng Qimin, “spatial decision support technologies in urban earthquake emergency software system,”World Earthquake Engineering, 2006 (02) , pp. 89–97.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 238-241

Design of Automobile Anti-theft and Alarm System Based on MCU and Information Fusion

Zhang Feng
Dept. of Electronic Engineering, Sichuan University of Science & Engineering, Zigong, China
E-mail: zfzj@suse.edu.cn

***Abstract*—An automobile anti-theft and alarm system based on MCU and information fusion is introduced in this paper, in which multi sensors are used to collect the information about automobile's situation firstly, and a MCU is used as the central processing and control center, the FCM and the neural network algorithms are adopted to fuse the multi-sensor information respectively, finally, the automobile's safety information is sent to the mobile phones of automobile owners through the GSM network in order to achieve the real-time monitoring for automobiles.**

***Index Terms*—Information Fusion, Neural Network, MCU, Anti-theft and Alarm**

I. INTRODUCTION

At present, the anti-theft device on the market processes passively alarm signal in many cars. Generally, it determines safety state of automobile depending on a single signal from sensor, therefore it have some defects, such as less reliability, existing phenomenon of misreport and failing to report. These defects bring many unnecessary troubles to user. It is very difficult to identify precisely safety state of automobile using only a single sensor because automobile have many different information due to its unsafe situations (for example, illegal starting, being damaged, handing entire car and false failure phenomenon, etc.). We can precisely judge and describe state of automobile if we combine information from multi-sensors that are complement each other in space and time and redundant information according to a rule of optimization. An automobile anti-theft and alarm system is designed using MCU as control center of this system in this paper. In order to obtain precise automobile's safety information all data from different sensors is fused in two levels by using the FCM and the neural network algorithms in this system. Deciding whether or not to start brake system is made depending on automobile's safety information and finally, the automobile's safety information is sent to the mobile phones of automobile owners through the GSM (Global System for Mobile Communications) network.

II. STRUCTURE OF ANTI-THEFT SYSTEM

A. Design of system structure

Modular structure is adopted in this anti-theft system, and in order to meet diverse demands for different customers the number of sensors or modules may be increased or decreased according to user's requirement. This system consists of detection modules, CPU and control modules, alarm modules, auto ignition control modules and communication modules. The structure of system is shown fig.1.

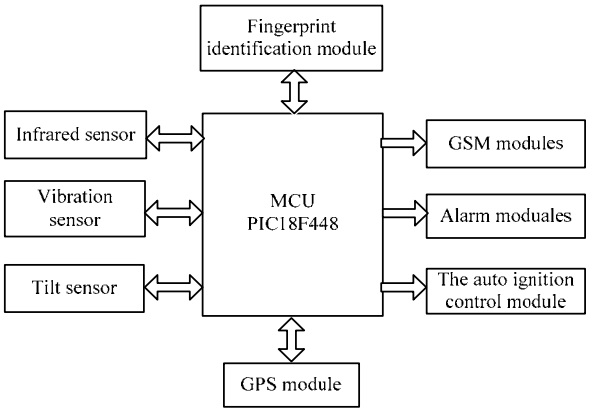


Figure 1. The structure of the anti-theft and alarm system

The CPU and control modules consisted of MCU PIC18F448 and its external circuits is used to receive the information from all detection modules, fuse different information by a corresponding method, and then make a judgment on automobile's safety condition on the basis of the result of information fusion. It can transmit different level alarm signals to alarm modules, executable modules and GSM modules, respectively. The auto ignition control module in the system can directly cut off automobile's ignition circuit to enable automobile can't work when automobile is being stolen. The detection module comprising many items (infrared sensor, vibration sensor, tilt sensor, fingerprint identification module and GPS module) is used to collect the automobile's real-time safety state and then transmit these signals to CPU and control module. The alarm module is used to receive instruct from CPU and give acoustic and optical alarm signals. User can monitor and control his automobile due to the communication module. The system transmits the alarm signals to automobile owner via short message service (SMS) using GSM network and user also realizes remote control to his automobile by the same means.

B. Detecting automobile's safety information

Situation due to automobile being stolen consists of illegal starting and being damaged and handing entire car. In these cases, the phenomena of vibration or leaning will occur in auto body and they will accompany some human biological information, for example, nocturnal radiation whose centre wavelength is from 9 to 10μm. In order to obtain an exact alarm signal, collecting information must be finished by a sensor corresponding to specific

phenomenon due to being stolen and be fused using a correct algorithm. The selection for the sensors complies with the following rules in this system to get enough information: selecting correctly and optimum integration sensors to meet low cost and high precision and robust, realizing the sharing and complement of information by exerting the advantage of different sensors, rational distribution of sensors on the board to eliminate detection dead area and increase the reliability of system. On the basis of these rules, the following detection modules of anti-theft system are chosen in this system to detect the real-time state of automobile:

1) Pyroelectric infrared sensor: Pyroelectric infrared sensor is sensitive only to the infrared radiation whose center wavelength is from 9 to 10μm and can find the radiation information from human body.

2) Tilt sensor: detecting whether or not to change between auto body and its initial position using Tilt sensor, we can make a conclusion that the automobile is being carried as a whole if this change appears with a certain frequency and reaches a presupposing value.

3) Vibration sensor and GPS module: The vibration sensor made from acceleration detection sensors can detect the tilt information of being stolen by carrying as a whole for automobile. The control center gets the automobile's address via reading data from GPS system after obtaining the tilt information. Since the drift of GPS locator, the GPS module can be started and send the address to owner by SMS only when change value is greater than a threshold.

4) Fingerprint identification module: Fingerprint identification system will be triggered and identify the operator's identity when the automobile's engine is forced to be started. The anti-theft system can lock the engine by starting automatically the spark control system and send an alarm message to owner via GSM module if its operator can't pass the fingerprint identification system.

III. MULTI-SENSOR INFORMATION FUSION

Multi-sensor information Fusion is actually an information processing paradigm that simulates the human brain to process a comprehensive matter. Its basic principle is to combine redundant information and data from different sensors, which are complement in time and space by an optimization algorithm, on basis of fully utilizing multi-sensor resources and correct using information, and then to get the consistency of interpretation and description about observed thing. In this paper, clustering and neural network algorithm are used to fuse the information from multi-sensors for feature level and decision-making level, respectively, to realize the real-time detection of automobile's situation.

A. Feature level fusion

The infrared sensor, tilt sensor and vibration sensor are used to detect automobile's state in time sequence and collect the feature information corresponding to different automobile' state, and then all feature information is fused using three ART-2 neural networks by time-fusion method. The ART-2 neural network is a excellent data clustering approach and its structural chart is shown in fig.2.

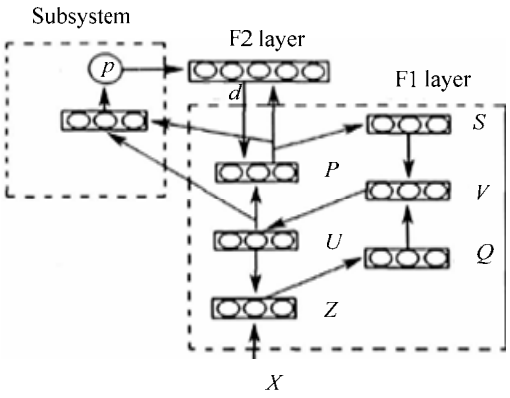


Figure 2. The basic structure of ART2

1) Fuzzy C means clustering(FCM)

FCM is a data processing unsupervised algorithm. In FCM, objective function is iterated according to the weighted member in C classification center and all samples collected in order to get the optimal classification of data sets. The result of clustering is that the grades of membership between every sample and all classifications are obtained and their sum is 1. Assuming $X=\{x_1,x_2,\cdots,x_n\}$ expresses n sets containing unclassified samples. Where $x_i \in R^P, i=1,2,\cdots,n$, P is feature dimension of samples. FCM clustering approach can classification X into C subsets ($S_1,S_2,\cdots,S_C$). If $V_1,V_2,\cdots,V_C$ express clustering center of corresponding to set, respectively, and u_{ij} expresses grade of membership between x_i and S_j , and $U=\{u_{ij}\}$ is $n\times c$ matrix, error sum of squares is employed as a objective function in FCM algorithm as follow:

$$J(U,V)=\sum_{i=1}^{n}\sum_{j=1}^{C}u_{ij}^{m}d_{ij}^{2} \qquad (1)$$

Where: u_{ij} meets a constraint condition as follow:

$$\sum_{j=1}^{C}u_{ij}=1, 1\le i\le n; u_{ij}\ge 0, 1\le i\le n, 1\le j\le C \quad i=1,2,\cdots,C \qquad (2)$$

$m\in[1,\infty)$ is the fuzzy weighted index and is used to control the fuzzy degree of matrix U. The fuzzy degree will increase with m and the optimal range is from 1.5 to 2.5 (m=2 in this paper).

$$d_{ij}^{2}=\left\|x_i-v_j\right\|^{2}=(x_i-v_j)^{T}A(x_i-v_j) \qquad (3)$$

Where, d_{ij} is the distance between a sample and the clustering center, A is a symmetric matrix (A is a unit matrix in this paper). $J(U,V)$ is the sum of squares of the weighted distance between samples in different classifications and the clustering center, the smaller $J(U,V)$ means better clustering. FCM algorithm is an iterative process to minimize its objective function. By Lagrange multiplier method we can obtain

$$v_j = \sum_{k=1}^{n} (u_{ij})^m x_k \Big/ \sum_{k=1}^{n} (u_{ij})^m \tag{4}$$

$$u_{ij} = (\frac{1}{d_{ij}})^{\frac{2}{m-1}} \Big/ \sum_{j=1}^{c} (\frac{1}{d_{ij}})^{\frac{2}{m-1}} \tag{5}$$

U and the elements of matrix V are modified until the minimum value of J(U,V) is obtained.

2) FCM-based feature-level information fusion

Window length of the sample model is set to 60s. The interval between two times sensor information collections is set to 3s. Therefore the number of data in a pattern window is 20. The information pattern features for infrared sensor, tilt sensor and vibration sensor are expressed as follows, respectively:

$X_1 = [T_1, T_2 \ldots T_{19}, T_{20}]^T$; $X_2 = [P_1, P_2 \ldots P_{19}, P_{20}]^T$; $X_3 = [H_1, H_2 \ldots H_{19}, H_{20}]^T$

Because dimension of X1 or X2 or X3 is 20 (N=20), the number of input pin in ART2 network is also 20. In experiment, other parameters in ART2 network are set as follows: Contrast constant is 12 (a=b=12), the adjustment subsystem constant c=0.25, field gain of output layer equal 1(d=1), filtering threshold is 1/200.5 ($\theta = 1/N^{0.5} = 1/20^{0.5}$), filter transformation function is expressed as following:

$$f(x) = \begin{cases} 0, & 0 \le x \le \theta \\ x, & x > 0 \end{cases} \tag{6}$$

Pattern vector X1, X2 and X3 are written into ART2 network. The information from every sensor is clustered into four typical classifications, which is represented as follow using coding: 00(no risk), 01(smaller risk), 10(risky) and 11(alarm). After fusing, fusing space C for infrared sensors, tilt sensors and vibration sensors is obtained.

$$C = [C_1, C_2, C_3, C_4, C_5, C_6] \tag{7}$$

Where, C1 and C2, C3 and C4, C5 and C6 are the coding of information from infrared sensors, tilt sensors and vibration sensors, respectively. GPS and GSM modules will be started and send the automobile's address to owner via GSM network when the information from the vibration and tilt sensors is conformed to be "alarm" in order to avoid that the automobile is stolen by carrying whole body. On the other hand, fingerprint identification and spark control modules will be started when the information from the infrared sensors is conformed to be "alarm" to prohibit starting the engine.

B. Decision-making level fusion

Three-layer BP neural network based on BP is applied to this system to process the former fusing result, it can predict the probability of being stolen for automobile by integrating the information from different sensors. Input layer in BP neural network consists of 6 neurons corresponding to six messages from three sensors (infrared, tilt and vibration), namely, six variables in vector C (from C1 to C6), on the other hand, out layer consists of only one neuron corresponding to a three-dimension vector B (b1, b2, b3). Element b1 in B represents the safety situation of monitored automobile (0: normal, 1: alarm). The b2 and b3 combine to represent the working state of three sensors (00: normal, 01: abnormal infrared sensors, 10: abnormal tilt sensors, 11: abnormal vibration sensors). In addition, the selection of nodal points of hidden layer neuron network based on BP can be carried out through empirical formula as follow:

$$r = (m+n)^{1/2} + (1 \sim 10) \tag{8}$$

Where, r, m and n represent the number of the neuron in hidden layer, input layer and output layer, respectively. In this system, r is equal to 12 so that it can meet to the requirement of detecting precise of vehicle condition and the training time of network. Its network structure chat is shown in fig.3.

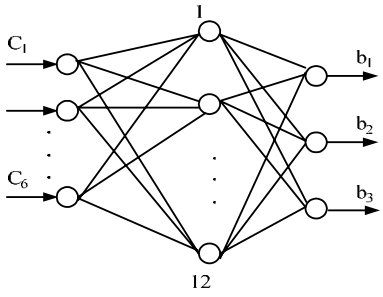


Figure 3. BP neural network structure

C. Simulating results

If the transfer functions between two layers in three-layer BP neural network is Gaussian function, then

$$Y_j = \sum_{i=1}^{12} W_{ij} Z_i + b_j \qquad (j = 1,2,3) \tag{9}$$

$$Z_i = \exp[-d_i(X, C_i, B)] \tag{10}$$

$$d_i(X, C_i, B) = (C_i - X)^T B(C_i - X) \quad (11)$$

Where, Y_j is the output of neuron j in the output layer, W_{ij} is a weighted value from neuron i in the hidden layer to neuron j in the output layer, b_j is the offset of neuron j in the output layer, Z_i is the output of neuron i in the hidden layer, d_i is the Euclidean distance between the input vector X and the vector C_i corresponding to neuron i in the hidden layer, B is the diagonal matrix of RBF (Radial Basis Function) bandwidth. According to formula (9), (10) and (11), we can obtain:

$$Y_j = \sum_{i=1}^{12} W_{ij} \exp\left\{-(C_i - X)^T B(C_i - X)\right\} + b_j \quad (12)$$

The ultimate purpose for us training the neural network is to modify its weighted value and then minimize the error sum of squares of network output, namely,

$$J = \min \sum_j \left\| y_j - y_j^m \right\|^2 \quad (13)$$

Where, y_j^m is the output of the neural network, y_j is the desired output. Simulation samples and the results of training are shown in the following table:

TABLE I. TABLE1 SIMULATION SAMPLES AND THE RESULTS OF TRAINING

samples	infrared (mm)	tilt (degree)	Vibration (mm)	network output
1	60	1	0.1	000
2	55	1.5	0.9	100
3	49	4	14.5	111
4	58	6	1.1	110
5	28	0.8	0.25	101

In table 1, sample 1 represents that the automobile is safe and three sensors are normal, sample 2 represents the automobile is to be risky and all sensors are normal, sample 3, 4 and 5 represent the automobile is still to be risky as well, the difference is that sample 3, 4 and 5 mean the abnormal phenomenon of the vibration sensor, tilt sensor and infrared sensor, respectively.

IV. CONCLUSIONS

An automobile anti-theft and alarm system based on MCU and information fusion is designed in this paper and we can realize the full-view monitoring of the automobile's safety state by using it. The comprehensive analysis for many kinds of vehicle condition from the different sensors is accomplished on basis of the fully utilizing the differences and complementarities of functions for different sensors, the accuracy of alarm is improved, since the system can identify actively and reliably the automobile's safety state by taking advantage of two-level information fusing on basis of FCM clustering and neural network. Compared with the traditional anti-theft system including only a one sensor, this system has some advantages as follows: low cost, high precision, expandability, real time and good robust, etc.

ACKNOWLEDGMENT

The research is supported by science and technology research project of Sichuan University of Science & Engineering (No.2007ZR004).

REFERENCES

[1] Gao xu wei, "A GPS/GPRS System of Vehicles," Master thesis, Dalian university of technology, 2005.

[2] He you, Wang guo hong, Lv da jin, "Multi-Sensor Information Fusion and Application," Electronic industry press, 2000.

[3] Xiao ying hui, Ou yang jun, "Research and Design Intelligent Vehicle Anti-theft Alarm System Based on More Integration of Information Technology," Computer & Digital Engineering, vol. 37(3), 2009, pp.114-116.

[4] Chen yong, Li yun xia, Lu xia fu, Li feng hua, "Multi-sensor intelligent wheelchair obstacle avoidance system based on information fusion technology," Digital Communication, 2009(04), pp. 58-61.

[5] Wang zhi gang, Fu xin, "Multi-sensor Information Fusion and Its Application," Electro-Optic Technology Application, vol. 23(3),2008, pp. 71-75.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 242-244

Application of Virtualization Technology in High-Performance Computing

Yan Junhao[1]，Xue Mingxia[2]

[1]College of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, China
Email: yanjh@hpu.edu.cn
[2]College of Accounting, Jiaozuo University, Jiaozuo, China
Email: xmxxmx16888@163.com

***Abstract*—In software, hardware, data centers and cloud computing, the figures of virtualization technology can be seen. Where the virtualization technology has not really been involved in is in high-performance computing. By analyzing the specific applications of virtualization technology and the possible challenges it will face if applied in high-performance computing, this paper highlights the virtualization technology application prospects in high performance computing.**

***Index Terms*— virtualization technolog, high-performance computing, Application prospects**

I. INTRODUCTION

Compared with the extensive application in x86 server field, virtualization technology's application in high performance computing is relatively scarce [1]. As we all know, the extensive use of virtualization technology in x86 servers has a necessary condition - the utilization and load of the server's CPU is low. In high-performance computing, the main focus is on the implementation of parallel high-density high-load task, and the processor is almost close to full capacity all the time. At the same time, high performance computing workloads is more influenced by memory bandwidth, I/O bandwidth. It seems that virtualization and high performance computing can be described as "incompatible". This view about the application of high performance computing virtualization technology is also universal among the majority [2].

As all PC servers are not suitable to be virtualized, it is not true that in high performance computing all applications are not suitable to be virtualized. To some extent, virtualization is just the effective way to enhance the extra value of high performance computing, and even the best choice to solve the traditional difficulties high performance computing has been faced with.

II. VIRTUALIZTION TECHNOLOGY AND HIGH PERFORMANCE COMPUTING

Virtualization technology first appeared in IBM mainframe systems in the 60s and 70s of the last century to support high-level software sharing underlying hardware resources, to provide users with multiple applications running environment. At present, the virtual infrastructure can improve x86 server utilization from 5%-15% to 60%-80%, and in tens of seconds to complete the resources allocation of new applied system [3]. Virtualization technology has been functioning well in PC and application server, which has produced a series of successful virtualized software like VMware, Virtual PC and Xen. The hardware manufacturers is joining the ranks of virtualization, prompting the advancement of virtualization technology. The hardware includes VT-X in Intel, VT-i and VT-d technology, Pacific in AMD and so on.

HPC (short for High-Performance Computing) is a branch of computer science, focusing on parallel computing and developing software, being committed to developing high performance computers [4]. With the rapid development of information society, HPC has become the third pillar of scientific research, following after the theoretical science and experimental science. In some emerging disciplines, such as new materials technology and biotechnology, high performance computers have become an indispensable tool for scientific research. With further research and increased competition, HPC is adopted to solve scientific and practical problems in production more and more. The development of HPC application not only prompts the innovation of science and technology, but the progress of society. The level of HPC's application is becoming a key indicator in measuring a nation's comprehensive national strength and international competitiveness.

III. THE POSSIBLE APPLICATION OF VIRTUALIZATION TECHNOLOGY IN HPC

With the development of HPC and further study of virtualization technology, the combination of these two is bound to become closer and closer. It is possible for virtualization technology to provide the difficulties in HPC with new methods [5]. Viewing from the respect of application, the combination of these two may appear in the following several forms:

A. Improve development efficiency.

HPC applications are closely related to computing environments and the behavior under the operation scale of different circumstances appears different. With

Yan junhao (1980-), Male, Han, Jiaozuo Henan, master, lecturer, Research area: Computer Network, DataBase Technology.
Project number: 2010 YSR Fund of HPU No. Q2010-54

virtualization technology, large-scale virtual application development environment can be built on small-scale systems, so that application programs can be developed and optimized under the environment which is much closer to the final system environment, and application programs can be transported to run in large-scale systems more quickly an easily.

B. Integrating heterogeneous resources

With the development of the application requirements and HPC, high performance computer systems are gradually developing towards the heterogeneity. How to manage and use heterogeneous systems efficiently is the major technical problem that HPC system software and application program developers face. Virtualization technology holds inherent advantages in integrating heterogeneous resources. Making use of virtualization technology to abstract and manage the underlying heterogeneous hardware resources can effectively hide the heterogeneous characteristics of the hardware platform and provide users with a unified system environment, making it convenient to use heterogeneous systems.

C. Providing customized Appliance

Different HPC application programs require different system environments, such as application-optimized operating system environment, a specific version of the compiler and the communication library, etc., which takes too much time and effort in application deployment; the system environment is also difficult to optimize, and system performance are not full. The adoption of virtualization technology contributes to the resolvement of the problems mentioned above. Virtualization supports packaging customized operating system in advance and optimized application running environment with the binary code into the VM (short for Virtual Machine) image, which is known as appliance. Through the direct deployment of Appliance, the fast deployment of HPC application programs can be realized and better performance can be reached.

D. Improving reliability and fault tolerance of the system

With the constant expansion of HPC, the continuous complexity among different parts, the failure rate of the system hardware is also growing. Checkpoint is often used to solve this sort of problems, keeping the intermediate results and then restarting the system. Checkpoint is traditionally made in the way of programming the code by users themselves. Because it would involve a number of border issues, there is higher requirement and a big challenge for users. By adopting virtualization technology, the state of a virtual machine can be well preserved, providing us with a clean border. Besides, in the virtualization system, due to isolation of each VM in the Q nodes, software errors, such as operating system or application failure only directly affects one VM, or even hardware failure, such as CPU, Memory, and equipment failures only affect the VMs they are assigned to. When a failure appears, it can be restored quickly through VM migration, restarting the VM and other methods, even without interrupting running applications.

E. Improving the security of HPC systems

System security is very important to HPC applications like the data center, and the isolation among VMs and self-testing capability provides a platform for establishing security system. As VMM((short for Virtual Machine Manager) only provides the abstraction and management of underlying hardware with some simple functions, it is more reliable and secure as opposed to a full-featured operating system kernel. VMM is not subject to interference from malicious code; the isolation between the VMM and authorized self-test feature are fully credible. VMM can check the credit of the loaded VM and the application programs loaded onto the VM, and it can authorize a VM to examine the VM state, such as scanning for viruses. Besides, VMM can also monitor the communication and state between VMs to ensure its correct running.

Only from the perspective of applications, the combination of virtualization technology and HPC may appear these several form of possibilities. In fact, high-performance computing hardware itself, especially the rapid improvement of processor performance, makes it more probable for virtualization technology to enter the HPC, functioning as a "good wife". In the past, CPU utilization of HPC is almost 100%, but with the latest Intel and AMD processors launched, the performance of HPC has been improved unprecedentedly. As the number of single-core CPU increases, even if in a single computing node, the application will not necessarily occupy all the Core (core). Thus, the remaining core can meet some program application whose requirements of the I/O and bandwidth are not particularly high [6]. While how these applications can be put onto the core at idle and isolated correspondingly, how to ensure these operations do not conflict with the original, for which virtualization technology can provide a better solution. With the rapid development of Intel's VT technology, it is turned into reality to improve the IT structure and enhance the value of IT to become by adopting virtualization technology.

IV. THE CHANLLENGES TO FACE AND THE STUDIES TO CARRY OUT

Although, virtualization technology may bring great possibility to increase the value of HPC, by farther adoption of virtualization technology in HPC is scarce, which is mainly due to the following several aspects:

F. Performance overheads brought by virtualization

Traditional server virtualization brings extra performance overheads. In a virtualization system, VMM run at the highest privilege level, and VM and Guest 0S run as the user-level VMM. This leads to the fact that in running, Guest OS must be embedded into VMM when faced with the privilege operation. This approach requires for the implementation of context switching, and would result in longer delay in accessing devices, which are

unacceptable for HPC applications that are sensitive to the system performance. Hence, the virtualization technology which is just intended for HPC system needs developing and it is necessary to optimize VMM technology based on the requirements of HPC.

G. The efficient coordination of many VMMs

The virtualization technology intended for HPC system is different from the traditional server virtualization technology. In traditional server virtualization, only one single VMM is needed to abstract the underlying hardware. While in the multi-dimensional heterogeneous HPC system, one separate VMM is needed in every node and this VMM only virtualizes that single node. In the whole system, a large number of interrelated VMMs run, and they also work in coordination, forming a unified virtualization environment of large-scale system level. Therefore, it is necessary to study efficient VMM coordination mechanism, including the technological problems like the coordination management of a large number of VMMs, coordination deployment, the efficient communication of VMs across physical nodes, VM migration, and so on.

H. The management of a large number of VM

In order to support the operation of HPC program, it may be necessary to deploy thousands of VMs at one time. While, traditional server virtualization technology can tackle the deployment and management of a small number of VMs. Thus, how to support the dynamic deployment of a large number of VMs, how to allocate the necessary hardware resources fast according to application requirements, how to start VM quickly with lower system overhead, how to manage a large number of VMs in operation, all of which are the important technical issues to achieve HPC virtualization.

I. Programming model and the support of software environment

The traditional programming model and the software environment supporting application development and operation are all directly intended for non-virtualization system; while, virtualization technology abstracts hardware system, changes the organization morphology of resources users see, so that the traditional programming model and software environment is in no way able to meet users' requirements about virtualization system. Hence, it is necessary to develop new programming model directed at virtualized HPC system and the corresponding optimized software environment intended for virtualized system, such as parallel compiler, linker, debugger optimization tools, parallel libraries, etc.

In a word, the application of HPC usually requires high performance; while on the other hand, virtualization will definitely lead to the damage of performance, which is almost unacceptable to most of HPC application. Therefore, it is necessary to develop efficient virtualization technology particularly intended for HPC system, such as the virtual model for high-performance computer systems, application development and deployment mechanism on large-scale virtualized systems, programming model and corresponding software environment supporting virtualization system [7]. With the continuous development of virtualization technology, the underlying hardware provide virtualization with increasing support, lowering the performance overhead of virtualization [8]. The trend of the high-performance computer architecture developing towards a multi-level, multi-granularity isomerization will further promote the development of virtualization technology.

V. CONCLUSION

To some extent, what hasn't been turned into reality is of certain possibility. Though analyzing the application prospect of virtualization technology in HPC, the chances and challenges virtualization may face in HPC are emerging easily. It is believed that HPC in the future will set a stage for virtualization. By discussing these chances and challenges as well as constant examination of the practice, both virtualization and HPC can be improved greatly. Meanwhile, the value of IT will increase gradually with the improvement of technology.

ACKNOWLEDGMENT

Thankful to 2010 Youth Science Research Fund of Henan Polytechnic University for its support of the project and its affilicated article.

REFERENCES

[1] The Disputes of virtualization and High Performance Computing, http://server.51cto.com/HPC-98598.htm

[2] Opportunities and challenges of virtualization in high-performance computing, http://tech.sina.com.cn/b/2009-07-07/14563242377.shtml

[3] VMware, History of Virtualization, http://www.vmware.com/virtualization/history.html

[4] WIKIPEDIA, "High-performance computing", http://en.wikipedia.org/wiki/High-performance_computing

[5] Timothy Prickett Morgan,Virtualization and HPC - Will they ever marry?

[6] Advanced Micro Devices. AMD64 Virtualization Codenamed \Pacica" Technology, Secure Virtual Machine Architecture Reference Manual, May 2005. http://www.theregister.co.uk/2008/11/21/virtualization_and_hpc/

[7] John Paul Walters, A Comparison of Virtualization Technologies for HPC, http://www.computer.org/portal/web/csdl/doi/10.1109/AINA.2008.45

[8] Khalid, Use of Server Virtualization in HPC Environments, http://www.hpccommunity.org/blogs/khalid/use-server-virtualization-hpc-environments-84

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 245-248

Multiplex Transmission of Data and Video Signals in Fiber Optic Communication System

Zhang Chang-sen[1], Zhang Ming-ke[2]
[1] School of Computer Science & Technology, Henan Polytechnic University, Jiaozuo, china
Email:zhangchangsen@hpu.edu.cn
[2] School of Computer Science & Technology, Henan Polytechnic University, Jiaozuo, china
Email:mingkemolly@sina.com

***Abstract*— A new transmission scheme is given for cutting costs, transmitting signals as much as possible and improving performance. This paper begins with a discussion of the characteristics of FPGA, A/D and D/A converter and HP1032/1034, followed by under the premise without increasing transmission bandwidth, five video signals and sixty data signals are transmitted simultaneously. A new method to solve this problem is presented, which is simple and practicable. Specific method is that each one video signal and twelve data signals are attached by FPGA, which is equivalent to using the bandwidth of one video signal without compression, completing one video signal and twelve low-speed data signals composite transmission. Then twice multiplexing/de-multiplexing is used, to achieve multiplex transmission of the digital video and data in a fiber. The results prove that the system with high quality, anti-jamming, is more suitable for long distance video transmission monitoring system.**

***Index Terms*— FPGA, HP1032/1034, TDM, fiber optic communication**

I. INTRODUCTION

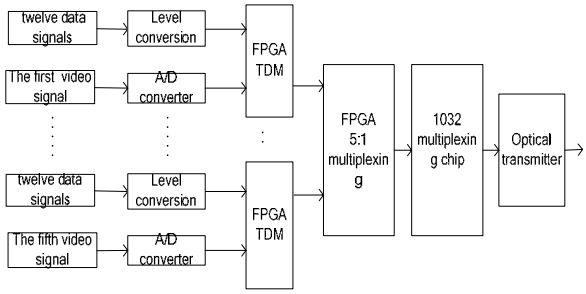
Recent years we have witnessed a very rapid growth of fiber optic communication. Nowadays it has become one of the main means of communication. Fiber optic communication falls into analog and digital fiber optic communication. The shortcomings of analog fiber optic communication are serious signal distortion, low transmission quality, unstable system performance and simultaneous transmission of multiple signals prone to producing mirror image and cross interference. So now we often use digital fiber optic communication .It has several advantages over analog fiber optic communication, which are low loss, wide bandwidth, electronic magnetic in EMI, good privacy and long lifetime.

In the actual industry application field, people often need to transmit multiple video image signals and multiple data signals in the meantime. But in the uncompressed video transmission system, it is not easy to complete the simultaneous transmission of multiple video and data signals. For example, a system transmits two video signals and twenty-four data signals. With using 12-bit code and plus twenty-four data signals, there will be forty-eight signals to transmit. Taking HP1032/1034 as an example, this chip can expand seventeen data signals multiplexing/ de-multiplexing transmission at most. At present of all the multiplexing /de-multiplexing chips, there is almost no 48 I/O multiplexing /de-multiplexing chip. If through increasing the number of fiber to complete the simultaneous transmission of multiple signals, the costs will increase too. By our use of the TDM (time division multiplexing) and WDM (wavelength division multiplexing) technologies, each one video signal and twelve data signals are attached by FPGA. Then five video signals and sixty data signals are transmitted simultaneously. This is accomplished by twice multiplexing and de-multiplexing.

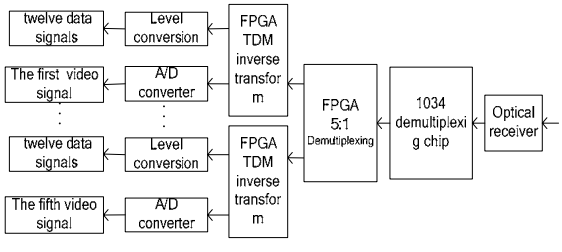
II. SYSTEM FRAMEWORK AND WORKING PRINCIPAL

The whole system is composed of the FPGA core control module, video A/D and D/A conversion module, 1032/1034 serializer/deserializer chip and optical transmitter/receiver. The system framework of transmitting circuit and receiving circuit are shown in Fig.1.

The working principal of the system is as follows. Optical transmitter and receiver system falls into transmitter and receiver. After the A/D (analog/digital) conversion of five video signals from the transmitter, one analog video signal is converted to twelve digital signals, and five analog video signals are converted to 60 digital



(a) Optical transmitter



(b) Optical receiver

Figure 1. The system framework

signals, which are sent into FPGA. In the meantime, 60 low-speed data signals are also sent into FPGA after being converted to TTL level. Through the designed logic module, completing multiplexing of data and video each twelve data signals are attached to one video signal. So it becomes 60 signals. Followed by to achieve 5:1 multiplexing, 60 signals are changed into twelve signals. The twelve signals, one data synchronization signal, one video synchronization signal enter HP1032 chip to complete the second multiplexing. They are changed into one serial signal, which is sent to the receiver by fiber. At the receiving end, through the reverse process of the above, it completes restoration of the original signal.

III. HARDWARE COMPONENTS

A. FPGA control device

The control part of FPGA is the center of the whole system. In order to meet the high-speed, multi-stream real-time processing, requiring the core controller of the system must have high frequency and response capabilities [4]. As embedded systems, processor must be low power. At present, FPGA processors have control over the industry, consumer electronics, communication systems and other kinds of market. With the low-cost, low power consumption, small size, multifunction and more powerful data processing capability, EP1C6Q240C8 of Altera is a very good choice.

B. A/D and D/A converter

In this experiment, one video signal is converted into twelve digital signals, we choose AD1674. The AD1674 is a complete, multipurpose, 12-bit A/D converter, consisting of a user-transparent onboard sample and hold amplifier (SHA), 10 volt reference, clock and three-state output buffers for microprocessor interface. The AD1674 is compatible with the industry standard AD574A and AD674A, but it includes a sampling function while delivering a faster conversion rate. The on-chip SHA has a wide input bandwidth supporting 12-bit accuracy over the full. Its key features are that it has complete monolithic 12-bit 10ms sampling ADC; it has on-board sample-and-hold amplifier; it has industry standard pin out; it has 8-bit and 16-bit microprocessor interface [5]. Block diagram of digital video is shown in Fig.2.

The A/D converter circuit at the receiving end mainly completes the conversion of digital video signal to analog, and reduces the standard video signal. The DAC used here is ADI's high-speed AD9708. The AD9708 is the 8-bit resolution member of the TxDAC series of high performance, low power CMOS digital-to-analog converters (DACs). The TxDAC family, which consists of pin compatible 8-, 10-, 12-, and 14-bit DACs, is specifically optimized for the transmit signal path of communication systems. All of the devices share the same interface options, small outline package and pin out, thus they provide an upward or downward component selection path based on performance, resolution and cost. The AD9708 offers exceptional ac and dc performance while supporting update rates up to 125 MSPS. At the receiving end, digital video is reduced to analog video which is shown in Fig.3.

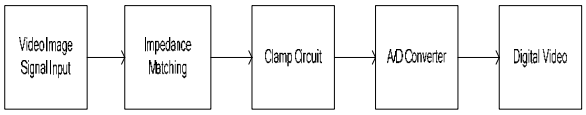


Figure 2. Digital video

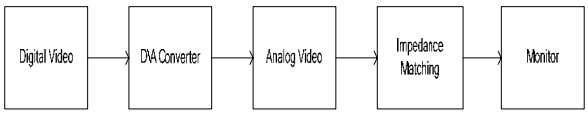


Figure 3. Block diagram of receiver

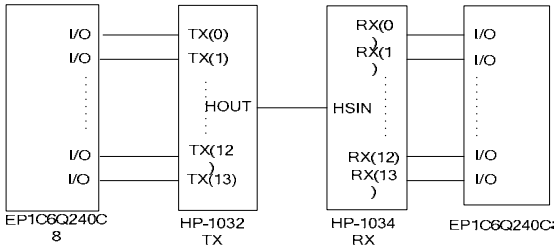


Figure 4. Interface connection

C. HP1032 / HP1034 serializer / deserializer chip

HP1032 transmitter (TX) and the HP1034 receiver (RX) are general-purpose gigabit serializer/deserializer chip used to build point to point data link. They provide users with a parallel digital signal transmission "virtual flat cable" interface; Parallel signal loading TX, passed to the RX through a serial channel, serial channel can be coaxial cable or optical fiber; RX data recovery in the original parallel format. They hide their code, multiplexing, clock extraction, de-multiplexing and decoding complexity on the user. Mainly used in backplane connecting, bus control, digital video transmission, dedicated high-speed data transmission and so on. Its key features are that it has standard TTL interface, the signal width of 16 or 17; it has high-speed serial baud rate of 260M~1400M (User-selectable); it has on-chip encode/decode; it has transmitter/receivers on-chip phase-locked loop, to provide frame synchronization; it has alternative flat signal cable. HP1032/1034 and FPGA interface connection is shown in Fig.4.

IV. SYSTEM FUNCTION

D. Time division multiplexing function based on FPGA

Using the characteristics of the video image signal, low-speed data can be attached to the video signal for transmission, which is equivalent to using the bandwidth of one video signal without compression, completing one video signal and twelve low-speed data signals composite transmission. One video signal through A/D is transformed into twelve digital signals which are sent to FPGA. At the same time, twelve low-speed data signals

are converted into TTL standard level digital signals and are also sent into the FPGA. Through logic module that has been designed, it completes the multiplexing of data signals and video signals.

As mentioned above, what system needs to complete is the low-speed data is attached to the video signal for transmission, which depends on the characteristics of the video signal. Composite video signal not only has image information but also has line synchronization, line blanking, field synchronization and field blanking. Line synchronization is 15625 per second and field synchronization is 50 per second. So the cycle of line synchronization is 64 μs . It is feasible that low-speed data is transmitted in appearing synchronous head. Because of level clock of synchronous head staying the same, it is not only beneficial for control sequencing, but also does not occupy image information transmission time slot. SNR of the system will not reduce because of multiplexing and it is possible to transmit low-speed data in the cycle of 64 μs [3].

E. 5:1 multiplexing function based on FPGA

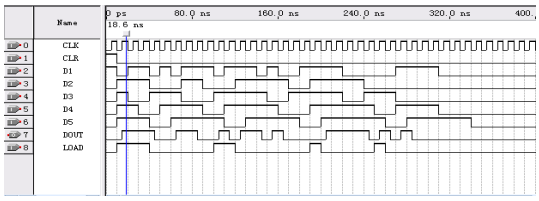
One analog video signal by the A/D converter will encode to twelve digital video signals, such as D1(1), D1 (2) D1(3) ... D1(12). Five analog video signals will produce 60 digital video signals. To achieve 5:1 multiplexing function, the first bit D1(1) of the first digital video signal after the A/D conversion, the first bit D2(1) of the second digital video signal after the A/D conversion, the first bit D3(1) of the third digital video signal after the A/D conversion, the first bit D4(1) of the fourth digital video signal after the A/D conversion and the first bit D5(1) of the fifth digital video signal after the A/D conversion conduct 5:1 multiple connection; the rest can be done in the same manner, the twelfth bit D1 (12)of the first digital video signal after the A/D conversion, the twelfth bit D2(12)of the second digital video signal after the A/D conversion, the twelfth bit D3 (12)of the third way digital video signal after the A/D conversion, the twelfth bit D4(12)of the fourth digital video signal after the A/D conversion and the twelfth bit D5(12)of the fifth digital video signal after the A/D conversion conduct 5:1 multiple connection, and then repeating the multiplexing [2] [6].

Multiplexing process is mainly that the parallel signals are converted into serial signals. Transition between parallel signals and serial signals circuit is composed of 5-bit parallel in / serial out shift register.

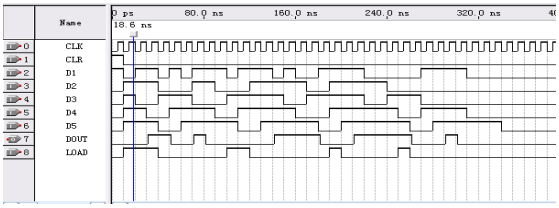
In Fig.6, clk，clr，D1…D5，dout and load are the clock signal input, clear end, data signals input, data signals output, signal load respectively. Using quartus Ⅱ simulation software, simulation results are shown in Fig.5.

F. 1:5 de-multiplexing function based on FPGA

De-multiplexing design is that after inverse transform of the transition between parallel signals and serial signals one serial signal separate five parallel signals by 1:5. Mainly serial in/parallel out shift register realizes. Serial signal transmitted from sending termination is received by receiving termination and under the action of cp clock edge it moves back step by step. After reaching to five bits, it is output in parallel.



(a) Functional simulation



(b) Timing simulation

Figure 5. 5-bit parallel in / serial out shift register simulation results

The 5-bit serial in /parallel out shift register circuit symbol with synchronous clear is shown in Fig.7. clk is the clock signal input, din is data signals input, clr is clear side, Dout[4...0] is data signals output.

The VHDL code of 5-bit serial in /parallel out shift register with synchronous clear [1]

```
Library ieee;
Use ieee.std_logic_1164.all;
Use ieee.std_logic_unsigned.all;
Entity sipo is
Port(clk:in std_logic;
din:in std_logic;
clr:in std_logic;
   Dout:out std_logic_vector(4 downto 0));
End;
Architecture one of sipo is
  Signal q:std_logic_vector(5 downto 0);
begin
Process(clk)
begin
If clk'event and clk='1'then
 If clr='1'then q<=(others=>'0');
Elsif q(5)='0' then
 Q<="11110" & din;
Else q<=q(4 downto 0)&din;
End if;
End if;
End process;
Process(q)
Begin
If q(5)='0'then
    Dout<=q(4 downto 0);
Else dout<="ZZZZZ";
End if;
End process;
End;
```

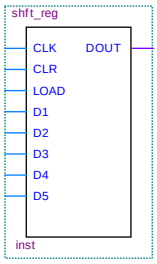


Figure 6. 5-bit parallel in/serial out shift register circuit symbol

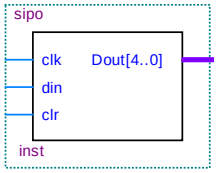
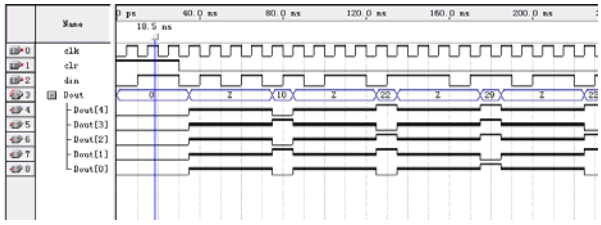
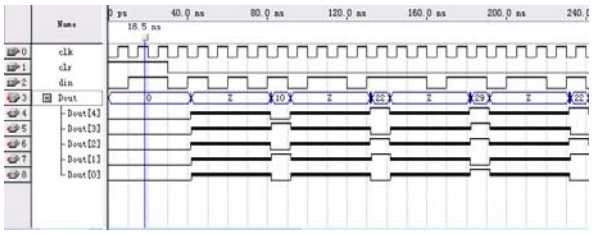


Figure 7. 5-bit serial in /parallel out shift register circuit symbol



(a) Functional simulation result



(b) Timing simulation result

Figure 8. 5-bit serial in/ parallel out shift register with synchronous clear simulation results

The timing simulation results are shown in Fig.8. Every 5-bit serial input signal is a group. Set six shift register to form serial in/ parallel out shift register, five of which are used for displacement and storage data in series, one of which is as a symbol used to record whether all five data move in register. Once the shift registers detect the five data all moved in, five data signals are output in parallel immediately. When the serial data is in the process of moving into the registers, parallel output signals of the shift register maintain a specific value. In this design, it is set high impedance.

From the waveform we can know that, when five data signals of the din all move into the shift register, the first five data signals of din are output. When we enter '01010', the output is '01010' which is converted to unsigned decimal, and it is '10'. When we enter '10110', the output is '10110' which is converted to unsigned decimal, and it is '22'.

V. CONCLUSIONS

This article describes the use of programmable logic device to complete the simultaneous transmission of multiple video image signals and multiple data signals. The approach is using TDM capabilities of FPGA to achieve one video signal and twelve data signals composite transmission, using shift register to complete the design of 5-bit serial in/ parallel out, using HP1032/1034 serializer/deserializer chip to complete the design of two times' multiplexing/de-multiplexing. In the end we complete the simultaneous transmission of multiple video image signals and multiple data signals. In this paper, block diagram and realization of simulation software quartus Ⅱ are given. The results obtained are in good agreement with the calculated values. The system with the feature of high quality transmission, strong anti-jamming ability is more suitable for long distance video transmission monitoring system, which can be used in the video surveillance systems and security systems.

REFERENCES

[1] ZHOU Run-jing, Tuya, ZHANG Li-min, Based on QuartusΠ the FPGA / CPLD digital system design example, Beijing:Publishing House of Electronic Industry, Aug.2007，pp.190-192.

[2] Li Quan, Wen Ying, "Using FPGA and serializer/deserializer chip HP1032/1034 the design of digital video multiplexer / demultiplexer," Electronics, Feb. 2003, pp.43-46.

[3] BAO Jian-xin, WANG Cheng, CAO Jia-nian, "Using TDM to realise multiplex transmission of the data and video signals in fiber optic communication systems," Applied Science and Technology, 2004, pp.28-30.

[4] Li Liang, HU Yi-liang, HAN Rui-zhen, "Design of Fiber Optic Transmission System Based on FPGA," CHNA DIGITAL CABLE TV, 2008, pp955-956.

[5] Chen Lan-gu, "Multiple digital video fiber optic transmission system design based on FPGA," East China Normal Univeisity,2009.

[6] Zhang Chang-sen, HUANG De-xin, "The design of video optical transmission and control based on FPGA and CY7B923/933," Optical Communication Technology, vol.12, 2009.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 249-252

Design and Implementation of a Novel ActiveLow-Power RFID Tag

Su Yu-na, XuYan-ping
School of Computer Science and Technology, Henan Polytechnic University ,Jiaozuo, Henan
Suyuna0296@163.com, xubaby@hpu.edu.cn

***Abstract*—The active radio frequency identification (RFID) tag with far reading distance, high reliability, low cost, low consumption, and long life is designed use the latest low power-consuming single-chip MSP430F2012 and wireless data transmission chip CC1100, and the hardware and software implementation are presented in this paper.**

***Index Terms*—RFID, low consumption,active tag**

I. INTRODUCTION

As a new kind of automatic identification technology, radio frequency identification (RFID) without contact, the fundamental principle is to realize the automatic recognition of the static or moving objects by radio frequency （RF ）signal and the spatial coupling and transmission characteristics. Usually, RF system consists of RF tag, reader and host. At present, most RFID systems adopt passive tags which get power from the reader by RF signal, it is beneficial to reduce the label size and cost, but the reading range and data storage capacity are limited. While the active tag with battery can provide larger range of reading ability and higher reliability. Now the breakthrough of low power consumed IC technology created favorable conditions for the development of small size, active and low consumed tag. The design aimed to implement active and low consumed tags based on MSP430F2012 and CC1100 with the lowest possible hardware cost.

II. DESIGN ANALYSIS

Since active RFID tags use the battery as supply power, system is very strict with low consumption performance to prolong the service life of battery. Low-power design requires both the choice of components and optimized reasonable run timing, most of the time keeps the circuit in standby mode on the premise of completed function tags.

A. Choice of MCU

MSP430 series MCUS are 16-bit ones & Flash-type that of ultra low power consumption, with low supply voltage, small leakage current of I/O port, has 0.5mA of standby current and 250mA/MIPS of operate power consumption, which becomes a recognized low consumption microcontroller in the industry. MSP430F2012 MCU is adopted in this paper.

B. Choice of RF chip

RF chip choose is the most crucial part of the RFID card, it directly related to tags' read range and reliability, but also the power consumption. Wireless transmitter CC1100 with small size, low consumption, supports programmable control; with internal address decoder, modulate processor, clock module and so on, is very easy to use. The design selects CC1100 with operating frequency of 915MHz, FSK modulation, data rate of 100kbps.

C. Battery supply

Voltage regulator circuit between battery and device is omitted in this paper, battery is used as supply power directly. It saves quiescent current brought from voltage regulator circuit, prolongs the service life of battery. To adopt battery as supply power, the key point is to solve the random wrong operation because of incomplete reset, which resulted from mechanical contact with the battery wires will produce power supply noise when replace the battery. While brown-out reset (BOR) is integrated in MSP430F2012, full reset will be implemented when voltage below the safety operate range, this function can solve the above problem better.

III. HARDWARE DESIGN OF TAG

A. SystemStructure of Tag

The active tag should have the following characteristics: miniaturization, low cost, low consumption, high reliability, adjustable reading, distance battery-powered, and so on. The block diagram of the tag is shown in Fig.1.

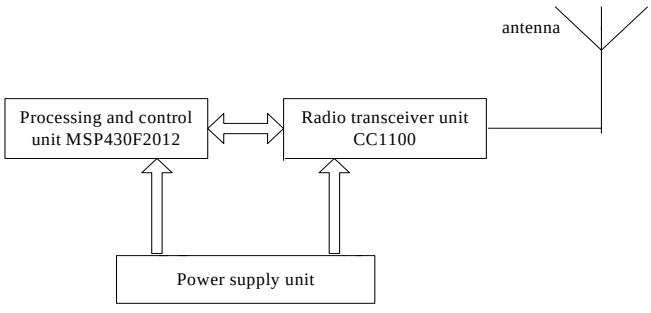


Figure 1 Block diagram of system structure

RFID tag consists of processing and control unit, radio transceiver unit and power supply unit. Of these, the processing and control unit with its own memory, which is responsible for the operation of RFID tag, data deposition and processing. The radio transceiver unit contains RF chip and antenna, to achieve information transmission between active tag and reader. The power supply unit supply power for RFID tag.

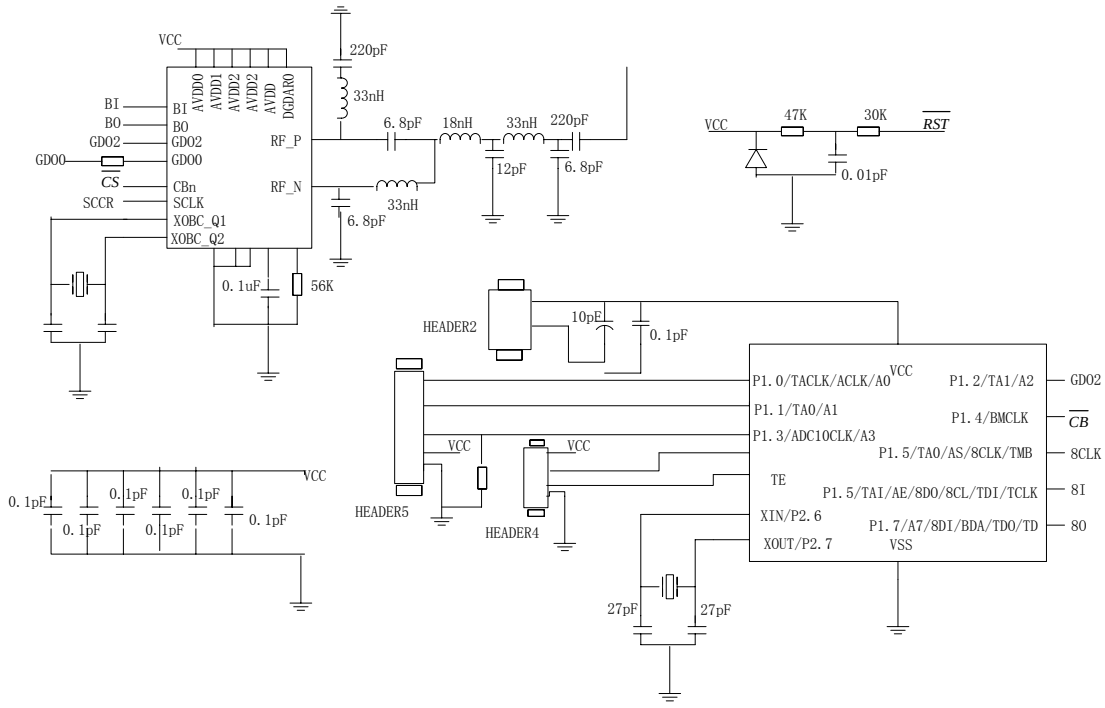


Figure 2 Schematic circuit diagram of hardware of MSP430F2012 control CC1100

B. Design Principle Diagram

The circuit principle diagram of active low consuming tag is shown in Fig.2.

IV. SOFTWARE DESIGN OF TAG

A. Design of Low-Consumption Workflow

The software development of system uses C language on Instruction Address Register (IAR) Embedded Workbench V3.41A to program, conflict of running speed, data flow, system performance and low consumption design is solved through combining intelligent operation management of MCU and CC1100 module with power save mode control, the current consumption of each functional module is reduced to the minimum, active state is limited to the minimum requirements[1]. After optimization, system can get very low power consumption, as well as higher performance. Specific procedure as follows:

The most important factor to reduce power consumption is the application of maximize Low-power mode 3(LMP3) time of MSP430 clock system, therefore, MCU usually keep in LMP3 low-power mode, while after the initialization configuration CC1100 will be in standby mode.

a) External 32 KHz crystal is used as ACLK, which also is the clock source of timer. It not only reduces power consumption, but also ensured stability and accuracy of the timer[2].

b) If produced receive or send interrupt, Single Chip Micyoco is activated to control CC1100 operate.

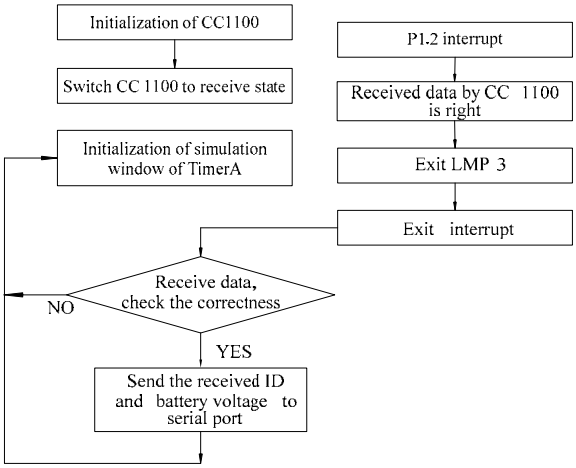


Figure.3 Workflow chart of CC1100

c) In order to prevent MSP430 'Run away' caused by external interference affects the digital device, internal watchdog module is started, ensure system to keep running normally.

Fig.3 is the workflow chart of CC1100 in RFID system.

B. Realization of Driver

1) SPI driver of MSP430

MSP430 communicated with CC1100 by standard Serial Peripheral Interface (SPI), which includes two data lines—SPI Bus Master Output/Slave Input（MOSI）and SPI Bus Master Input/Slave Output （MISO）, and the clock line CLK, which adopted by master to keep clock synchronization with slave. SPI is divided into master

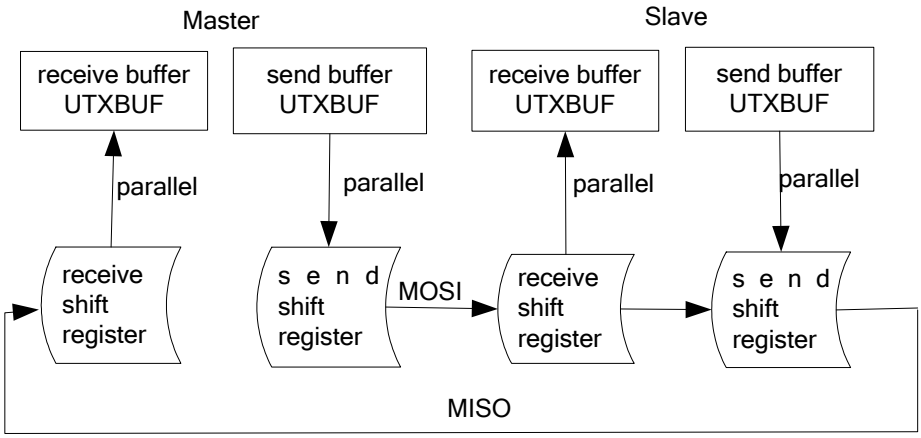


Figure 4 Schematic diagram of SPI

mode and slave mode, slave mode is completely passive mode, data dispatching is host control. Actually, there are four registers take part in work, as shown in Fig.4. Host write data to UTXBUF of send buffer, store data in parallel to send shift register. Once data are written in UTXBUF, immediately shifted from MOSI line to the receive shift cache of slave, data in the shift register of slave is shifted to the receive shift cache of host through MISO shift, and then read in parallel to the receive cache[3]. It means that SPI can not only read data but also write data.

Part of the code is as follows:

```
//achieve reading and writing, CC1100 specified the value of address at the same time
Char_Spi_Read_Write(char data)
{
      // disabled SPI interruption
      IE1&=~UTX IEQ;
      IE1&=~URX IEQ;
      TRXBUFO=data
      //wait for data transmit and receive finish
 While(((IFG&UTXIFG0)==0)//(IFG &URXIFG0)==0));
    //return read value, can give up if not need
          Return RXBUFO;
    }
```

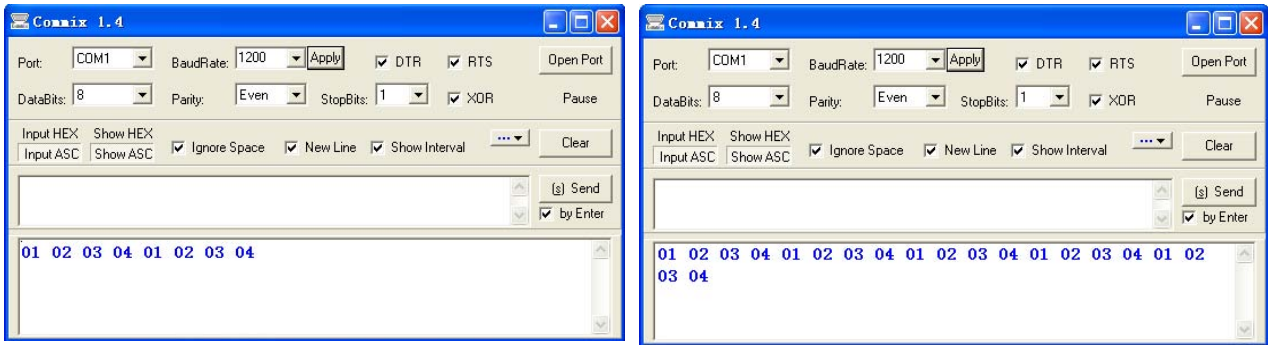
2) Drive of CC1100

There are two special configuration pins and one shared pin in CC1100, to output internal state information which is useful to control software. These pins can be used to generate interrupt on the MCU, the special pins named GDO0 and GDO1, the shared pin is SO in SPI interface. The default configuration of GDO1/SO is status output, which will be general pin by chose arbitrary control options[4] .When CSn is low-level, this pin is a general SO pin. Under synchronous and asynchronous continuous mode, GDO0 is used as a continuous TX data input pin in the transmission mode. Use software Smart RF Studio to get the optimal register settings, evaluation of performance and function[5].

V. SYSTEM SIMULATION

Combined with configured reader, we get that the program is correct through IAR online simulation, then burning the program by emulator, use commix1.4 of serial debugging tools to test, test results that send two times and five times are respectively shown in Fig.5 (a) and (b)[6].

Form the debug results in Fig.5, we know that when active tag is awakened, the ID number will be sent to reader, the reader sends it to serial debugging software of host computer by RS232 interface, to show the ID[7] . In Fig.5, (a) is the result received when send four active tags two times, while (b) is the result when send five times. The simulation and debug results proved the feasibility of RFID design.



(a) Send four active tags two times (b) Send four active tags five times

Figure.5 Result of debug active read-write system

VI. CONCLUSIONS

This design fulfilled the ultra-low power characteristics of MSP430 MCU, took measures to limit power to CC1100 module which consumes is lower by means of software, and increased the reliable operate time of system. Meanwhile, to a great extent, this active RFID tags solved such identification problems as long-distance, big flow, anti-interference, high-speed and reduced the cost. RFID tags designed in this paper and configured readers form persons or goods recognition and location system, which is widely used in mining, industrial production, road traffic, national defense security and so on.

REFERENCES

[1] GAO Tian-bao, WANG Jing-chao, ZHANG Chun, LI Yong-ming, WANG Zhi-hua. "Design and realization of a portable RFID reader," Application of Integrated Circuits, May, 2008.

[2] SUN Wei-ming, SHI Jiang-hong. "Wireless Sensor Networks Consists of MSP430 and CC1100," Microcontrollers & Embedded System, Aug. 2007.

[3] SHANG Liang, LI Wen-feng, LI Bai-ping. "Design of active low power RFID tag based on MSP430F2012 and CC1100," Electronic Component & Device Applications, vol. 20, 2008.

[4] HU Da-ke. "Principles and Applications of ultra-low power & 16-bit MSP430 family microcontroller," Press of Beihang University, Beijing, June 2000.

[5] LI Qiang, HAN Yi-feng, XIE Wen-lu, HE Hao. "Design of Ultra-low Power Bandgap Voltage Reference and Its PSR Analysis," Chinese Journal of Semiconductors, vol.25, 2004, pp.1474-1478.

[6] Liao Ping, Jing Huan-huan, He Zhen-wei Zhou Shou-qin. "Design of Data Acguisition System Based on RFID Technology for Yard," COMPUTER MEASUREMENT & CONTROL ,May,2009.

[7] Lai Xiao-zheng,Liu Huan-bin."Microstrip Parasitic Antenna on Papery Substrate for RFID Tag," JOURNAL OF SOUTH CHINA UNIVERSITY OF TECHNOLOGY, May,2008.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 253-255

Local Adaptive Image Enhancement Based on HSI Space

Sima Haifeng[1], Liu Lanlan[2]
[1]College of Computer Science Technology, Henan Polytechnic University, Jiaozuo, China
E-mail: smhf@hpu.edu.cn
[2]College of Emergency Management, Henan Polytechnic University, Jiaozuo, China
E-mail: liulanlan@hpu.edu.cn

***Abstract*—Based on the color perception characteristics of eyes, the paper proposed nonlinear transformation of color image enhancement algorithm by the various components of the HSV color space. Conversion original image space from RGB to HSI color space, Extract H, S, I components, on which the color and saturation enhancement processing component with local filtering, and then Synthesis the HSI image and converted to RGB space, The results show that the algorithm can improve the color distortion, increase the recognition of color images, raise the information clarity of image.**

***Index Terms*—transformation, HIS space, enhancement, convolution**

I. INTRODUCTION

Visual perception system will be restrict by many factors, such as lower color image resolution, insufficient brightness , some local image difficult to identify etc. So we need to resolve these problems. we need to take enhancement processing on image, to improved recognizable of image, to highlight the different objects in original image, to improve the resolution effect, lay foundation for the pattern recognition.

Image enhancement is a key step in image processing technology, the theoretical study and practical application have been one of the wide attention. Image Enhancement and there are many kinds, some enhanced operations can be directly applied to any image, while others only apply to specific types of images. Some algorithms need to enhance the image in crude approach, because they need to extract more information from the image. It is adversely that there is no single standard for enhancement. Many different types of image or scene can be enhanced as the image data, Different types of images, has been corresponding to enhancement of its own feature, Some enhancement is only suitable for certain special types of image enhancement. Enhancement results depending on the occasion and requirements of measurement. Image enhancement is the key steps from the image processing to image analysis. So image enhancement results have a direct impact on the image understanding. In recent years, researchers put forward variety of algorithms for the enhancement of color images. Most of the algorithms are achieved in RGB color space. First segregated the image into RGB space enhancement three components, and then use some logic algorithm will combine the three components of the edge. However, the three RGB components are highly related in color image[1]. such as when the light is changed, RGB three components will simultaneously change. In RGB space, many mature approaches used vector space method. The main idea is to image each pixel is the RGB space as a three-dimensional vector, then the whole color images is considered as a two-dimensional with Three components vector field. The main disadvantage of RGB color space of color perception is not uniform, that is, the distance between the two colors is not equal to the color perception between the two color differences cannot be directly estimated from the RGB color value in color , saturation and brightness perception properties. That the distance between the two pixels is not equal to the color differences in color perception , It is difficult to estimate the color, saturation and brightness perception properties from the RGB color values[2]. To overcome shortcomings of the RGB color space, we can select color space with color visual properties in color image processing. HSI color space is one of a kind, it uses color (Hue), saturation (Saturation) and brightness (Intensity) to characterize the three components of color. Each component is independent with others, and agree with the humans feel. Measurement of HSI color difference is the important question in enhancement, but there is not much research on this issue[3-4]. In this paper proposed nonlinear transformation of color image enhancement algorithm by the various components of the HSV color space. conversion original image space from RGB to HSI color space, Extract H, S, I components, on which the color and saturation enhancement processing component for local filtering, and then Synthesis the HSI image and converted to RGB space, The results show that the algorithm can improve the color distortion, increase the recognition of color images, raise the information clarity of image.

II. COLOR SPACE CONVERSION

HSI color space is based on the human visual system, with the color (Hue), color saturation (Saturation or Chroma) and brightness (Intensity or Brightness) to describe the color space. HSI color space can be describe by a conical space model (Fig. 1).

HIS color space the cone model is very complicated, however it really can express the variations clearly of

color, brightness and color saturation. A lot of algorithms used HSI color space in the image processing and computer vision space , they can be dealt with separately and are independent of each other. Therefore, in the HSI color space can greatly simplify the image analysis and processing.

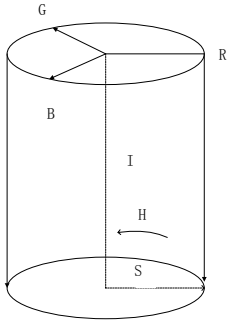


Figure 1 HIS space

Conversion formula from RGB to HIS space is show as follows:

$$\theta = \cos^{-1}\left[\frac{\frac{1}{2}\left[(R-G)+(R-B)\right]}{\sqrt{(R-G)^2+(R-B)(G-B)}}\right]$$

$$H = \begin{cases} \theta & G \geq B \\ 2\pi - \theta & G < B \end{cases}$$

$$S = 1 - \frac{3\min(R,G,B)}{R+G+B}$$

$$I = \frac{1}{3}(R+G+B)$$

III HIS ENHANCEMENT

HSI color space of the three components constitute the basis of the image. In the same resolution, the display quality of image depends on the distribution of the overall contrast and color difference calculation. We first deal with the brightness component I, After analyzed the whole image, according to the distribution characteristics of I uniform treatment. Calculated as follows.

Firstly, compute the overall brightness mean value M of the Intensity. if M is greater than a threshold value of brightness means meet the basic requirements, or to enhance the brightness . In order to increase color and contrast resolution, to meet the requirements for the component luminance image I refer to the Gaussian function mapped to a more reasonable space, which makes the brightness level changes in the brightness of the two is more obvious. The objective space brightness range is [0-255].

$$f(x) = \frac{1}{\sqrt{2\pi}\,\delta} e^{\frac{(x-\mu)^2}{2\delta^2}}$$

To examine S and H component's features, these two components are a pair of orthogonal variables according to the transformation from the RGB. While H and S are consistent with characteristics of visual perception, we use linear function on H and S to be and the corresponding color space to improve the resolution results.

Adjustment on S and H under the context of I component. While H is relative stabilization，so do not take it into account. components S of the processing using the look convolution function image is divided into D*D non-overlapping region A (A1 ... AN). First test for the border region of Ai, to compute chroma means of both sides of the subarea according to the border. If the difference of two subarea is more than threshold value P, adjust saturation, else the color difference is small, does not change the saturation value. During the detection of the border using CANNY[5] Gradient operator for sub region boundary detection.

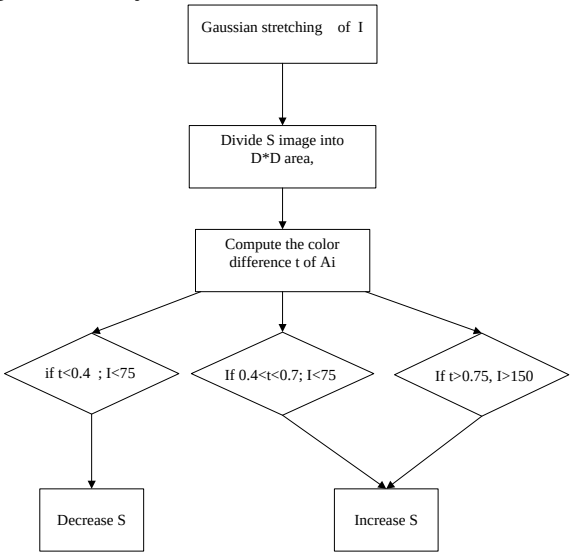


Figure 2 Enhancement process of H,S,I

IV EXPERIMENT ANALYSIS

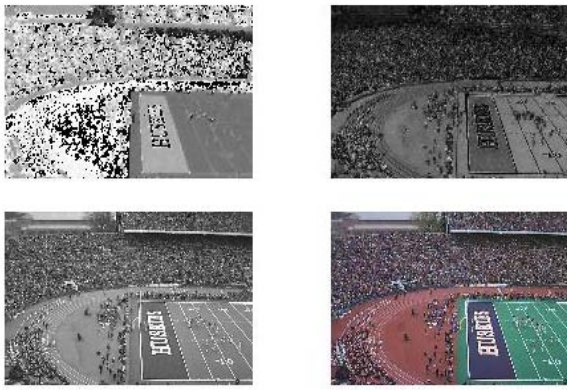


Figure 3 the HIS Components after Enhancement

This method was programming in matlab7.0. It used different types of scene for detection enhancement , were randomly selected for enhanced effect is remarkable, compared to the same state filtering algorithm, to save computing time, and that does not lose some image detail, can be applied to more sophisticated image analysis and processing. The Fig. 3 show the HIS

components of the HIS space after enhancement, the image is from NASA databases. and Fig. 4 show the Synthesis image and converted to RGB space.

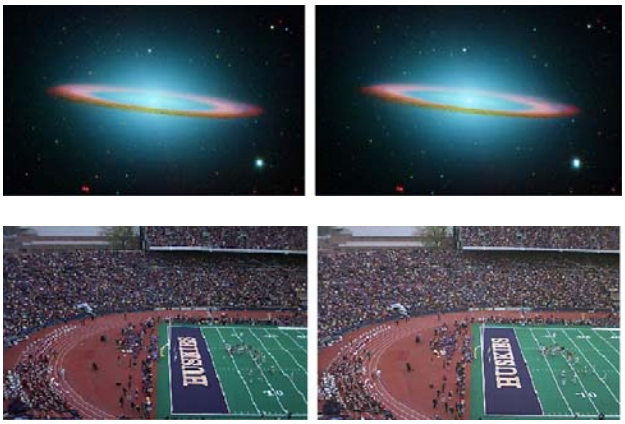


Figure 4 Result of Enhancement

V CONCLUSION

In the paper, we introduce an enhancement method based on HIS space. It is a sort of method of gray transform and spatial filtering. Firstly, this method imports Gaussian for gray transformation during the Intensity process. It has made an effective improvement in the color enhancement with three components of HIS space on the foundation the original method. while, it can avoid the disadvantages under RGB space.

REFERENCES

[1] Tsagaris V Anastassopoulos V. Multispectral image fusion method using perceptual attributes[A].Image and Signal Processing for Remote Sensing IX Proceedings of SPIE [C]. 2004,5328: pp. 357-367.

[2] Yang Yongyong and Lin Xiaozhu. Research on the Comparison of Color Image Enhancement Techniques. Journal of Beijing Institute of Petro-Chemical Technology [J]. vol14,No.3,2006, pp. 43-47.

[3] Hu qiong. Color Image Enhancement Based on Histogram Segmentation. Journal of Image and Graphics.[J] vol14.sep, 2009. pp. 1776-1780.

[4] Han Li-na, XiongJie and Geng Guo-hua. Using HSV space real-color image enhanced by homomorphic filter in two channels. Computer Engineering and Applications[J], 45(27) ,2009, pp. 18-20.

[5] Canny, J.A Computational Approach To Edge Detection, IEEE Trans. Pattern Analysis and Machine Intelligence, 8:679-714, 1986.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 256-258

Design of Adaptive Equalizer Based on Variable Step LMS Algorithm

Wang Junfeng[1], Zhang Bo[2]
[1]School of Computer Science and Technology,Henan Polytechnic University, Jiaozuo, Henan; China
wangjunfeng@hpu.edu.cn
[2]Schoo of Mathematics and Information Science, Henan Polytechnic University, Jiaozuo, Henan; China
zhangbo@hpu.edu.cn

***Abstract*—Adaptive equalizer is important in transmission of wireless communication. The equalizer using least mean square (LMS) algorithm is adopted. Simulation results show that step size influences the algorithm convergence and stability, which will significantly affect the performance of adaptive equalizer. The requirement of step for convergence speed, time-varying tracking accuracy and convergence precision is contradictory. Therefore, a variable step LMS algorithm is presented in this paper. Simulation results show that the convergence speed and stability of the variable step algorithm are superior to ordinary LMS algorithm; moreover, the variable step algorithm is proper to be applied in channel equalization in low SNR.**

Index Terms—**Adaptive Equalizer, LMS, Variable step**

I. INTRODUCTION

Multi-path effect exists in transmission of wireless data communication. The signal through different paths of different delay, is received in the same time, causing intersymbol interference. Moving communications carrier and surrounding objects (vehicles, etc.), result in the dissemination of environmental changes with time, that is, ISI caused by multi-path effects also changes with time. Adaptive equalization is a technology used to resolve inter-symbol interference.

Adaptive equalizer is used to make attends to time-varying unknown channel, so it needs a special algorithm to update the equalizer coefficients to track the channel changes. A detailed study of adaptive algorithm is a complex work.

As is described in [1]:

The disadvantage of zero forcing algorithms is that great noise gain may be appeared in the deep channel n frequency. As zero forcing equalizer completely ignores the effect of noise, it is not commonly used in wireless link.

The equalizer using least mean square (LMS) algorithm is more stable than the zero forcing equalizer. The criterion is that the mean square error (MSE) between the desired output value and the actual output value of the equalizer minimizes.

II. TRADITIONAL LMS ADAPTIVE EQUALIZATION ALGORITHM

The linear equalizer [2] is showed in Fig.1.

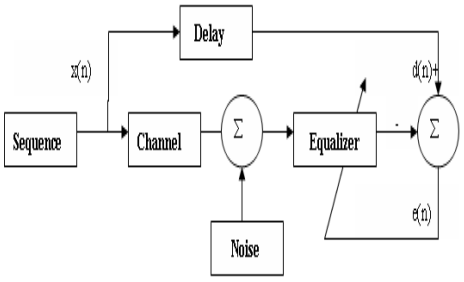


Figure.1 Linear adaptive equalization system

It can be seen, the signal x (n) after the AGWN channel is used as the input signal through the filter with different w (n), e(n) is the error between the output of the filter response y(n) and the expectation signal d(n) with delay of signal x(n) . The filter adjust the values of w(n), according to feedback error e(n) and adaptive algorithm. With the sample x (n) of the continuously is updated, e(n) becomes smaller and smaller, and w (n) gradually is close to the nominal value, similar to the ideal channel characteristics. The filter plays the role of an inverse filter in the whole process actually .

The capability of the equalizer is assessed by convergence speed and convergence stability.

The structure of the adaptive filter is showed in Fig.2.

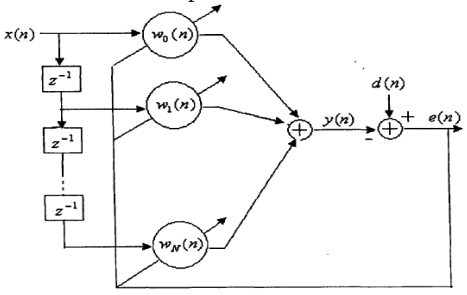


Figure 2 schematic diagram of adaptive filter

The iterative formulas of steepest descent method based on least mean square algorithm (LMS algorithm) are defined as follows:

$$y(n) = w(n)^T x(n) = x^T(n)w(n) \quad (1)$$

Where, x (n) is the filter input; y (n)is filter output and w (n) is filter weights

$$e(n) = d(n) - y(n) \quad (2)$$

Where，d(n) is a reference signal and e(n) is the error between d(n) an y(n).

$$w(n+1) = w(n) + 2\mu e(n)x(n) \quad (3)$$

Where，μ is the step size.

Convergence condition of the LMS algorithm is limited to 0 <μ <1/λmax. λmax is the largest eigenvalue of the autocorrelation matrix for input signal. The performance of the algorithm is influenced by the value of μ[3].

In LMS adaptive algorithm, the requirement of step factor for convergence rate, time-varying tracking accuracy and convergence precision is contradictory. In the range of convergence, convergence speed is faster with greater μ. But the μ value is too large, oscillation will occur during the convergence; smaller μ value can reduce steady-state the noise, improve the accuracy of convergence. However, the decrease of μ value will reduce the convergence speed and tracking speed.

III. NEW VARIABLE STEP SIZE LMS ALGORITHM

1. Principle of variable step size LMS algorithm

The contradiction between the convergence speed and the convergence precision fixed step LMS algorithm can be solved in the variable step LMS algorithm. In the initial stages of adaptive and tracking phase, a larger step size is used in order to have fast convergence speed, when the algorithm is in the steady state, smaller step is used for a small steady-state error.

Some approximation is used as a measure to control step size in adaptive processes. Simple and effective method is to use the adaptive error signal in the process, trying to establish some kind of function between the step size and the error signal

Currently, the main variable step size algorithm is to establish the nonlinear relationship between the step size and the error signal to adjust the step. The working principle is: the error is large in the initial iteration stage along with a larger step size to speed up the convergence rate; when the error is close to zero, a smaller step is accessed to achieve smaller stable-state error.

2. New variable step size LMS algorithm

In the process of convergence, e (n) decreases and approaches zero value gradually; μ value changes similar to e (n); and when e (n) = 0, μ = 0. Therefore, monotone and smooth cure of mathematical function between e(n) and μ can be concluded. The curve is through origin with μ changing by adjust e(n). It is studied that arc-Tangent curve is consistent with the variation of step factor. Therefore, variable step size LMS algorithm based on arc-tangent function is presented in this paper, called atan-LMS algorithm

The relationship between u and e(n) is established as follows:

$$\mu(n) = \text{atan}(e(n)) \quad (4)$$

For better variation, factors of α, β and γ are introduced[4]:

$$\mu(n) = \beta \text{atan}(\alpha e(n)^{\gamma}) \quad (5)$$

Where, the shape of the curve is controlled by α which influences the increase speed of step; the range of function is controlled by β; the speed of decline curve is determined by γ.

3. Parameters Analysis

(1) α for different values, β = 0.024, γ = 2

As can be seen from Fig.3 that the step increases with the increase of α in the same error case, which can speed up the convergence speed of the adaptive algorithm. But, when the error is small, the step changes largely cause the poor stability of the algorithm. It Can be seen from the figure, the step when α = 4 is close to the corresponding step when α = 8 , but better stability is achieved. So α = 4 is adopted.

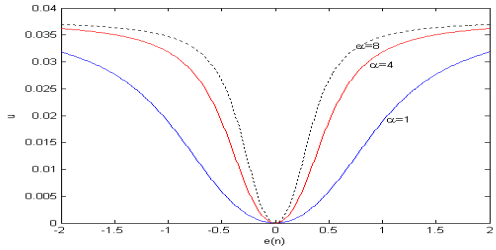


Figure.3 step with different α

(2) β for different values, α = 4, γ = 2

As can be seen from Fig.4 that the step increases with the increase of β in the same error, it can accelerate the convergence speed of the adaptive algorithm. The initial step size and range of the step is determined by β.

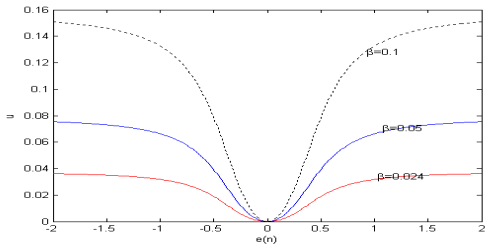


Figure.4 step with different β

(3) γ for different values, α = 4, β = 0.024

As can be seen from Figure 5, step changes with less difference in the initial stages, but when the error is small, the step changes gently with the increase of γ. The larger is γ value, the stronger is the complexity of the algorithm. Usually γ = 1or 2 is adopted.

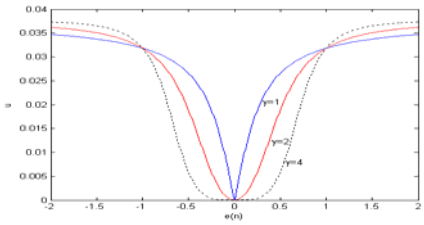


Figure.5 step with different γ

4. Simulation

x[n] is a random bipolar sequence, which value is randomly +1 and -1. Random signal is transmitted by channel which property may be a FIR filter with three-

coefficient [0.3,0.9,0.3]. Channel output is added white Gaussian noise with the variance of 1. the response of the FIR adaptive equalizer with 11 order is x [n-7][5] .

An evaluation of the merits of adaptive equalization algorithms, is mainly focused on the convergence speed and tracking capability for time-varying channel.

Experiments are performed by choosing the reasonable parameters. The u value of the LMS algorithm is 0.06.The parameters of the atan-LMS algorithm are defined as: β=0.04, α = 4, γ = 2. The length of the training sequence is 1000.The curve value is confirmed by the mean value of the error in 20 independent experiments

(1) Convergence speed

Simulation results in Fig.6 prove both algorithms have the almost same convergence speed, but the atan-LMS algorithm is more stable than the LMS algorithm.

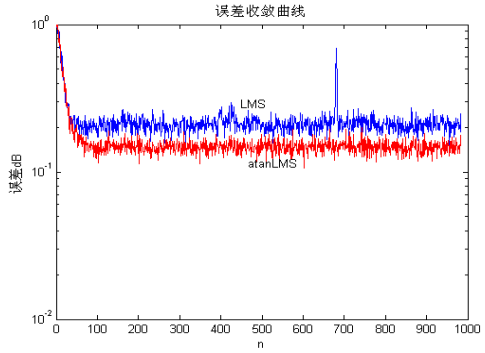
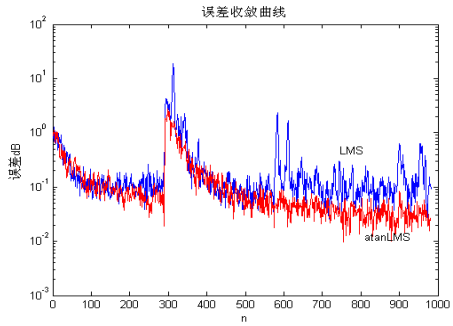
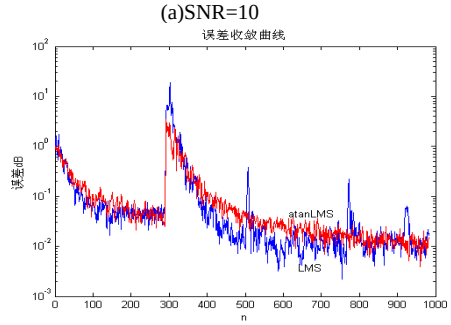


Figure.6 Convergence performance of LMS and atan-LMS

(2) Tracking capability



(a)SNR=10



(b) SNR=15

Figuew7 Tracking capability of atan-LMS with different SNR

Channel parameters change from [0.3, 0.9, 0.3] to [0.8, 0.6, 0.4]] in the first 300 iterations number. Tracking capability for time-varying channels is studied in different signal to noise ratio.

Fig.7 shows that when the SNR is low, the tracking capability of the atan-LMS algorithm is superior to the LMS algorithm.

IV. Conclusion

In this paper, the algorithm of the adaptive equalizer is studied. Through analyzing the principle of the LMS algorithm, a variable step algorithm is presented in which step factor is amended by arc-tangent function. Simulation results show that the variable step algorithm is superior to the ordinary LMS algorithm. The variable step algorithm is suit for channel equalization in mobile communication technology in low SNR.

V. Reference

[1] WANG Tian-lei, Developments in Adaptive Equalization Algorithm Research, Journal of Wuyi University [J],2009, 02(23),pp37-42

[2] DIAO Shu-lin; ZHONG Jian-bo, Analysis and Application of a Time Domain Adaptive Equalizer, Radio Engineering of China[J], 2009,09(39),pp44-47

[3] Shen Fu-min. Adaptive signal processing [M]. Xi'an: Xidian University Press, 2001

[4] ZHONG Hui-xiang, ZHENG Sha-sha, FENG Yue-ping, A Variable Step Size LMS Algorithm in Smart Antennas Based on Hyperbolic Tangent Function, JOURNAL OF JILIN UNIVERSITY (SCIENCE EDITION)[J].2008.5(46),pp935-939

[5] Wang Junfeng, A Variable Forgetting Factor RLS Adaptive Filtering Algorithm; International Symposium on Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications, 2009

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 259-262

The E-Mail Categorization and Filtering Technology Based On eEP

Yan Li[1], Xiguang Dong[2]
[1]Department of Information Engineering, Henan Technical College of Construction , ZhengZhou City,China
Email:liyanunique@126.com
[2] Department of Mathematical and Physical Science, Henan Institute of Engineering , ZhengZhou City,China
Email:nydxg@163.com

***Abstract*—The volume of junk emails on the Internet has grown tremendously in the past few years and is causing serious problems. Content-based filtering is one of mainstream technologies used so far. This paper has had a deep study in the content of emails and come up with a better idea to get the features which make it even convenient to e-mail classify as well. This paper uses the classification algorithm by essential emerging patterns CeEP to the junk email examine, and carries out a new categorization and filtering algorithm ECFEP of emails based on the EP. The experiments show, the new feature extraction methods and the combination CeEP classification is a very efficient method of classification, and The classification efficiency of the algorithm ECFEP is higher than currently several better classification algorithm.**

***Index Terms*—E-Mail Categorization, Feature Extraction, Essential Emerging Patterns**

I. THE BASIC CONCEPT

DB-based training data set contains N samples e-mail (T1, T2, ..., TN), is divided into two known types of C1, C2, and a given sample of each class mail. Classification in the mail, all the samples are all text. Although the text of the title, abstract and key words containing important classified information, but not all of the text contains such information. Therefore, assuming that all the text of this article contains only the contents of the message body, hereinafter referred to as message content.

We use Spam Corpus —PU Series corpus set. Its received from a provider of real-time e-mail. Corpus to retain only those e-mail the title and the body of the plain text content.

Many classification algorithms[1] have message content will be mapped to n-dimensional space, in which n equal to the contents of all e-mail appear in the number of different words. After the initial filtering, mail content in the different number of words remains high. Feature extraction task is to delete that information for the classification of small, "unimportant" words. After feature extraction, the message appears in the text known as the characteristics of the word. In the following discussion, sometimes referred to as the characteristics of items, the term "word", "Feature" and "item" will be mixed use.

Feature extraction, each message is a collection of items. So that W = (w_1, w_2,, ..., w_n) is the message content of the items appeared in The Complete Works. W subset $X \subseteq W$called itemsets. If the itemset X in T appear in the text, then T contains X.

Definition 1

a set of training data set D is a subset of DB. Itemset X in D on the degree of support $\sup_D$ (X) = $count_D(X)/|D|$, which $count_D$ (X) is a data set D contains a sample of the number of X, and | D | is the total number of D in the samples.

If D is a collection of C_i types of training samples, $\sup_D$ (X) recorded as $\sup_i$ (X), it is the itemset X in the category C_i training focused on the frequency samples.

Definition 2

Given two dierent classes of datasets D and D' ,the growthrate of an itemset X from D to D' is defined as GrowthRate(X)= $gr_{D'\to D}(X)$:

$$gr_{D'\to D}(X)=\begin{cases} 0 & \text{if } sup_{D'}(X)=sup_D(X)=0 \\ \infty & \text{if } sup_{D'}(X)=0, sup_D(X)\neq 0 \\ sup_D(X)/sup_{D'}(X) & otherwise \end{cases}$$

If the data sets D and D 'are non-spam and junk e-mail collection of samples, $gr_{D'\to D}(X)$ recorded as $gr_i(X)$, it is a set X from a non-junk mail to junk e-mail support (frequency) significant changes in the extent of the measure.

Denition 3

Given a growth rate threshold ρ >1, an itemset X is said to be ρ-Emerging Pattern[2](ρ-EP or simply EP)from a background dataset D to a target dataset D' . Itemset X is eEP (essential EP) of D , if (1) X is D, EP, (2) X in D, the support is not less than pre-specified minimum support threshold ξ, and (3) X subset of any really not satisfy the conditions (1) and (2).

When D and D 'are non-spam and junk e-mail when a collection of samples, D of the EP / eEP also known as spam EP / eEP. In fact, eEP is the "shortest possible", the most ability to express the EP. During the discussion after the feature extraction, we will discuss in detail the eEP based on the establishment of e-mail classifier.

II. THE PRE-PROCESSING AND FEATURE EXTRACTION OF THE E-MAIL TEXT

E-mail in the feature extraction, we first set of data pre-processing e-mail:

(1) an e-mail data sets together an e-mail messages to a large document, remove the message headers, message content, only in part. These e-mail documents to each message type label started, and then began to use -1 as the content of signs, followed by the message body, and finally to -11 as the end of each message;
(2) each repetition of the word removed from the e-mail message body to retain only a duplication of the word.

feature extraction for the message:
(1) Statistics for each word in a normal e-mail and spam that exist in the frequency into the hash table;
(2) for the hash table appear in the zero-spam and email in the normal word zero times we have it mapped to a fixed two special symbols (special symbols can be used in place of any number), and then remove the e-mail the body of each repeat of the special symbols, special symbols so that each retained only one in the message body in order to shorten the length of the message body;
(3) According to a word in the greater difference between two types of messages in terms of its more important, more important the higher frequency of these two principles, for the hash table in a normal e-mail spam and there are times for the non-zero term, we Among several types of frequency for the larger number of frequency x, the smaller number of frequency y, proposed formula:

$$F(m)=\left(\frac{x}{x+y}\right)^{\alpha}(x-y)^{\beta}(\alpha>0,\beta>0)$$

α, β balance these two principles, One of value set δ , δ known as the balance factor, the balance of the different factor δ, by descending order of the results on the order of different threshold δ extract large in the *c* word as our word feature extraction (0 <*c* <1), the number of feature words set to 70, each time the characteristics of the word is greater than the number of messages in order to select a different value of *c*;
(4) feature extraction of speech after the message body and words in the match to retain the characteristics of the word, by deleting the non-feature of the word, match result, the body of each message contains only the characteristics of the word (that is, Feature items).

III. EEP-BASED E-MAIL CLASSIFICATION AND FILTERING ALGORITHMS ECFEP

Feature extraction, all of the messages are a collection of characteristics, and training data sets characteristic of DS is a collection of multiple sets.

A. Mining eEP

For the establishment of the e-mail-based classifier eEP, first of all need to dig eEP. The steps are as follows:

(1) get set minimum support threshold ξ and minimum growth rate ρ;

(2) for *i* = 1,2, Ci on behalf of two types of e-mail (spam and non-spam):
Training data set will be divided into categories Ci and Ci samples set;
Mining Mining eEP category Ci and the Ci-type eEP;

literature[6] gives the detailed steps eEP excavation, this article is no longer cumbersome. A large number of experiments showed that growth ξ = 1%. However, the minimum support threshold rates ρ depend on the data distribution. In general, for the easy classification smaller can take larger values (greater than 5), contrary of data sets, better value can be set up through repeated value should be taken (2~5). classifier, according to the classification of the samples tested from the accuracy to determine appropriate adjustments. See literature [5].

Some characteristics may not appear in any eEP, they do not work the classification of unknown samples.

Definition 4 Characteristics of the definition of *w* is a key feature of an effective, if *w* appears in at least one of eEP.

B. Sorting

eEP distinguish between a good performance. *X* is Ci-based category eEP, its growth rate of gri (*X*). This indicates that the focus in the classification of samples, *X*-type Ci samples in the frequency (support) is a non-Ci samples in the frequency of the gri (*X*) times. If the *X* in question appeared in *T* classification of mail, from a statistical point of view, *T* is the possibility of Ci-type *T* does not belong to Ci category gri (*X*) times.

E-mail to be classified in order to determine the type of *T*-owned, Ci each category eEP are trying to determine whether the *T*-type Ci. *X* is Ci-based category eEP. If *X* is not appear in *T*, then *X* can not determine whether the *T*-type Ci to judge. If *X* appears in *T*, *X* will be the probability $\frac{gr_i(X)}{gr_i(X)+1}$ Determine the type *T* belong to Ci, and to the probability $\frac{1}{gr_i(X)+1}$ Determine the type *T* does not belong to Ci.

In order to classify e-mail *T*, ECFEP combination of Ci to C_i-type and non-eEP of each category to determine the calculation of *T* scores are Ci-type, score (*T*, Ci). ultimately determine the type *T* belong. The PS (*T*, Ci) = (*X* | *X* is Ci category eEP, and *X* in *T* appear in), NS (*T*, Ci) = (*X* | *X* non-Ci category eEP, and *X* appear in *T*) . ECFEP the following steps to classify e-mail *T*:

a) the deletion of *T* in the absence of an effective focus on the characteristics of the word appears;

b) For i = 1, 2,

For PS (*T*, C_i) and NS (*T*, C_i);

By computing *T* under the category C_i is the score score (*T*, C_i)

$$score(T,C_i)=\sum_{X\in PS(T,C_i)}\frac{gr_i(X)}{gr_i(X)+1}+\sum_{X\in NS(T,C_i)}\frac{1}{gr_i(X)+1}$$

c) *T* was placed under the category of the highest scores.

IV. ANALYSIS OF EXPERIMENTAL RESULTS AND EVALUATION

Experimental data sets using public spam corpus set PU series, provided by the Greek scholar Androutsopoulos. Its received from a provider of real-time e-mail. Corpus to retain only those e-mail the title and the body of the plain text content. Providers in order to protect the privacy of e-mail corpus, It will be different in different words in place of integers. PU Series corpus currently consists of PU1, PU2, PU3 and PUA four corpus. The average corpus of each PU is divided into 10, that is, part1 to part10. At present, we mainly corpus PU1, PU1 corpus of each check in a ten to 10 fold cross-validation (cross validation), PU Series corpus shown in table 1.

Table 1 spam database(unit: letter)

Data set	Non-spam numbers	Spam numbers	Totall numbers	Remarks
PU1	618Pr	481Pr	1099	Encryption Forms
PU2	579Pr	142Pr	721	Encryption
PU3	2313Pr	1826Pr	4139	Encryption
PUA	571Pr	571Pr	1142	Encryption

Spam is usually classified using the performance evaluation of text classification relevant indicators. Specifically, based pos is the total number of spam, t_pos is the correct classification of spam (junk e-mail really) a few, and f_pos was wrongly classified as spam (false spam) number, then the following evaluation different indicators can be used to measure the spam filtering performance of the system:

(1) Recall : Recall that the rate of spam;

$$recall = \frac{t\_pos}{pos}$$

Recall reflects the filtration system's ability to find spam. The higher recall rate, "slipping through the net" less spam.

(2) Precision: that is, spam precision;

$$precision = \frac{t\_pos}{t\_pos + f\_pos}$$

The accuracy of the filter response system "to find the" junk e-mail capabilities, precision higher miscarriage of justice would be a legitimate message as spam the possibility of the smaller.

(3) Accuracy: that is, for all mail (including junk mail and legitimate e-mail) on the rate of the contractor. Accuracy, that is, for all mail (including junk mail andError rate: err = 1 legitimate e-mail) the rate of the sentence wrong.

(4) F-measure:

$$F = 2\frac{recall \times precision}{recall + precision}$$

F-measure is the recall rate and accuracy of harmonic average, it will recall rate and precision into a comprehensive indicator.

In addition, spam filtering is often used in the Fallout, Miss rate and so on.

In order to verify our proposed method of feature extraction and classification of CeEP after combining the classification system in the classification of the efficiency of e-mail, the paper has done three sets of experiments: (1) parameters for different values ECFEP filtering algorithm for classification and evaluation results ; (2) with parameters fixed to the assessment of changes in the growth rate of trend indicators; (3) Comparison with other algorithms.

Experimental environment for Pentium 4 CPU, 256MB RAM, 80GB hard drive, the operating system to Microsoft Windows XP, programming software for the Microsoft Visual C++. Net 5.0. Experiment using 10 fold cross-validation approach to the statistical results of the classification of mail. That data sets will be divided into ten mutually exclusive subsets of pay or "discount" DB_1, DB_2, ..., DB_{10}, roughly equal the size of each pack. Training and testing carried out 10 times. i times in the first, DB_i used as a test set, a subset of the remaining are used for training classifiers.

A. for different values of parameters ECFEP the results of the evaluation algorithm

Experiments found that Balancing factor, α and β values is very important because the balancing factor α and β values affect the order of characteristics. too small or too big can not extract the characteristics of a good item. *c* value of the selected test results for the equally important, the *c* values for different characteristics of the selected items have a great impact. After a large number of experiments show that, α and β, better select for

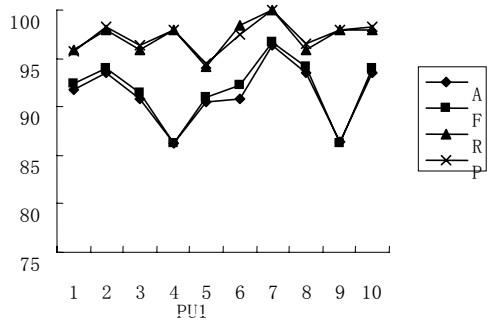


Fig.1 α，β were 2，1/2 PU1 Experimental evaluation results

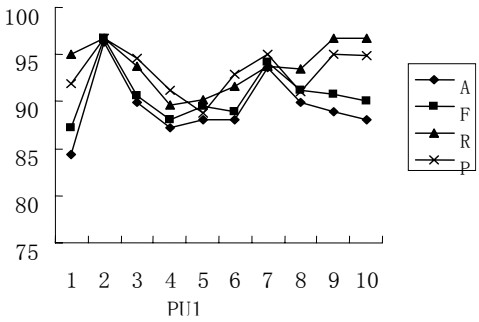


Fig.2 α，β were 3，1/3 PU1 Experimental evaluation results

α and β, countdown each other, and the value of 2 and 3. c the value of a good choice for between 0.30 to 0.60. And the growth rate r of the classification threshold evaluation of the results of the impact is not great. Therefore, the main consideration in the experiment balance factor α and β value of the classification results. Figure 1 Figure 2, respectively, better values are given access (α = 2, β= 1 / 2) and (α= 3, β = 1 / 3), when our results.

Compare Figure 1, Figure 2 we can see that the balance factor, respectively 2,1/2 of α and β at the time of the recall rate and accuracy are respectively higher than the 3,1/3. however, accuracy and F-measure of 3,1/3 are respectively higher than the 2,1 / 2. Also found in PU1 corpus set part2 and part7 four evaluation criteria are the highest. respectively 2,1/2 α and β, the recall rate and precision of part7 were 100%. With the higher recall rate and the omission of the less spam, the higher the accuracy of miscarriage of justice would be a legitimate message as spam the possibility of the smaller. The two evaluation criteria to improve our spam classification and filtering efficiency is very important, so ECFEP algorithm to achieve a better classification's ability to spam.

B. Comparison with other algorithms

PU1 corpus in the same set, we use the current favorable Naive Bayesian (Nbayes) classification algorithm (reference[9] The evaluation results) as well as the more popular KNN algorithm (reference[10] The evaluation results), Decision Tree Algorithm (Decision Tree) (reference[8] The evaluation results) and Bayesian neural networks (Bayes network) algorithm (reference [11] The evaluation results) the results of ECFEP algorithm and compare the results of Table 2 below (Table 2 the results are listed in spam and non-spam

Table 2 Five algorithms classification and filtering evaluation results

	A	*F*	*R*	*P*
Nbayes	0.9318	0.9415	0.9191	0.9578
KNN	0.977	0.9324	0.935	0.9298
Decision Tree	0.891	0.8805	0.882	0.879
Bayes Network	0.892	0.9362	0.9466	0.926
ECFEP	0.9532	0.9526	0.9538	0.9517

classification of the results of the weighted average):

A representative of the accuracy, F on behalf of F-measure, R on behalf of the recall rate, P on behalf of precision.

From Table 2 we can see that the recall classification ECFEP and F-measure are the highest. The precision of ECFEP is the same as the highest precision of the Naive Bayesian algorithm. KNN algorithm is to the highest rate of accuracy. Decision tree of the evaluation index have reached the minimum. For spam classification algorithm the recall rate and the improvement of accuracy is particularly important, we can see ECFEP algorithm has a very high classification and spam filtering capabilities.

V. PROSPECTS

In this paper, e-mail classification and filtering methods are discussed. A new feature extraction method are proposed. Combined with the basic exposure model based on (eEPs) classification algorithm CeEP, we realized the EP-based classification and filtering e-mail filtering algorithm ECFEP. Experiments show, ECFEP is a very efficient method of e-mail classification and filtering. The next step will be to expand spam ECFEP used for other data sets. Spam ECFEP applied to other areas of data sets.

REFERENCES

[1] Pang-Ning Tan, Michael Steinbach, Vipin Kumar ,Fan ming, Fan hong jian. Introduction to Data Mining. Beijing: posts and telecom press, 2006,259~293.

[2] G. Dong, X. Zhang, L. Wong *et al*. CAEP: Classification by aggregating emerging patterns. Proc. of the 2nd Int'l Conf. on Discovery Science (DS'99). Berlin: Springer-Verlag, 1999. 30~42.

[3] I. H. Witten and E. Frank (2005) "Data Mining: Practical machine learning tools and techniques", 2nd Edition, Morgan Kaufmann, San Francisco, 2005.

[4] Fan H and Ramamohanarao K. Bayesian Approach to use Emerging Patterns for Classification. In Proc of 14th Australasian Database Conference. Adelaide, Australia: Australian Computer Society, 2003. 39~48.

[5] Fan Ming, Liu Mengxu, Zhao Hongling. Classification by Essential Emerging Patterns. Computer Science (in Chinese), 2004, 20 (11): 211~214.

[6] Fan Ming, Wei Fang. Mining Essential Emerging Patterns for Classification Computer Science (Supplement) (in Chinese), 2004, 307~309.

[7] Xu Hong Tao, Fan Ming , Zan Hong Ying. Text automatic Categorization by Emerging Patterns. Journal of computer research and development(Supplement) 2005. 9, 351~355.

[8] Wang Bin,Pan Weng feng. A Survey of Content-based Anti-spam Email Filtering. Journal of Chinese information processing , 2005, 19(5).

[9] Zou lei, Hu Yan Sheng, Cui De Xuan *et al* . An anti-spam filtering algorithm based on cost minimization. Huangzhong university science and technology(Nature Science Edition) ,2005.12, 33, 352~355.

[10] Lin Chen, Li Bi Cheng. New effective method for spam filtering. Computer Application, 2006,26(8), 1980~ 1982.

[11] Liu Zhen, Zhou Ming Tian. Spam filtering algorithm based on supervised Bayesian parameter estimation. Computer Application, 2006, 26(3), 558~561.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 263-266

Convergence of Internet Congestion Control

Lina Zhang [1], Ya Li [2]
1 School of Computer Science and Technology
Henan Polytechnic University, Jiaozuo, P.R. China
Email: zln@hpu.edu.cn
[2] School of Computer Science and Technology
Henan Polytechnic University, Jiaozuo, P.R. China
Email: liya@hpu.edu.cn

***Abstract*—Internet congestion control is inherently high-dimensional, nonlinear, dynamic, and complex. In this paper, We study a simplified model with one resource. We discuss rate of converge for the one resource model, completely characterizing convergence to equilibrium for the region of stability.**

***Index Terms*—congestion control; convergence; internet**

I. INTRODUCTION

With the rapid development of the technique of communication networks, especially the internet, and the increase of the requirement for networks, it becomes more and more important to provide congestion control and avoidance algorithms, and analyze the dynamics of these algorithms. Dramatic process is being made in developing such a theoretical framework to investigate and solve the problem of internet congestion control. Congestion control is a representative example of how ad hoc solutions, although being successful at the system reveals their deficiencies as the evolution of the system reveals their deficiencies. Congestion control mechanisms were introduced in the transmission control protocol (TCP) protocol to fix the defects that led in October 1986 to the first of a series of "congestion collapses." Despite its profound success, there are currently strong indications that TCP will perform poorly in the future high-speed networks. Simulations and real measurements indicate that as the bandwidth delay products increase within the network, the slow additive increase and the drastic multiplicative decrease policy of the TCP protocol cause the system to spend a significant amount of time typing to probe for the available bandwidth, thus leading to underutilization of the available resource. It has also been shown analytically that as the bandwidth delay products increase, TCP becomes oscillatory and prone to instability. Moreover, TCP is grossly unfair towards connections with high round-trip delay. Finally, it has been shown that in networks incorporating wireless and satellite links, long delays and non congestion-related losses also cause the TCP protocol underutilize the network. The theoretical framework used in most of the recent studies. In the fore-mentioned framework, the congestion control problem is viewed as a resource allocation problem where the objective is to allocate the available resource to the competing users without the input data rates at links exceeding the link capacity. Through an appropriate representation, this problem is transformed into a convex programming problem. A utility function is associated to each flow and the objective is maximize the aggregate utility function subject to the capacity constraints. congestion control algorithms can then be viewed as distributed iterative algorithms that compute optimal or suboptimal solutions of this problem.

Control problem in communication networks are inherently high-dimensional, nonlinear, dynamic, and complex. However, they have become increasingly important today due to the explosive expansion and growth of traffic in the internet. Congestion control in the internet is an extremely important and challenging problem, which has been the main subject of intensive studies over the last decade [1-11]. Congestion occurs in the internet when the users of the network collectively demand more resources like bandwidth and buffer space than the network has to offer. Unless appropriate action is taken to control network congestion, the network can end up in a state of persistent overload, potentially leading to congestion collapse [1].

The congestion control algorithms for the internet generally can be classified into two types: one is the algorithms implemented in its transmission control protocol(TCP), such as TCP Reno and TCP Vegas, which adjust transmission rates of sources based on available feedback congestion marks[1]. The other is the algorithms, such as Drop Tail and RED, used as the active queue management (AQM) at the link nodes, which how to drop arriving packets when the network is overload [12]. However, it is believed that the existing congestion control algorithms which are based on "trial-and-error" methods employed on small testbeds may be ill-suited for future network where both communication delay and network capacity can be large [13]. This has motivated research on theoretical understanding of TCP congestion control and the search for protocols that scale properly so as to maintain stability in the presence of these variations. In particular, an optimization-based framework that provides an interpretation of various congestion control mechanisms is developed [2, 3]. The advance in mathematical modeling of congestion control have stimulated the research on the analysis of the behavior, such as stability, rate of convergence, robustness and fairness, of currently developed Internet congestion control protocols as well as the design of new protocols with higher performance[14,15].

In this paper we consider a simper network consisting one resource and one route. In the paper we discuss rate

of convergence for the network model, completely characterizing convergence to equilibrium for the region of stability. The paper is organized as follows. In section I, we give an introduction for the internet congestion control. Next, we give an end-to-end internet congestion algorithm in section II. In section III, we discuss the rate convergence for the one resource model. We give conclusions about the internet congestion control in section IV.

II. AN END-TO-END CONGESTION CONTROL ALGORITHM

We consider a network with a set, J, of resources. Let a route r be a nonempty subset of J, and denote the set of all routes by R. Associate a route r with a user and let $x_r(t)$ be the sending rate of user r, which models the number of packets generated by user r at time t. In an end-to-end internet congestion control algorithm, each user tires to adjust its sending rate based on available feedback information so that the overall network satisfies some desirable properties. The primal algorithm is given by

$$\frac{d}{dt}x_r(t) = \kappa_r\left(w_r - x_r(t)\sum_{j\in r}\mu_j(t)\right), r \in R \qquad (1)$$

where κ_r is a positive constant and the congestion indication signal $\mu_j(t)$ at resource j is generated by

$$\mu_j(t) = p_j(\sum_{s:j\in s} x_s(t)), j \in J \qquad (2)$$

in which the congestion indication function $p_j(\cdot)$ is increasing, nonnegative, and not identically zero.

Algorithms (1) and (2) can be interpreted as follows. Suppose that user r generates packets $x_r(t)$ at time t. s is a route that through resource j and $y = \sum_{s:j\in s} x_s(t)$ represents the total follow through resource j. $p_j(y)$ can be view as the probability a packet at resource j receives a "mark"—a feedback congestion indication signal and $\mu_j(t)$ is the probability a packet at resource j receives a mark at time t. If we assume a packet may only be marked at most once, then $\sum_{j\in r}\mu_j(t)$ is the probability a packet from user r received a mark at time t and $x_r(t)\sum_{j\in r}\mu_j(t)$ is just the expected number of marks received by user r at time t. Eq.(1) corresponds to a rate control algorithm for user r that tries to adjust its sending rate $x_r(t)$ so that the expected number of marks received by user r will tend to a target value w_r.

Systems (1) and (2) has a unique equilibrium point, $x^* = [x_1^*, ..., x_R^*]^T$, given by

$$x_r^* = \frac{w_r}{\sum_{j\in r} p_j(\sum_{s:j\in s} x_s^*)} \qquad (3)$$

and this equilibrium point is globally asymptotically stable.

In order to investigate the influence of propagation delays on the stability of congestion control algorithms, we consider adding delays to (1) and (2) as follow. Given a router r, for each resource $j \in r$ we define a forward delay $d_{jr}^{\rightarrow}$, and a return delay $d_{jr}^{\leftarrow}$. The forward delay is the delay incurred in communication from the user to the resource; the return delay is the delay incurred in communication from the resource back to the user. In the current internet, each route is subject to a roundtrip delay. We model this delay by assuming each route has an associated delay D_r, such as $d_{jr}^{\rightarrow} + d_{jr}^{\leftarrow} = D_r$ for each $j \in r$. Consider now the following delayed difference equations analogous to the primal algorithm, where we assume that $d_{jr}^{\rightarrow}$ and $d_{jr}^{\leftarrow}$ are integer valued:

$$x_r[t+1] = x_r[t] + \kappa_r\left(w_r - x_r[t-D_r]\sum_{j\in r}\mu_j[t-d_{jr}^{\leftarrow}]\right) \qquad (4)$$

for $r \in R$, where, for $j \in J$,

$$\mu_j[t] = p_j\left(\sum_{s:j\in s} x_s[t-d_{js}^{\rightarrow}]\right). \qquad (5)$$

We consider a simple network consisting of one resource and one route. The difference equation describing this system is:

$$x[t+1] = x[t] + \kappa(w - x[t-D]p(x[t-D])) \qquad (6)$$

where $D = d^{\rightarrow} + d^{\leftarrow}$.

III. RATE OF CONVERGENCE

For the simplified case where we have only one resource, we can study rate of convergence to the stable point. We consider this problem, via the theory of differential-difference equations. For convenience, we restate a result due to Hayes [16].

Lemma 1(Hayes) All the roots of $be^{\lambda} + c - \lambda e^{\lambda} = 0$, where b and c are real, have negative real parts if and only if: (1)b<1;and (2), $b < -c < \sqrt{a_1^2 + b^2}$, where a_1 is the root of $a = b\tan a$ such that $0 < a < \pi$. If b=0, we take $a_1 = \pi/2$.

We study the rate of convergence of the system through the following differential-difference equation:

$$\frac{d}{dt}x(t) = \kappa[\omega - x(t-D)p(x(t-D))] \qquad (7)$$

where $D = d^{\rightarrow} + d^{\leftarrow}$. This is just the continuous analog of the discrete-time equation (6). We can study the stability of this simple system via linearized version. Let the fixed point be (x, p) where $p = p(x) = 1 - e^x(1-x)$ and $\omega = xp$. Then linearizing with $x[t] = x + y[t]$, we obtain:

$$\frac{d}{dt} y(t) = -\kappa(1 - e^x + xe^x + x^2 e^x) y[t-D],$$

neglecting higher order terms. The corresponding characteristic equation, obtained by substituting $y = e^{st}$, is thus:

$$s = -\kappa(1 - e^x + xe^x + x^2 e^x) e^{-sD},$$

which after substituting $\lambda = sD$, reduce to:

$$-\kappa(1 - e^x + xe^x + x^2 e^x) D - \lambda e^{\lambda} = 0 \qquad (8)$$

The fixed point is locally stable if all roots of the above equation have negative real part. For each D, we are interested in the maximum value of κ such that the system is locally stable.

Theorem 1: The system (7) is locally stable if:

$$\kappa(1 - e^x + xe^x + x^2 e^x) < \frac{\pi}{2D}$$

And unstable if:

$$\kappa(1 - e^x + xe^x + x^2 e^x) > \frac{\pi}{2D}$$

Proof: A direct application of Lemma 1 to equation (8) with b=0 and $c = -\kappa D(1 - e^x + xe^x + x^2 e^x)$.

Define $a = \kappa(1 - e^x + xe^x + x^2 e^x)$; then the linearization of (7) is :

$$\frac{d}{dt} y(t) = -a y(t-D) \qquad (9)$$

When $a < \pi / 2D$, all roots λ of (8) satisfy Re λ <0 (by Lemma 1). Let λ^* be a root of (8) such that $\mathrm{Re}\,\lambda^* > \mathrm{Re}\,\lambda$ for all other roots λ. Then the rate of convergence to the stable point is equal to $|\mathrm{Re}\,\lambda^*| D^{-1}$. The following theorem characterizes the rate of convergence of (9), for a in the region of stability $(0, \pi(2D)^{-1})$.

Theorem 2: the maximum rate of convergence for the system (9) is D^{-1} when $a = (eD)^{-1}$. The rate of convergence is monotonic increasing and convex from 0 to D^{-1}, and the convergence is nonoscillatory, for $a \in (0, (eD)^{-1}]$. The rate of convergence is monotonic decreasing from D^{-1} to 0 for $a \in [(eD)^{-1}, \pi(2D)^{-1})$.

In particular, the maximum rate of convergence to the equilibrium point x of (7) is achieved if and only if:

$$\kappa(1 - e^x + xe^x + x^2 e^x) = \frac{\pi}{eD}$$

and in this case, the equilibrium is nonoscillatory.

Proof of theorem 2 follows immediately from the following lemmas.

Lemma 2: For $a \in [(eD)^{-1}, \pi(2D)^{-1})$, the rate of convergence of the system (9) decreases monotonically from D^{-1} to 0.

Proof: Let the root be $\lambda = -\gamma + \delta i$ where $\gamma > 0$. Then from equation (8):

$$\gamma e^{-\gamma} \cos\delta + \delta e^{-\gamma} \sin\delta = aD \qquad (10)$$

$$\gamma e^{-\gamma} \sin\delta - \delta e^{-\gamma} \cos\delta = 0 \qquad (11)$$

Notice that if (γ, δ) satisfies the above equations, then so does (γ, δ). Hence, without loss of generality, we may assume that $\delta \geq 0$. Equation (11) gives:

$$\gamma = \delta / \tan\delta \qquad (12)$$

which on substitution into (10) yields:

$$\delta e^{-\delta/\tan\delta} / \sin\delta = aD \qquad (13)$$

Observe that $\delta \in [2n\pi, (2n + \frac{1}{2})\pi]$ where $n \in Z_+$ since $\tan\delta \geq 0$ and $\sin\delta \geq 0$ from equation (12)-(13). Also, from (12) γ is monotonic decreasing in δ from 1 to 0 for $\delta \in [0, \pi/2]$. Then the LHS of (13) is monotonic increasing in δ from e^{-1} to $\pi/2$ for $\delta \in [0, \pi/2]$. Hence, a root of equations (12)-(13) exists for $\delta \in [0, \pi/2]$ which is decreasing in γ as a increases from $(eD)^{-1}$ to $\pi(2D)^{-1}$. It remains to show that this is the root with the smallest γ. Notice that if $\delta / \sin\delta$ is larger, then γ has to be larger for equation(13) to be satisfied. But $\delta / \sin\delta \leq \pi/2$ for $n = 0$ and $\delta / \sin\delta \geq \pi/2$ for $n > 0$, so taking $n = 0$ yields the root with the smallest γ. Then, since the rate of convergence is γD^{-1}, we have the required result.

Lemma 3: The rate of convergence of the system (9) is monotonic increasing and convex from 0 to D^{-1}, and the convergence is nonoscillatory, for $a \in (0, (eD)^{-1}]$.

Proof: Consider the function $-\lambda e^{\lambda}$ when λ is real and negative. Suppose we maximize the function with respect to λ. We have:

$$\frac{d}{d\lambda}(-\lambda e^{\lambda}) = -e^{\lambda} - \lambda e^{\lambda}$$

which equals zero when $\lambda = -1$, giving a maximum of e^{-1}. Moreover, $-\lambda e^{\lambda}$ is convex increasing in λ, rising from an asymptote of 0 to e^{-1} on the interval[-1,0]; This characterization implies that that equation (8) has 2 negative roots when $a < (eD)^{-1}$ and 1 negative root when $a = (eD)^{-1}$. Moreover, the modulus of the smaller

root (in modulus) increases from 0 to 1 as a increases from 0 to $(eD)^{-1}$, and is clearly a convex function of a.

It remains to show that taking λ be real yields the root with the smallest (in modulus) negative real part; this will also show that the convergence is nonoscillatory. First note that the modulus of the smaller real root is not greater than 1. Suppose there are complex roots for $a \le (eD)^{-1}$. Then using the same argument as in the proof of Lemma 2, we see that there are no roots for n=0, $\delta \neq 0$, since the LHS of equation (13) is greater than e^{-1}.Also, for $n > 0, \delta / \sin \delta > \pi / 2$, so the complex root, if any, will have $\gamma > 1$. This completes the proof.

IV. CONCLUSIONS

This paper has studied the simple model for a resource. Especially we achieve the rate of convergence of congestion control. Congestion control is very complex, even congestion control of a simple network is challenging. The problem of congestion control is still far from solved. We hope this paper can stimulate more research in this area.

ACKNOWLEDGMENT

The work was supported by Youth Fund of Henan Polytechnic University (Q2009-39).

REFERENCES

[1] V.Jaccobson, Congestion avoidance and control, proceedings of ACM SIGCOMM'88, Stanford, CA, 1988, pp. 314-329

[2] F.Kelly, A. Maulloo, D.Tan, Rate control in communication network: shadow prices, proportional fairness, and stability, J.Oper. Res. Soc. 49(1998) 237-252.

[3] S.H. Low, D.E. Lapsley, Optimization flow control-I: Basic algorithm and convergence. IEEE/ACM Transactions on networking, 7(1999), 861-874.

[4] F. Paganini, J. Doyle, S. Low, Scalable laws for stable network congestion control, conference on Decision and control, 2001, to appear. http://www.ee.ucla.edu/˜paganini

[5] L. Peterson, B. Davie, Computer Network: A Systems Approach, 2nd Edition, Morgan Kaufmann, Los Altos, CA, 2000

[6] R.J. Gibbens, F.P. Kelly, Resource pricing and the evolution of congestion control, Automatica 35(1999) 1969-1985.

[7] L. Benmohamed, S. Meerkov, Feedback control of congestion in packet switching networks: the case of a single congested node. IEEE/ACM Transactions on networking 1, 1993,no. 6:693-708

[8] L. Benmohamed, S. Meerkov, Feedback control of congestion in packet switching network: the case of multiple congested nodes. In: Proceeding of the American Control Conference, 1194, pp. 1104-1108

[9] F. Bonomi, D. Mitra, J.B.Seery, Adaptive algorithms for feedback based flow control with in high-speed, wide-area data networks. IEEE Journal on Selected Areas in communications 13: 1267-1283.

[10] S. Chong, R. Nagarajan, Y.T. Wang. First-order rate-based flow feedback control with dynamic queue threshold for high-speed wide-area ATM networks. Computer Network and ISDN System, 1998, 29:2201-2212

[11] K.W. Fendick, M.A. Rodrigues, A. Weiss, Analysis of a rate-based feedback control strategy for long haul data transport. Performance Evaluation, 1992,16: 67-84.

[12] S. Floyd, V. Jacobson. Random early detection gateways for congestion avoidance. IEEE/ACM Transactions on networking, 1(1993), 397-413.

[13] S. H. Low, & R.Srikant. A mathematical framework for designing a low-loss, low-delay internet. 2002. http://comm.csl.uinc.edu/ŝrikant /pub.html.

[14] G. Vinnicombe. Robust congestion control for the internet.2001. http://www-control.eng.cam.ac.uk/gv/internet/index.html.

[15] S. H. Low, F. Paganini, & J.C. Doyle. Internet control. IEEE Control System Magazine,2002, 22(1),28-43

[16] N. D. Hayes. Roots of the transcendental equation associated with a certain differential-difference equation. Journal of the London Mathematical Society. 1950. 25:226-232.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 267-269

An Improved Naive Bayes Text Classification Algorithm In Chinese Information Processing

Lingling Yuan[1]
[1]School of Media and Literature/Henan Polytechnic University, Jiaozuo, China
lianna@hpu.edu.cn

***Abstract*—In Chinese information processing, Naive Bayes is a simple text classification method that is easily implemented. Its core is the realization of the calculating posterior probability algorithm and the effectively reducing dimension for feature words. This paper improved Naive Bayes text classification from the calculating posterior probability and the reducing dimension of feature words of text. The result of experiment indicated that the improved method is of the higher efficiency than the original algorithm.**

***Index Terms*- Naive Bayes; text classification; feature word; multi-variable Bernoulli model**

I. INTRODUCTION

With the development of the Internet, web applications such as e-commerce and information search are applied widely, and the information on network is expanding rapidly. It gives us a wealth of resources, almost any information we looking for can be found on the Internet. Every day we face a lot of information from the Internet, including the useful and spam. So it is currently a major challenge in information technology how to quickly find the correct information that the user need from the mass information. As an important carrier of information, most of network information is stored in the form of text. Therefore, text classification has become a hot research topic in this field.

The target of text classification is to categorize texts into one or more appropriate sorts that are based on analysis of the text content. Text classification system that is based on artificial intelligent technology can automatically classify many texts according to the semanteme of text, which makes text information easier to use. Text classification technology will gradually combine with some information processing technologies such as search engine and information filtration, which will improve the quality of information service effectively.

At present, the mainly common text classification methods are the Bayes Classification Algorithm, K-Nearest Neighbor(KNN), Neural Network(NN), Support Vector Machine (SVM), the Decision Tree, Linear Least Squares Fit (LLSF), etc [1-5]. The Bayesian classification algorithm which is recognized as a simple and effective method of text classification is focuses on in this paper.

II. NAIVE BAYES CLASSIFIER

A. Principle

The Naive Bayes classifier is a simple probabilistic classifier based on applying Bayes' Theorem (from Bayesian statistics) with strong (naive) independence assumptions which assumes all of the features are mutually independent.

It uses a Bayesian algorithm for the total probability formula, the principle is according to the probability that the text belongs to a category (prior probability), the text would be assigned to the category of maximum probability (posterior probability).

In simple terms, a naive Bayes classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature [6].

Suppose the training sample set is divided into k categories, denoted as $C = \{C_1, C_2, \cdots, C_k\}$, the prior probability of each category is denoted as $p(C_j)$,where j = 1,2, ..., k.

For an arbitrary document denoted as $d_i = (w_1, \cdots, w_j, \cdots, w_m)$, whose feature words are denoted as w_j , where j = 1,2, ..., m, belongs to a specific category C_j . To classify the document d_i , is to calculate the probability of all documents in the case of a given d_i ,i.e. the posterior probability of category C_j , calculated as follows:

$$p(C_j \mid d_i) = \frac{p(d_i \mid C_j)\, p(C_j)}{p(d_i)} \qquad (1)$$

Bayesian text classification is to maximize the value of formula (1). Obviously, for all the categories given, the denominator $p(d_i)$ is a constant. Therefore, solving the maximum value of formula (1) is converted into solving the formula followed:

$$\max_{C_j \in C} p(C_j \mid d_i) = \max_{C_j \in C} p(d_i \mid C_j)\, p(C_j) \qquad (2)$$

According to Bayesian hypothesis, the feature words $w_1, \cdots w_j \cdots, w_m$ of $d_i = (w_1, \cdots, w_j, \cdots, w_m)$ are independent, the joint probability distribution is equal to the product of the probability distribution of the various feature words, i.e.

$$p(d_i \mid C_j) = p(w_1, \cdots, w_j, \cdots, w_m \mid C_j) = \prod_{i=1}^{m} p(w_i \mid C_j) \quad ,$$

therefore formula (2) becomes as follows [7,8]:

$$\max_{C_j \in C} p(C_j \mid d_i) = \max_{C_j \in C} p(C_j) \prod_{i=1}^{m} p(w_i \mid C_j) \quad (3)$$

This is the formula for the classification.

Where, the value of $p(C_j)$ is the sample size of category C_j divided by the total number of training set samples, denoted as n. There are many ways of to calculate $p(w_i \mid C_j)$, the simplest way is $p(w_i \mid C_j) = \frac{N_{ic} + 1}{N_c + M + V}$, where N_{ic} is the number of training document with the feature attribute W_i among the category C_j, N_c is the training document number of the category C_j, V is the total number of the categories, M is used to avoid the problems caused by too small N_{ic} [9].

B. The advantages and disadvantages of Naive Bayesian classifier

Naive Bayes text classification algorithm is a text classification algorithm which is simple, easy to implement and has superior performance. The independence assumptions which assumes the entire feature words are mutually independent shows all the statements between the feature words are not related, thus the implementation of the algorithm is greatly simplified [10].

Currently in the field of text classification, there are a lot of improved algorithms based Naive Bayes text classification algorithm from reducing the dimension of feature words to improving the classification algorithm itself, these improvements have greatly improved Naive Bayes text classification [11-15], but the application of the Naive Bayes text classification is limited as these improvements only be done in a particular area or for a particular category [16]. In addition, for a small training set, a rare feature words appear very random. The Naive Bayesian formula is a continued product, so the rare feature words in the text plays a dominant role [17-18]. For example, a feature word in the text only appears once in a special category, the probability of this category is calculated as the formula: $p(w_i \mid C_j) = \frac{1+1}{N_c + M + V}$,the probability of other categories are calculated as the formula: $p(w_i \mid C_j) = \frac{0+1}{N_c + M + V}$,both of the value are approximately equal to 0, which will greatly affect the outcome of $\max_{C_j \in C} p(C_j \mid d_i) = \max_{C_j \in C} p(C_j) \prod_{i=1}^{m} p(w_i \mid C_j)$.

In this paper, the algorithm is improved both on calculating the posterior probability and reducing the dimension of the feature words in documents. The formula (3) is the core of Naive Bayes algorithm. Although the formula is very easy to implement, the feature word is compared with C_j each time, then probabilities is computed. However, the feature word is a string, the efficiency is lower. If the feature words are used to run through text classification, it is also influential in lowering the dimension of the document under test. Based on the above considerations, this paper had improved Bayesian classification algorithm with the following method.

III. AN IMPROVED NAIVE BAYES TEXT CLASSIFIER

A. To improve the estimated value of $p(d_i|C_j)$ by using multivariate Bernoulli event model.

The reason for using this model is which is characterized by not considering the number of the feature word occurrences in the text for calculating $p(d_i \mid C_j)$. Text vector is weighted by a Boolean value, for the document $d_i = (w_1, \cdots, w_j, \cdots, w_m)$, w_j take the values from {0,1}, where w_j = 1 means feature words appear in the text, then the weight is 1, otherwise the weight is 0.

B. To reduce the dimension of $d_i=(w_1,\ldots,w_j,\ldots,w_m)$

Through word segmentation, any document d_i may have duplicate feature words, for example, there are $w_i = w_j$, where $1 \le i, j \le m$. Because multivariate Bernoulli event model is characterized by features not considering the number of the feature word occurrences in the text for calculating $p(w_i \mid C_j)$, when $w_i = w_j$, we can remove repeated word, $d_i^{'} = (w_1, \cdots, w_j, \cdots, w_t)$,where $t \le m$, the more repetition number of a word in the document is, the smaller the dimension of $d_i^{'}$ than the one of d_i is.

Been improved, the formula (3) becomes:

$$\max_{C_j \in C} p(C_j \mid d_i) = \max_{C_j \in C} p(C_j) \prod_{i=1}^{m} (w_i p(w_i \mid C_j) + (1 - w_i)(1 - p(w_i \mid C_j)) \quad (4)$$

Another advantage of Multi-variable Bernoulli event model is that if the features word appearing in the text, take the item is $p(w_i \mid C_j)$, if not, take the item is $1 - p(w_i \mid C_j)$.

IV. EXPERIMENT RESULTS

The paper experiments with the Starter Edition text classification data made by Sogou laboratory. It has 9 categories, 17910 documents. In this paper, 8952 documents are used as training documents, while 8958 documents are test documents, in accordance with the 1:1 ratio in the experiment. The experiment result is evaluated using precision ratio, recall ratio and F1 values used in information retrieval system.

$$\text{precision} = \frac{\text{the number of correct category for documents}}{\text{the total number of category for documents by test}}$$

$$\text{recall} = \frac{\text{the number of correct category for documents}}{\text{the total number of category for documents}}$$

$$F1 = \frac{\text{precision} * \text{recall} * 2}{\text{precision} + \text{recall}}$$

The experiment results are shown in Table 1.

TABLE I. THE EXPERIMENT RESULTS

Category	N_{tr} [a]	N_{te} [b]	Unimproved			Improved		
			precision	*recall*	*F1*	*precision*	*recall*	*F1*
Economics	991	999	0.83	0.80	0.82	0.86	0.82	0.84
IT	995	995	0.91	0.92	0.92	0.94	0.96	0.95
Health	990	1000	0.70	0.72	0.71	0.78	0.76	0.77
Sports	992	998	0.84	0.81	0.82	0.89	0.92	0.90
Traval	994	996	0.79	0.83	0.81	0.83	0.81	0.82
Education	993	997	0.71	0.73	0.72	0.75	0.76	0.75
Recruitment	997	993	0.92	0.90	0.91	0.95	0.92	0.94
Culture	1002	988	0.68	0.73	0.70	0.74	0.72	0.73
Military	998	992	0.83	0.85	0.84	0.87	0.84	0.85

a. The Number of Trainning Documents. b.The Number of Test Documents.

The precision ratio of the entire test set is raised from 80% to 85%, the recall ratio from 81% to 83%, F1 from 81% to 84% after using the improved algorithm with the multi-variable Bernoulli event model, which demonstrate that the algorithm took very good results.

V. CONCLUSION

There are many ways to compute the posterior probability of Bayesian classification algorithm. This paper has improved the original algorithm by using multivariate Bernoulli model.

Through word segmentation, the division of the feature words in the text is very high which has a direct impact on efficiency. In this paper, the dimension of feature words is reduced by removing the duplicate word appeared in document, experiment results show that the algorithm is easy to implement and efficient. But time complexity and space complexity of algorithm are not considered in the improved algorithm, which would be researched in the next phase of our work.

REFERENCES

[1] LIU Ying, Analysis on Text Classification Using Naive Bayes, Computer Knowledge and Technology(Academic Exchange), 2007.12.11.

[2] LIANG Hong-sheng, XU Jian-min, CHENG Yue-peng, An Improving Text Categorization Method of Naive Bayes, Journal of Hebei University(Natural Science Edition), 2007.3(27).

[3] ZHANG Yu-fang, CHEN Jian-min, XIONG Zhong-yang, Improved Naive Bayes Text Classification Algorithm, Journal of Guangxi Normal University(Natural Science Edition), 2007.2(25).

[4] AN Yan-hui, DONG Wu-zhou, YOU Zi-ying, The text categorization study on improved Naive Bayes, Journal of the Hebei Academy of Sciences, 2007.24(1).

[5] Jingnian Chen, Houkuan Huang, Shengfeng Tian, Youli Qu, Feature selection for text classification with Naive Bayes, Expert Systems with Application, 36(2009)5432-5435.

[6] Kim.S., Han.K., Rim.H., Myaeng.S., Some Effective Techniques for Naive Bayes Text Classification, IEEE Transactions on Knowledge and Data Engineering, (2006)18(11),1457-1466.

[7] WANG Jun-ying, GUO Jing-feng, HUO Zheng, Design and Implementation of Chinese Text Categorization System, Microelectronics & Computer, 2006.23.

[8] Gao Yuan, Liu Da-zhong, A Comparison Study of Chinese Text Categorization, Science & Technology Information, 2008.2.

[9] YANG Ye, PENG Hong, LIN Jia-yi, Chen Shao-jian, The Bayesian Text Categorization Based on Extraction of Effectual Features, Systems Engineering, 2004.9(22).

[10] CHEN Jing-nian, HUANG Hou-kuan, TIAN Feng-zhan, QU You-li, Method of feature selection for text categorization with bayesian classifiers, Computer Engineering and Applications, 2008.44(13).

[11] YU Fang, JIANG Yun-fei, A Feature Selection Method for NB-based Classifier, Acta Scientiarum Naturalium Universitatis Sunyatseni, 2004.5(43).

[12] WANG Xiao, An Improved Bayesian Text Classification Model, Modern Computer, 2008.1.

[13] LIU Hua, An improved Bayesian text categorization system, Journal of Jinan University(Natural Science & Medicine Edition), 2007.1(28).

[14] LUO Hai-fei, WU Gang, YANG Jin-sheng, Way of text classification based on Bayes, Computer Engineering and Design, 2006.12.

[15] WEI Xiao-ning, ZHU Qiao-ming, LIANG Xing-yan, Using Bayesian in Text Classification with Participle-method, Journal of Suzhou Vocational University, 2008.1(19).

[16] WAN Di-Fei, FAN Xing-Huan, WANG Guo-Yin, Two-class Text Categorization Method Based on Naive Bayes and GA, Computer Science, 2008.4(32).

[17] Bai Liyuan, Xiao Le, Huang Hui, Ding Wei, A BAYES CLASSIFICATION ALGORITHM BASED ON BOOTSTRAP AVERAGING, Computer Applications and Software, 2007.9(24).

[18] YUAN Fang, YUAN Jun-ying, Naive Bayes Chinese text classification based on core words of class, Journal of Shandong University(Natural Science), 2006.3(41).

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 270-273

Braces Surface Generating Algorithm Based on the Surface of Triangles

Jin Jihong, Liu Shuzhi
Department of Computer and Information Engineering, Jiaozuo Teachers College, Jiaozuo, China
Email: jinjihong9082@163.com
shuzhi_liu163@163.com

***Abstract*—Along with the development of computer technology, the orthodontic tooth problem using computer-aided technology is getting more and more attention. The concept of invisible braces is popular with numerous patients when it put forward in medical field. The braces surface generating is crucial to overall production. This paper studied a Triangle-based transition of surface generation algorithm for transition curved surface in Dental orthodontics. Based on the tooth outer surface s_1 and s_2 which is obtained by offsetting s_1 a certain distance d, s_3 is generated by smoothly connecting s_1 and s_2. Though melting s_1, s_2 and s_3, tooth brace surfaces would be obtained. The algorithm is simple and straightforward, avoiding a large number of calculations. The simulation result showed the transition surface is natural and smooth, suited for generating small-scale transition curved surface.**

***Index Terms*—Virtual orthodontics, STL files, Transition surface, Triangular facet**

I. INTRODUCTION

With the rapid development of computer technology, computer-aided correction of abnormal tooth technology is getting more and more attention. First the patients teeth data could be obtained by some scan technology such as CT, MRI, or other such tomography [1]; Then the teeth data would be three-dimensional reconstructed ; The result of the process is a STL file, which coming from the corresponding teeth data; Next the STL file would be read, and then a digital model is be output; For teeth model, in order to simulate the whole process of correcting abnormal tooth[2], the first step is to get teeth segmentation, and then plan the orthodontics path, finally exert resistance force on deformity tooth to destination according to plan. The overall treatment plan, the patient was instructed to wear different braces in different periods, which can achieve the expected effect. Braces using rapid prototyping manufacturing system are a transparent flexible plastic, which does not affect patients' appearance. In addition, patients can also see the virtual treatment process before the real treatment, early know the treatment results. In addition to the "invisible", transparent braces by computer design also can control treatment time and treatment patterns, and in orthodontic treatment phase in particular to correct of certain deformity tooth.

The key technology is braces surface generation in using this method to produce teeth braces. This paper proposes a simple quick based on the surface of triangles in transition though the analysis of the current transition surface algorithm.

II. CURRENT TRANSITION SURFACE GENERATING METHODS OF RESEARCH ANALYSIS

A. Radius Transition Method

Radius transition method [3] can direct, simple to produce the transition and the ridge, connecting point, the contour line are automatically generated. The main problem is ridge generation in radius transition method, so many efficient algorithm emerge. In which SSI algorithm is commonly used**.** In this algorithm, taken the equidistant surface of the base surface as ridge line, the center of rolling ball along the ridge line and the radius of the ball is equal to the distance of original surface and equidistant surface, thus the transition surface will be swept by radius**.** Others use rolling ball to transition quadric surface, but sweep a tube surface which even in the relatively simple cases is high **a**lgebraic surface**.** With high complexity, this method is not benefit for calculation**.** Because curved surface shape only is arc using radius of transition, whose application scope is corresponding limited.

B. PDE Method(Partial Differential Equations)

PDE surface [4] using a group of elliptic differential equations produce curved surface. The thoughts originated from taken the transition construction problem as boundary value problems of partial differential equation. It can be found this method can easily obtain surface structure of practical problems. PDE surface can only be used for the quadrilateral region, and PDE surfaces shape control is imperfect. PDE surface shape is decided by boundary conditions and the choice of partial differential equations.

C. Energy Method Based on Physics

Energy method based on physics [5,6] taken transition tangent touch line and vector of the base surface and other boundary lines as constraints, use physical surface modeling technology to generate transition surface. The basic idea is: the surface as the elastic deformation of the thin shell, introducing the energy paradigm, establishing surface deformation control equation through mechanics

Jin Jihong,(1977-09), Jiaozuo teachers college

principle, and then using numerical method to get numerical solutions satisfied with certain constraints.

Compared with other methods, the energy method based on physics has incomparable advantages. First, it can be used to construct not only quadrilateral domain transition, but for n edge domain, whose characteristics of this are not easy to embody in other ways. Secondly, the structural transition surface require a few conditions, only a few boundary and its vector which can be only determined. The final, and is the most important is that this method can not only ensure the continuous transition of the border, and can ensure merge the generation region of transition surface, and the transition through some intermediate constraint line. These can well satisfy actual engineering requirements. But this method also exist some problems such as: excessive computation for structural transition surface, and difficult to local shape control.

III. THE TRANSITION GENERATION ALGORITHM BASED ON TRIANGLE SURFACE

In view of the above methods, this paper presents a new method to generate transition, which develop a simple and direct and can realize local control algorithm. The specifics are as follows. Based on the tooth outer surface s1 and s2 which is obtained by offsetting s1 a certain distance d, s3 is generated by smoothly connecting s1 and s2. Though melting s1, s2 and s3, tooth brace surfaces would be obtained.

Taken digital dental model obtained from tomography method as the research object, with STL file format for storing and triangles as mesh, the algorithm is this paper Concentration on.

Triangles are exact subdivision linear surface connected to edges and vertices of in 3-D space, where each edge contained in two triangles in most. A STL file, obtained from 3-D entity model after triangle treatment, is a file format applied for rapid prototyping (RPM). STL model is approximate the original CAD entity data model with many small spatial triangle plane. For three-dimensional entity description and explanation it has uniqueness. In addition, it is a collection of more disorderly triangle facets, each triangle facet made of four data items, namely that three vertex coordinates (x, y, z) and the normal vector of the triangle facets (lx, ly, lz), those abide by Right-Hand Rule. The data structure of STL file is very simple, no adjacency and connections.

The basic thought of based on triangles transition surface algorithm is with the teeth mobile real-time to constantly adjust triangles in the process of correction teeth. In the adjustment, two kinds of methods are available: triangles deforming method and increase of triangles. The former is not changing the topological structure of triangles, through the adjustment of the local scope to the shape of the triangles to generate transition method; the latter is to add triangles real-time while teeth move. These two methods were eventually needed according to local shape of late smooth processing. The two algorithms are simple, less computation, and satisfy the transition process of orthodontic tooth surface requirements, but only applicable to small range of transition.

A. Transition Surface Method Based on the Triangle Deforming

For digital triangular mesh model, the process of orthodontics, from the perspective of geometry, is local triangular mesh translation operation or rotating operation. Translation and rotating operation would make separation or overlapping between grids. The transition surface generation method based on the triangle mesh deformation is suitable for the separation and overlap between grids.

Triangular mesh deformation process is decided by the offsets from translation or rotation operation. The offset reflects on the offset vector of triangular mesh point, and the migration offset vector into other adjacent vector by triangle vertex, which make triangular mesh cause deformation, so recursive until the offset vector is less than a specified value. This process is the guarantee of the overall shape of the triangle mesh not big deformation; also meet the requirements of keeping local shape. Deformation process of separation triangular mesh is shown in Fig. 1.

In Fig.1: AB and BC belong to boundary; B is vertex need to move; Offset vector $\overrightarrow{Bb}$ is called for trends vector. Trends vector can cause F&G vertices (level-1 neighbors vertices of B) moving a fewer distance than B. The offset vectors named as $\overrightarrow{Ff}$ and $\overrightarrow{Gg}$, and we call them as result vectors of the trends vector $\overrightarrow{Bb}$. The same is that $\overrightarrow{Ff}$ and $\overrightarrow{Gg}$ will also cause the level-2 neighbors vertices of B, such as D&E moving a less than F&G. The offset vectors are called respectively as $\overrightarrow{Dd}$ and $\overrightarrow{Ee}$, etc. Next, $\overrightarrow{Ff}$ and $\overrightarrow{Gg}$ are taken as trends vectors of $\overrightarrow{Dd}$ and $\overrightarrow{Ee}$. On the other hand, $\overrightarrow{Dd}$ and $\overrightarrow{Ee}$ become result vectors of $\overrightarrow{Ff}$ and $\overrightarrow{Gg}$. And so forth, offset displacement of adjacent vertices will be calculated until modulus of result vector is less than a threshold. A relationship is abided by between results vector and trends vector:

$$\overrightarrow{Gg} = \mathrm{k}\,\overrightarrow{Bb}$$

There is many methods to obtain k value. For the simple aim, we can use the linear relation. In this paper, through Several tests, we determine the scope of k: 0.2-0.3. Mesh deformation process reasoned by overlap of triangular as figure2 shows.

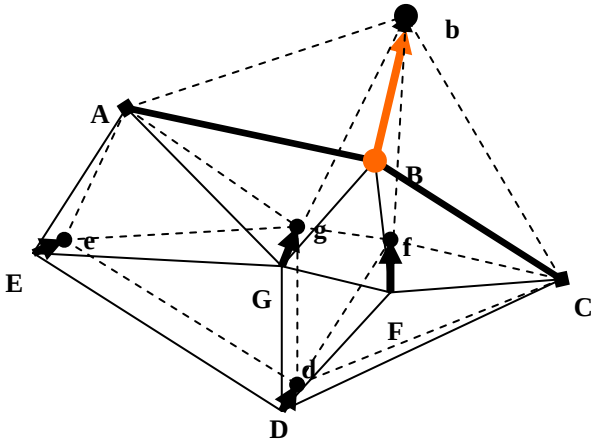


Figure 1. 3-D mesh stretching

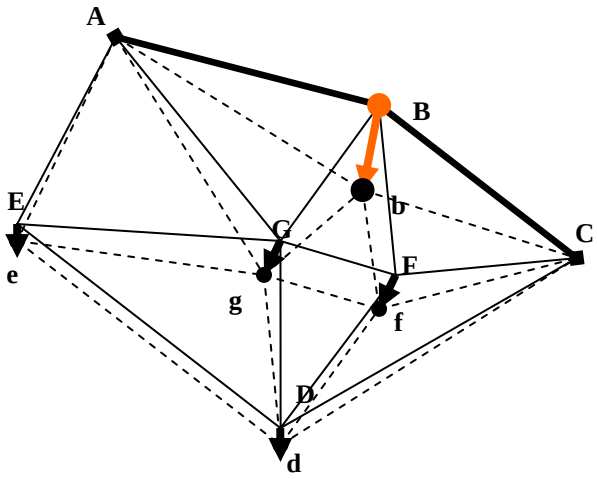


Figure 2. 3-D mesh compression

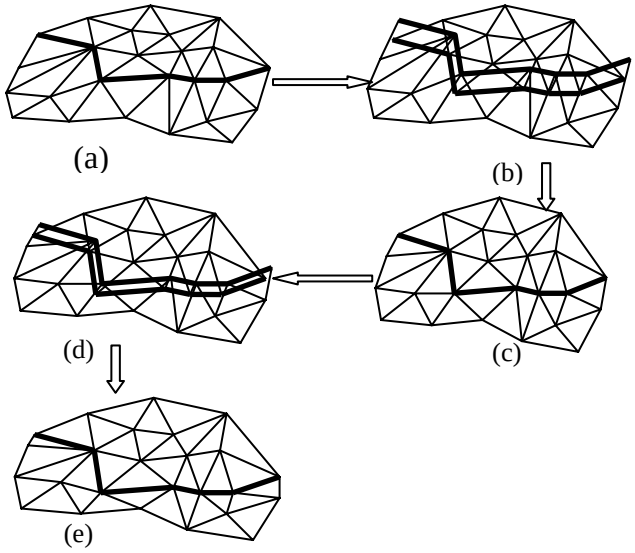


Figure 3. Triangular mesh twice compression effect

Fig.3 shows mesh deformation result after twice compression and overlap.

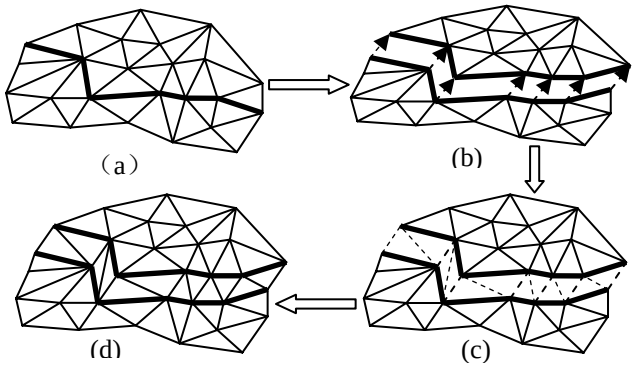


Figure 3. Transition surface method based on the increase of triangles

B. Transition surface method based on the increase of triangles

In translation operation or rotating operation, if there is not only grid separation but also grid overlapping, for separation, we use increasing triangles to generate transition surface which makes calculation speed, and for overlapping, we use grid deformation based on triangle.

Transition surface method based on the increase of triangles is realized when triangular mesh was divided and began to move, through real-time generation triangular mesh in the division area. In the process of separation, triangular mesh in expanding, there are always two vertices of edge.

Specific steps are:

Step 1: Sequentially store pair wise vertices, when the partition distance reach the standard, connecting corresponding two vertices, as shown in Fig.4 (b);

Step 2: One vertex connect the next according to the order of storage, as shown in Fig.4 (c). Thus, new triangular mesh is formed in the division area.

Step 3: By modifying the vertex, line, triangles of model, new triangles data will be added into digital teeth model.

Step 4: Instead of the original boundary, new triangles will continue to separate the grid.

Schematic diagram shown as shown in Fig. 4.

IV. THE SIMULATION ALGORITHM REALIZING

The algorithm realize in Microsoft Visual Studio c ++ 6.0 environment, with OpenGL graphical development platform simulating.

Orthodontics process needs multiple adjustments, and the transition surface generation also need teeth move many times to accomplish. The teeth move amplitude cannot too big. Specific mobile distance is based on the analysis on the medical. This paper makes observation and real-time adjustment.

Fig.5 shows the rendering of digital teeth model before orthodontics while Fig.6 is the rendering of teeth digital model after transition surface generation. In Fig. 5 and 6, from down to up, teeth move gradually increasing amplitude. From Fig. 6 can be found transition surface generated in the first and second tooth movement process is natural and smooth, can satisfy the requirements on the whole. But in Fig.5 from the bottom third tooth movement

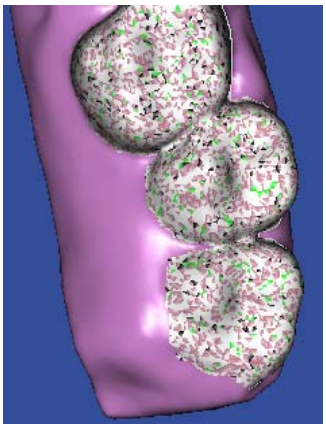


Figure 5. The triangles of teeth digital models

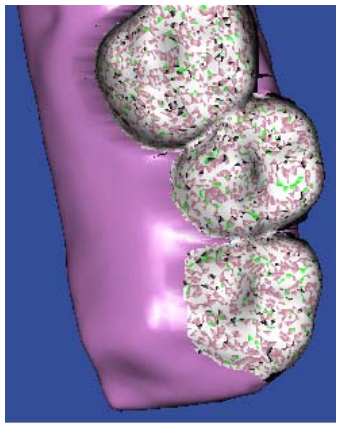


Figure 6. The triangles of teeth digital models with transition surface

range is too big, result in that the transition surface tension in Fig.6 is obvious, cannot satisfy requirements.

From the above analysis, it can be concluded that transition surface generation based on increasing triangles and deforming is adapt to a small range of teeth mobile, and the transition surface is natural, level and smooth. Therefore, in the use of orthodontic brackets, it cannot be eager to hope for success, exerting too much power, resulting in the teeth movement is too big, easy to cause the teeth and gums torn open.

V. CONCLUSION

Based on comparison with the several transition surface generation method, the paper puts forward a new kind of human-computer interaction transition method of generating surface. Then, outer surface of the tooth model generate a new tooth surface layer though offset, the new surface is taken as the model of the outer surface of the braces model. Two curved surface as base surfaces merged by the transition surface, so braces surface is achieved. The algorithm is simple, direct, avoiding a lot of calculation and transition surface is natural and smooth, applicable to such small range of orthodontic tooth.

ACKNOWLEDGMENT

The work described in this paper is supported by Natural Science Foundation of Shaanxi Province under Grant No.2004F37 are highly appreciated. Special thanks are given to Wang Jiping and Professor Li Zhanli.

REFERENCES

[1] Zhou. J.M, Bai. Y.X, and Wang. B.K, “Development Of a highly precise digital three-dimensional reconstruction system for dental cast”[J]. Beijing Journal of Stomatology, 2004,12 (1)

[2] Zou. B.j, Lv .G.F, Zhou. H.Y, and Sun. J.G, “Design and implementation of a virtual operation system for beautifying lady’s face. “[J], Journal of system simulation, 2003,15(6):898-901

[3] Wu. G.L, Lin. J.P, and Li. C.X, “ Research on blending for vertex of multi-surface in surface modeling”[J]. China mechanical engineering, 2002, 13 (5) : 12-14

[4] Ma. L, Zhang. X, and Zhu. X.X, “Transition surface generation With partial differential equations “[J]. Journal of engineering graphics, 1995, 16 (1) : 1-8.

[5] Peng. F.Y, Zhou. Y.F, and Zhou.j, “Algorithm of surface smoothing based on extended energy minimization.” [J]. Journal of huazhong university of science and technology, 2002,30 (2) : 5-8

[6] Zhao. Z.Z, Lin. H, Shi. X.M,”Optimizing data model of parts with energy function”[J], Machinery design and manufacture, 2000, (2) : 24-26

[7] Farouki. R.T, “The Approximation of Non-degenerate Offset Surfaces” [J]. Graphical Models and Image Processing, 1986, 3(1):15-43

[8] Piegl. L. A, Tiler. W, “Computing Offsets of NURBS Curves and Surfaces” [J]. Computer Aided Design, 1999, 31(2):147-156

[9] Wallner. J, Sakkalis. T, and Maekawa. T, “Self-intersections of offset curves and surfaces “[J]. International of Shape Modeling.2001, 1:1-22

[10] Jung. W. H, Shin. H. Y, and Choi. B. K, “Selfintersection removal in triangular mesh offsetting “[C] . CAD’04 conference, Thailand, 2004:477-484

[11] Pottman. H, Lü. W, and Ravani. B, “Rational ruled surfaces and their offsets”[J]. Graphical Models and Image Processing, 1996, 58(6): 544-552

[12] Jang. D. G, Park .H, and Kim. K, “Surface offsetting using distance volumes “[J]. The International Journal of Advanced Manufacturing Technology, 2005, 26:102-108

[13] Farouki. R. T, “ Exact offset procedures for simple solids” [J]. Computer Aided Geometric Design, 1985, 2(3): 257-279

[14] Martin. R, “Principal Patches A New Class of Surface Patch Based on Differential Geometry “[C].in: Hagen P J, eds, Proc.Eurographics’83, North-Holland, Amsterdam

[15] Pottmann. H, “Rational curves and surfaces with rational offsets” [J]. Computer Aided Geometric Design, 1995, 12(2): 175-192

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 274-278

3D Rapid Modeling for the Foundation of Steel Headframes

Xu Wenpeng[1], Qiang Xiaohuan[2]
[1] College of Computer Science & Technology, He'nan Polytechnic University, Jiaozuo, China
Email: wpxu@hpu.edu.cn
[2] College of Surveying & Land Information Engineering, He'nan Polytechnic University, Jiaozuo, China
Email: qiangxh@hpu.edu.cn

***Abstract*—We present a rapid modeling approach for the foundation of steel headframes, in which modeling are quickly constructed by some design parameters. The design of Headframes' foundation is always a cumbersome and difficult process. Currently it need design in 3D environment and often need finite element analysis according to its importance while traditional design is focused on 2D construction drawings. Based on the analysis of its design process, we points out that the core work of its 3D modeling is vertices calculation. 3D coordinates calculation method of vertices is proposed, and finally a system implementation is given. The results show that a complex modeling task of headframes' foundation can be quickly implemented with our system. It can effectively reduce the design difficulty and increased design efficiency.**

***Index Terms*—CAD, Rapid Modeling, Steel headframes foundation**

I. INTRODUCTION

In the mine lifting system, headframes are an important building, which are bearing the head sheave to provide the lift height of the cage, unloading bend channel, falling protector. They are generally composed with two parts: the guideframe, the backstay. The foundation is a significant part of headframes, which transfer the upper load to the ground and maintain the overall stability of main body. The guideframe is established on the bearframe of the well neck, which is the foundation of the guideframe. As for the backstay, its foundation is commonly constructed on natural ground and need design seperately. Therefore the foundation of headframes is referred to the backstay foundation in general.

There are two aspects in the design of headframes foundation which are mechanical calculation and construction drawing. The task of mechanical calculation is to obtain the key dimension of the foundation using hand-calculated traditionally. Then the detailed dimension of each part is calculated from the key dimension. Based on these dimension, construction drawing can be ploted. Some mechanical calculation work has been replaced by finite element analysis(FEA) software like ANSYS, ABAQUS etc. Its workload and difficulty have been effectively reduced. As for construction drawing part, it need convert the foundation information between 3D and 2D, sometime it need calculate unfolded drawing of the foundation formwork. The whole process still lack the professional software support and the work is very complicated and error-prone. With economic development and energy demand quickly increasing, the mine building needs better and faster to complete, which makes the traditional design method can not meet the mine construction.

In engineering and product design, the computer can help designers responsible for computing, information storage, drawing and other work. We call these work as Computer aided design(CAD) . CAD can reduce the heavy workload, shorten the design cycle and improve design quality. There is few work about headframes foundation CAD. Shi sanyuan[1,2] makes some research about 2D CAD on headframe foundation and its result can directly to generate the construction drawing. Xu laiyong[3] uses AutoCAD to draw the wireframe model of the foundation, then generate the final construction drawing. Considering the importance of the foundation, the foundation need test its strength and stiffness to meet the design requirement through FEA. Therefore, it is not enough with construction drawing for design. We need rapid establish 3D model of the foundation too and this is the purpose of this paper. It has practical signnificance to reduce the design workload and improve the design efficiency of headframe foundation.

II. MODELING ANALYSIS OF HEADFRAMES FOUNDATION

A. Design process and analysis

The load which the foundation beared is from the backstay and its direction is same as the axis of the backstay column. As shown in Fig.1, where symbol 1 represents the guideframe, symbol 2 represents the backstay and symbol 3 is the foundation of the backstay. The foundation can be divided into two components structurely: main body and bottom slab. Main body is a irregular hexahedron while bottom slab is a cuboid. According to the basic difference between the specific type, the foundation can be divided into single and double vertical-plane categories which single vertical-plane foundation has only one vertical plane as the side face in its main body and double vertical-plane foundation has two adjacent vertical planes, as shown in Fig 2 and Fig. 3. Compared from two figure, double vertical-plane foundation can be more economical with the same design parameters.

The design process of the foundation mainly has three steps according to headframes design reference[4]:

(1) Calculating load: According to the design requirement, three load need calculate: the force from the backstay column, self weight and the weight of the soil above, as shown in Fig. 4(a);

(2) Determining bottom slab size: Trial method is used to obtain the slab plane size b_1 , b_2 to meet the requirements of anti-sliding, anti-overturning and ground capability check accroding to the current national standard, “National Codes for Design of Building Foundation” (GB50007-2002);

(3) Determining the top face size: Considering the enough concrete pressure area to the force from the backstay column, the top face size a_1 , a_2 is determined, as shown in Fig. 4(a).

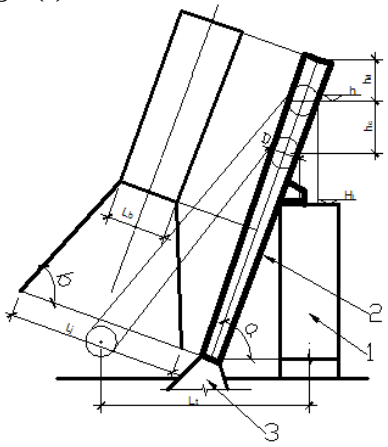


Figure 1. Steel headframes.

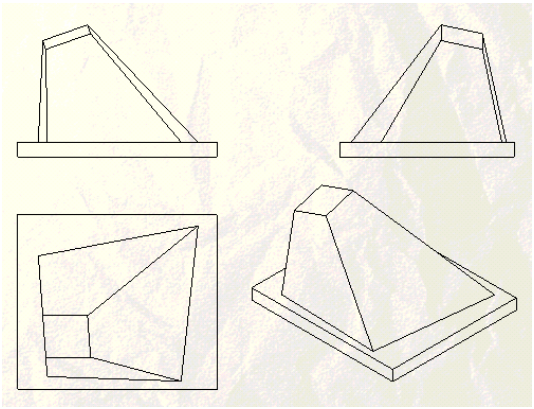


Figure 2. Single vertical-plane foundation.

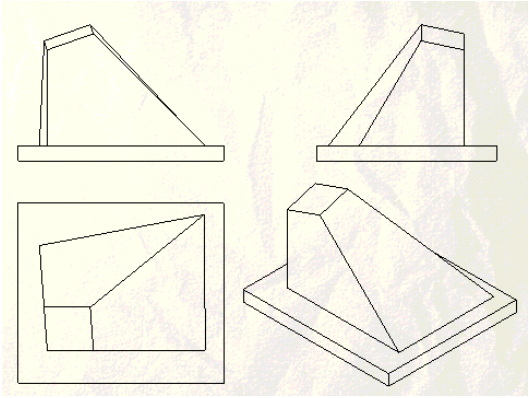
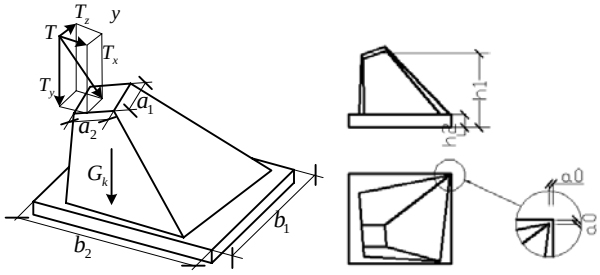


Figure 3. Double vertical-plane foundation.

Furthermore we can get the total height h_1 , the slab height h_2 from the overall design conditions. The length apart the slab edge a_0 is a construted dimension which is generally assign to 150mm, as shown in Fig. 4(b). The dip angle a and b can be get from the overall information of the headframe, as shown in Fig. 1. 3D model of headframe foundation can be constructed with the size information above.



(a) loads and sizes (b) height and a0

Figure 4. Design analysis.

B. Modeling process and analysis

Based on design analysis above and design mannual[5], 3D model of headframes foundation can be constructed by the following three steps, as shown in Fig. 5:

(1) Locating the top face of main body: The normal of the top face is parallel to the center axis of backstay column which can be obtained from two dip angle a and b from backstay column. Then the top face and its four vertices(v1, v2, v3, v4) can be located and constructing with the top face size a_1 and a_2.

(2) Constructing bottom slab: Bottom slab can be located according to that its center must be coincide with the center axis of backstay column. Then with the plane size b_1 , b_2 , and height h_2 , we can construct it and obtain its eight vertices from v9 to v16.

(3) Determining the bottom face of main body: The bottom face of main body is determined by single vertical-plane or double vertical-plane type. The model process is complicated compared with the work above. To take single vertical-plane type for example, it can be subdivided into four steps as following:

(a) To take v10 as a known vertex, v6 can be located for it is apart with a0 in both directions from v10.

(b) The point v5 is an intersect point, which is generated from intersecting with a vertical plane p1-4(plane including v1 and v4, the same below), a plane p1-2-6 and a horizontal plane p9-10-11-12.

(c) If the plane p2-3-6 intersects with the plane p9-10-11-12, we can obtain a intersected line. We can certainly find a point on this intersected line which it is apart from edge e11-12(edge connected by v11 and v12) with a0. And this point is v7.

(d) The point v8 is a point intersected from the plane p2-3-6, p9-10-11-12 and the vertical plane p1-4.

As for double vertical-plane type, another vertical plane is formed with v3 and v4. Then v7 is the intersected point from the plane p2-3-6, the horizontal plane p9-10-11-12 and the vertical plane p3-4. The other process is the same as single vertical-plane type.

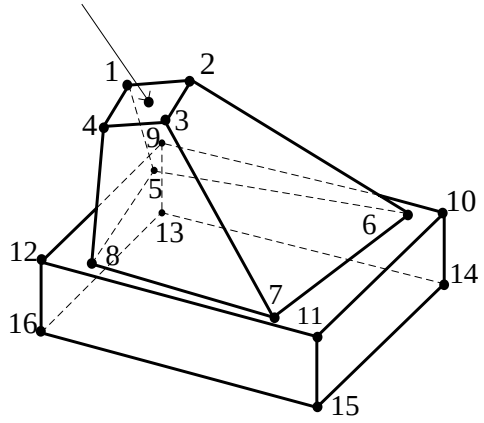


Figure 5. Vertices of single vertical-plane foundation.

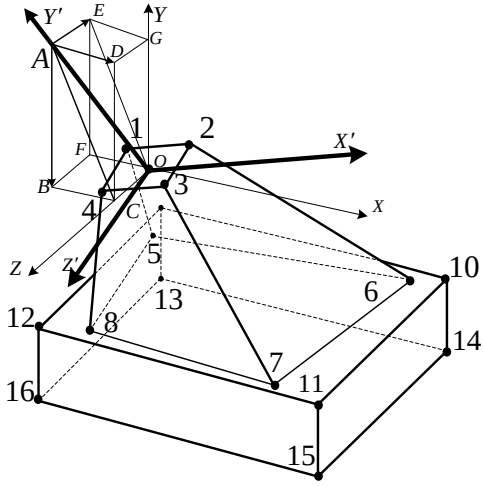


Figure 6. Vertices calculation.

III. VERTICES CALCULATION

Based on the analysis above, we can see that if we can locate from v1 to v16 we can easily construct the foundation model. Therefore the key task of model is calculation of these sixteen vertices.

With the coordinate origin set to the top center of main body, X axis parallelled to edge e11-12, Z axis paralleled to edge e10-11 and Y axis is vertical, a 3D coordinate system(CS for short) can be established, as shown in Fig. 6. Then the vertices can be divided into three types to calculate as following:

A. Vertices in the top face of main body

These vertices include v1, v2, v3 and v4. They used to located by projection method traditionally according to design mannual[5], which need convert between 2D and 3D and every vertex need consider. Here we present a transform method to calculate them: First we establish the transform matrix, then all the vertices new coordinates can be obtained from multiplying their old coordinates with the matrix.

For clearly description we establish another CS as old CS $OX'\ Y'\ Z'$, which set its origin at the point O, X' axis parallelled to edge e1-2, Z' axis paralleled to edge e2-3 and Y' axis is perpendicular to the top face, as shown in Fig. 6. Then the coordinates of v1 can be easily represented as $(-a_2/2, 0, -a_1/2)$ in $OX'\ Y'\ Z'$. However, we need calculate its coordinates with transformation matrix in OXYZ, which is named as new CS. According to [6], the transformation matrix can be obtained by transforming CS OXYZ coincided with CS $OX'\ Y'\ Z'$, which have two steps as following:

(1) CS OXYZ is rotated with angle ∠AOE in X axis and a transformation matrix is generated:

$$T_{X\theta} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{bmatrix}, \theta = \angle AOE = \pi/2 - b \quad (1)$$

Note that rotating angle is the angle ∠AOE, not ∠DOG.

(2) Then CS OXYZ is rotated with angle ∠EOG in Z axis, then another matrix is generated:

$$T_{Z\alpha} = \begin{bmatrix} \cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}, \alpha = \angle EOG = \pi/2 - a \quad (2)$$

Based on two matrix above, we can obtain the final matrix T:

$$T = T_{X\theta} \cdot T_{Z\alpha} = \begin{bmatrix} \cos\alpha & \sin\alpha & 0 \\ -\sin\alpha\cos\theta & \cos\theta\cos\alpha & \sin\theta \\ \sin\alpha\sin\theta & -\cos\alpha\sin\theta & \cos\theta \end{bmatrix} \quad (3)$$

Therefore the coordinate (x y z) of v1 can be calculated as the followed expression:

$$(x \quad y \quad z) = (-\frac{a_2}{2} \quad 0 \quad -\frac{a_1}{2}) \cdot T \quad (4)$$

The coordinate calculation of other vertices is same as v1.

B. Vertices of bottom slab

Compared to vertices above, the coordinates calculation of the vertices in the bottowm slab is more easy for the bottom slab is a box. The key is to locate the center of its bottom face, which is in the total force direction from the backstay column. According to this, if the force direction is regarded as a line, the line equation can be resolved. Then a point can be intersected from the force line and the plane which the bottom face is lied. The intersected point is the center of the bottom face. Based on the center and the slab height, we can calculate eight vertices in the bottom face and top face. The detailed process of calculation is leaved out here.

C. Vertices in bottom face of main body

According to modeling analysis above, v6 is related with v10 by the constructed length a0 and other three vertices is located by intersected from three planes. Therefore the core work is the intersected calculation between three planes.

Taken three planes equation as followed: $a_1x+b_1y+c_1z=d_1$, $a_2x+b_2y+c_2z=d_2$, $a_3x+b_3y+c_3z=d_3$, there is a intersected point between these three planes only when D≠0. And the coordinates of the intersected point is followed: $x=\frac{D_x}{D}$, $y=\frac{D_y}{D}$, $z=\frac{D_z}{D}$

where

$$D=\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}, \quad D_x=\begin{vmatrix} d_1 & b_1 & c_1 \\ d_2 & b_2 & c_2 \\ d_3 & b_3 & c_3 \end{vmatrix}, \quad D_y=\begin{vmatrix} a_1 & d_1 & c_1 \\ a_2 & d_2 & c_2 \\ a_3 & d_3 & c_3 \end{vmatrix},$$

$$D_z=\begin{vmatrix} a_1 & b_1 & d_1 \\ a_2 & b_2 & d_2 \\ a_3 & b_3 & d_3 \end{vmatrix}.$$

Based on the method above, v5, v7 and v8 can be obtained respectively.

IV. System Implementation

We have implemented a rapid protyping system of headframe foundation with C# as the development language and SolidWorks as the development platform. The system can quickly build 3D model of the foundation with simple input of design parameters. Then it can easily generate 2D construction drawing with help of SolidWork platform.

The System mainly consists of three modules: (1) design parameter input; (2) vertices calculation; (3) model building. System framework is shown in Fig. 7.

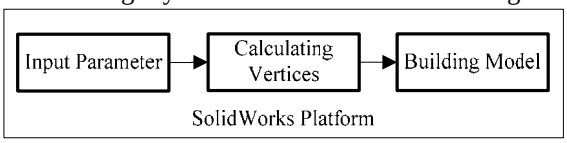


Figure 7. System framework.

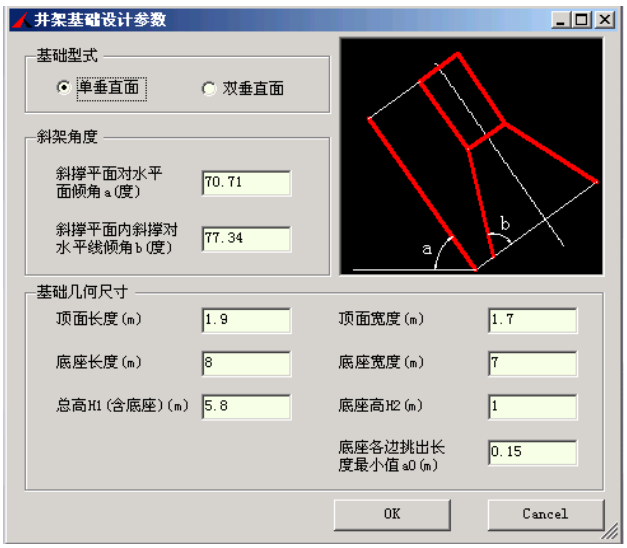
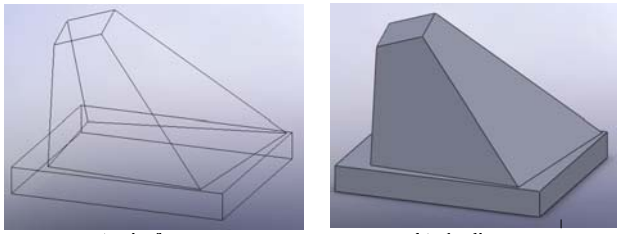


Figure 8. Interface of input module.

The main function of parameter input module is to realize the user interface which fulfill the input tasks of design parameters. According to the common design requirements, the parameters is divided into three categories: (1) the type of foundation: single vertical-plane or double vertical-plane; (2) the angle of backstay column to determine the force from the column; (3) the key size of foundation parts, such as the size of the top face, bottom slab etc. All the parameter need is detailed in Fig. 8.

Vertices calculation module complete the task of vertex coordinate calculation from v1 to v16. For there is some matrix and det in vertices calculation, we can add some matrix and det class according to object oriented programming to improve the code readability and maintainability.

Model building module establish 3D foundation model from the vertex coordinate of v1-16 according to modeling analysis, which can consist of two steps as followed: (1) Building main body which is lofted from the plane p1-2-3-4 and p5-6-7-8, which is shown in Fig. 6. (2) Building bottom slab: firtstly a Rectangle is created from v9 to v12, then the slab is obtained by rectangle extrued with its height. The final model is shown in Fig. 9. And a sketch drawing from it is easily generated, which is shown in Fig. 10.



a) wireframe b) shading

Figure 9. Single vertical-plane foundation.

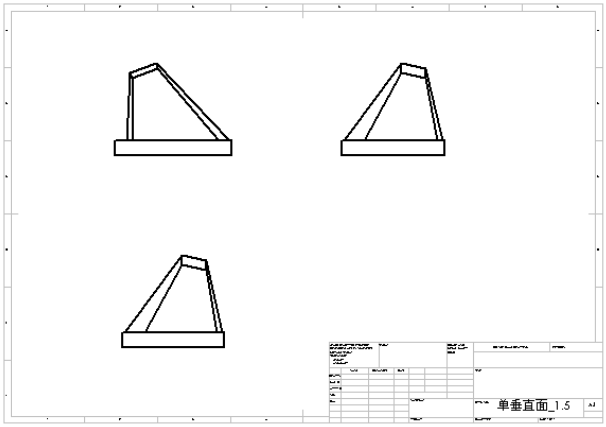


Figure 10. A sketch drawing of foundation.

IV. Conclusion

In this paper, we proposed a rapid modeling method for the foundation of steel headframes based on CAD technology according to the design need of headframes foundation. Compared with the traditional 2D method, this solution is more general and simple although its final

result is 3D model. It can be extended to other similar object construction. Next it need optimize the design result with FEA software.

ACKNOWLEDGMENT

We would like to thank the Henan provincial science research program (No.2008B520014), and the young backbone teacher Program of Henan Polytechniacl University (No.649064) for supporting this work.

REFERENCES

[1] Shi Sanyuan, Wang Baoyan, Feng Yixiao, "Computer Aided Design for the Footing of Steel Headframe," *Journal of Hebei Institute of Architectural Science and Technology*, 1999, (3), pp. 16-17.(in Chinese)

[2] Shi Sanyuan, Feng Yixiao, Wang Baoyan, "CAD Method for the foundation of steel headframes backstay," *Engineering Mechanics, supplement*, 1999, pp. 934-939. (in Chinese)

[3] Xu Laiyong, "2D CAD Method for headframes foundation," *Jiangsu Coal*, 1998, (04), pp. 34-35. (in Chinese)

[4] China Coal Construction Association, *Code for design of the mine headframes*, Beijng：China Planning Press, 2006. (in Chinese)

[5] China Coal Society, Coal Mine Construction Engineering Committee, *Construction manual of reinforced concrete buildings and structures*, Beijng：China Architecture & Building Press, 1995. (in Chinese)

[6] Chen chuanbo， Lu Feng, *Fundamentals of Computer Graphics*, Beijng：Electronic Industry Press，2007. (in Chinese)

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 279-282

Research on Virtual Digital Campus Platform

Lv Aili[1], Xue Mingxia[2], Zhao Lin[3]
[1]Modern Educational Technology Center, Henan Polytechnic University, Jiaozuo, China
Email:lvaili0210@yahoo.com.cn
[2]Accountancy Academy，Jiaozuo University
[3]Neusoft Institute of Information, Dalian, China
Email:xmxxmx16888@163.com, zlwozl@gmail.com

***Abstract*—In the course of construction, datacenters encounter many problems, such as low-efficient resource usage, difficult management and maintenance. In order to resolve those problems, this paper introduces virtualization technology. After proposing virtualization technology and its classification, it presents the design of digital campus architecture based on VMware Infrastructure, and discusses function design and structure of the platform. It presents how to carry out high reliability of system by prepare rules. With virtualization technology, server resources can be saved effectively; meanwhile, datacenter's construction and operating costs can be reduced.**

***Keywords*-Datacenter, VMware, Virtualization, Virtualization Technology**

I. INTRODUCTION

With the development of informatization, application systems and users based on campus network, such as education management platform, research management system, office management system and OA etc., are gradually increased. Fast growing applications demand more and more servers and storage resources. How to ensure that applications can operate long and steadily, how to maximize resources utilization, reduce difficulty of resource management, and establish a unified and green data center are the problems we must to deal with. Virtualization technology is a solution [1], which is a broad term, the main problems need to resolve include [2]: (1) multiple programs can share the same device by segmentation hardware; (2) software porting can take place between various operating systems; (3) old application systems can run on new computers.

With VMware Infrastructure virtualization scheme, this paper applies virtualization technology to informatization construction. It consolidates all storage resources into a unified and measurable resource pool [3], divides them into distributable unit precisely and distributes the resources according to the allocation principles and strategies pre-ordered.

II. VIRTUALIZATION PRESENTATIONS

Virtualization [4] is a technology, which takes a physical system (hardware, operating system (OS), and applications) and packages them into one or more independent partitions, and each partition can simulate a stand-alone server as needed. The substance of virtualization is to administrate and reallocate the computing resources by middle-tier, so as to realize the aim of maximizing the resource utilization.

As figure1 (a) shown, virtualization technology implements system virtualization by means of adding a thin Virtual Machine Monitor (VMM) on existing platform[5], for instance ,virtual processor, virtual MMU(Memory Management Unit) and virtual I/O system. From the point view of applications, programs running on virtual machines are same as running on corresponding physical computers, as shown as figure1(b).

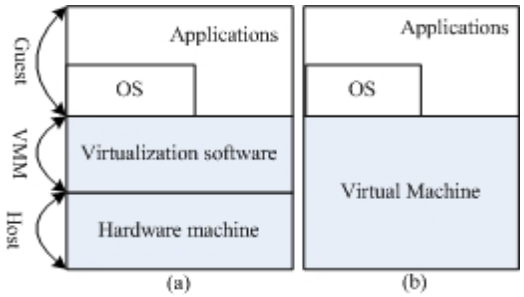


Figure 1. Virtualization software and VM layers.

A. Server Virtualization

Server virtualization [6] refers to running different operating systems within the same box. With this technique, a host can support various operating systems (Linux, Windows and UNIX etc.) without having to reboot to switch. Fundamentally, server virtualization technology takes a physical system (hardware, OS, and applications) and packages them into a portable, manageable virtual machine container. It can increase server utilization and reduce system's total cost of ownership, improve resources utilization, support High Availability (HA)and load balance, reduce operating system's dependence on hardware, save investment and maintenance costs.

Server virtualization is typically divided into two main types: full-virtualization—based on hardware level, and OS virtualization [7]. VMware is a representative of full-virtualization. It establishes a virtual platform between computer, storage and network hardware, so that all hardware can be unified into a virtual layer. The virtual layer provides bare-metal virtualization, which runs natively on the hardware. Virtual Machine (VM) provides a suit of virtual Intel x86-compatible hardware for OS

Lvaili(1980-),female, Han, Shouguang Shandong, master, lecturer, research area: Computer Network .

images running on the VM. The virtual hardware contains all the devices that a server should include (mainboard, CPU, memory, SCSI, IDE disk devices, interfaces and I/O devices).Moreover, each VM is encapsulated into a separate file, and it can be migrated flexibly.

OS virtualization lies on the top of host OS, which implements server virtualizations on the basis of host OS. That is to say, host OS allocates resources for virtual servers and keeps these servers independent each other. This approach can improve server consolidation rate and resource utilization greatly.

B. Storage Virtualization

Storage Virtualization [8] maps all the storage resources to a unit in logic; as a result, application servers only face a mapped storage volume, but not to mind the specific type of storage arrays. Meanwhile, users do not need to care about how storage arrays allocate spaces and process data, only to manage the virtual storage volume centrally.

This paper adopts storage virtualization technology based on networks. It segregates application servers and storage devices by way of installing software and hardware equipments — virtual server management platform and network middleware, in original SAN (Storage Area Network). And it administrates all the equipments fabricated by different firms and digital storage resources in a storage pool. In the pool, it builds one or more different sizes of virtual volume, and allocates the volume to application servers according to a certain read-write authorization.

C. VMware Virtualization Introduction

At present, one of mature data center virtualizations is VMware Infrastructure 3, which mainly includes three parts: ESX Server, Virtual Center (VC) and VMware Infrastructure Client.

VMware ESX Server builds on the hardware layer. It abstracts all resources, e.g. CPU, memory, internet and I/O devices, to multiple Virtual Machines. These Virtual Machines are independent relatively, they own their virtual resources—CPU, memory, and network card. Each VM runs various OS and applications on the basis of these virtual resources.

Virtual Center offers centralized management, automatic operation, resource optimization and high reliability for IT environment, which helps to improve the maintainability and high availability of IT environment.

VMware Infrastructure Client can manage a single physical server or connect to VC to manage servers and virtual machines. Operations VMware Infrastructure Client can carry out include: creating VM; powering up, powering off or rebooting VM; adjusting consumption ratio of CPU and memory resource; ghost; snapshots; migration and colony settlement; performance monitoring. And physical servers connected to utility storage can achieve cluster by means of Vmotion, HA and DRS (Distributed Resource Scheduler).

. VIRTUAL DIGITAL CAMPUS PLATFORM DESIGNATIONS

D. Virtual Digital Campus Platform Architecture

On account of current states of data center, this paper consolidates servers and storage resources with virtualization technology. It can improve data center's efficiency and processing capacity, make the system smooth and seamless upgrades. Figure 2 illustrates virtual digital campus architecture.

As shown in figure 2, virtual digital campus architecture is divided into three layers:

1) Virtualization hardware platform. The platform is the hardware basis of realizing virtual digital architecture, which includes ESX server, campus network and digital storage equipment.

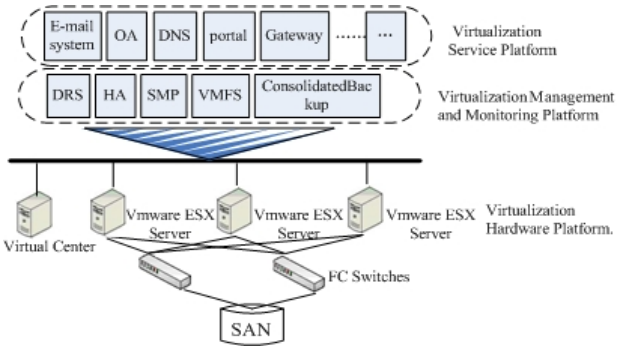


Figure 2. Virtual digital campus architecture.

As virtualization platform, the server assembling ESX need to run multiple virtual machines, so that its system burden is heavier and request for CPU and memory is higher. Its CPU adopts multi-path structure, selects and uses multi-core, and combining with Virtual SMP (Symmetric Multi-Processing) can improve working efficiency of VM effectively. ESX Server uses the method of preassignment system to manage memory. Its physical memory allocation depends on the number of VM and the size of memory preassigned to each VM. Meanwhile, ESX Server needs a certain memory to run its VM kernel and virtual service desk, therefore the size of physical memory should not be smaller than summation of memory all VM required add 1GB.

During the process of virtualization architecture design, memory planning is a key step. Many components of VMware Infrastructure, such as VMotion and HA, depend on share storage. In this paper, we link storage arrays through FC Switches and compose SAN, and dispose storage virtualization software to make all memory elements a virtual storage pool. Based on the virtual storage pool, the server virtualization platform can distribute memory space according to one's need.

2) Virtualization management and monitoring platform. The platform includes VMware ESX Server, VCMS(Virtual Center Management Server),VMware Infrastructure API(VI API)，which monitors, allocates and manages the physical hosts and virtual resources. Three physical servers install ESX Server to run VM. Files on the VM are stored in share memory by way of FC SAN, iSCSI SAN or NAS. ESX Server keeps communication with share memory continuously by

multipath. VCMS provides the basic allocation and management for physical hosts and virtual resources. They are software fundamental of virtual digital platform.

3) Virtualization service platform. The platform offers virtual application systems for users and implements various information application services for digital campus.

E. Function Information

Virtualization digital campus architecture is established on the basis of high performance servers, networks and SAN. In order to ensure application systems work safely and reliably, it means that any server and VM down will never lead application systems running on them to stop working, this paper predefines redundancy rules. The rules make certain that all application systems—ORACLE, DNS, file serves and WebSphere are not conflict with each other. And all applications can operate on any server at the same time. Redundancy rules are shown in table 1.

In actual runtime environment, each server has two redundancies, so that, system will never stop working in any extreme case. Once server is down, it can transfer in line with redundancy rules predefined in table 1.The transformation includes application systems (server and process), network resources (IP) and storage resources (volume). It makes processing unit (server node) and application system (WorkUnit) completely separate in logic, and realizes virtual applications and contents its actual functional requirement, enforces dynamic computing resource monitoring and management and dynamic optimization allocation.

TABLE I. SERVER REDUNDANCY RULES DEFINITION

server	First redundancy	Second redundancy	Third redundancy
DataServer	APPServer	FILEServer	WebServer
WebServer	APPServer	APPServer	DataServer
FILEServer	APPServer	DataServer	WebServer
APPServer	DataServer	WebServer	FILEServer

F. Implementation Steps and Details

1) Estimate

After estimating and analyzing the present environment, it confirms scope and target of virtualization, analyzes virtualization feasibility (if hardware meets the need of virtualization, if application can be virtualized, if network and storage need to make corresponding adjustment and reform).

2) Programming

Capacity planning: virtualization is carried out on existing servers or new servers. Storage planning: the document quantity and storage space of virtualization calling for, buying new storage devices or not. Network planning: the number of net card, network redundancy circuit and virtualization network policy.

Application planning: to grade application systems according to OS and service level, integrate existing tools of server management and monitoring.

Expansion planning: to think over CPU compatibility and plan for smooth transition.

3) Design

Virtualization overall structure design: cluster design based on HA and DRS, template design, template location mode, template management approach, template patch mode and template usage mode.

Storage partition: principle of storage partition, size of LUN (Logic Unit Number) and relationship between Virtual Machines and so on.

Network strategy: internet security policy, network bandwidth allocation and network load balancing policy.

Backup strategy: backup strategy of virtualization platform and VM, integration with third-party backup software.

Security policy: firewall settings and network isolation measures.

Migration policy: migration tools, migration methods and migration failed policy.

Consolidation policy: integration of existing management process, monitoring environment and automatic deployment tools.

4)Implementation

Installing VM components, disposing VLAN(Virtual Local Area Network), security policy and bandwidth limit, completing storage connectivity, creating partition, setting firewall and starting migration.

5) Operating and maintenance

Monitoring VM, managing virtualization architecture and proceeding optimization.

G. Stress-test Experiment

In order to verify practical running result, we design stress-test experiment based on one-card system. The experiment is divided into two groups: (1) basic configuration group, it runs multiple target systems on a single host; (2) virtual configuration group, it runs multiple virtual machine systems. Each virtual machine runs a target system. According to the processor type of VM system supported, VM is divided into two classes—single processor and symmetric multi-processor. We design an experiment, which takes pages return per second as the throughput evaluating parameters. Partial

TABLE II. EXPERIMENTAL DATA OF BASIC CONFIGURATION GROUP

4-way/CPU accounts	throughput	response time/s	threads/clients
1	231	69	4/18
2	229	70	4/18
3	230	71	4/20
4	228	78	4/21
5	220	76	4/22
6	215	82	4/23
7	211	81	4/24
8	212	85	4/25
1	268	116	4/32
2	309	102	4/32
3	360	85	4/32
4	353	88	4/32
5	342	90	4/32
6	341	91	4/32
7	337	92	4/32
8	335	93	4/32

TABLE III. EXPERIMENTAL DATA OF VIRTUAL CONFIGURATION GROUP

4-way/CPU accounts	throughput	response time/s	threads/clients
1	60	70	4/4
2	110	75	4/8
3	165	79	4/12
4	200	80	4/16
5	201	97	4/20
6	190	123	4/24
7	191	141	4/28
8	192	160	4/32
2	92	78	4/6
4	180	79	4/12
6	175	125	4/18
8	172	150	4/24

Experimental result indicates that, in the first experiment, four CPU and eight CPU servers reached the best operating efficiency when its target system account is three. The operating efficiency of virtual configuration group, running four virtual machine systems, is higher about 10 percent than basic configuration group. In the second group, single processor achieves the best operating efficiency when the VM account is four and eight, while, symmetric multi-processor achieves the best operating efficiency when the account is two and four. When the number of VM increase gradually, function of the second group decreases less than first group.

Based on above analysis, we can draw conclusions: (1) when VM account is equal to or less than CPU account, the operating efficiency is the highest; (2) application systems running on VM can content actual performance requirement; (3) key performance index of migrated VM system is higher than original physical system. In file system tests, VM CPU usage rate is 10 percent, which is higher about 7 percent than physical system.

CONCLUSIONS

Currently, datacenters have various problems: server resource utilization is low, which is about between 10% and 15% on average; system management is complex and maintenance management level is low; safety control and data backup is difficult. Using virtualization technology can resolve above problems effectively. Operation results of virtual digital campus platform practice that virtualization technology has advantages as follow:

- To maximize server resource utilization and reduce server amounts, server utilization ratio can come up to 70 percent or so.
- To improve resources allocation and increase resource utilization rate. Virtualization technology can carry out virtual partition on physical servers dynamically according to the resources different operation needed. And as a result, server utilization ratio is improved and resource allocation is more reasonable.
- VM exits the character of dynamic migration, so that it can be maintained and upgraded quickly. Not matter any server is down, VM can run continuously on other server. In consequence, stability and security of application systems are increased.

ACKNOWLEDGMENT

I would like to thank my colleagues on the virtualization technology team for their contributions, insights, and support.

This paper is supported by Youth Foundation of Henan Polytechnic University — Application of Virtualization Technology in Data Center of Campus Network.

REFERENCES

[1] Green Data Center: Virtualization management is key. e-Technology,pp:68,2009.

[2] Kohlbrenner E, Morris D, Morris B. Virtual Machines Core of Information Technology[Z]. 2005.

[3] XIAN Xiao-bing, SHEN Jun-yi, Research and Application of Virtual Digital Campus Architecture, Computer Engineering, pp:277-279, Vol.34,No.15,2008

[4] LIU Zhong-bao, JI Shen-hua, Application Prospects of Virtualization Technology in Commercial Bank, Financial Computer of China,pp:46-49,2008.

[5] DONG Yaozu, ZHOU Zhengwei. X86-based System Virtual Machine Development and Application [J]. Computer Engineering, pp : 71-73. Vol.32,No.13, 2006

[6] TANG Xiao-kang, Application of Server virtualization in Campus Network, Computer Era., pp:14-15,No.2, 2009

[7] CHEN Yu-ping, Digitized Campus Virtual Storage Technology, Technology and Innovation Management, pp: 656-658, Vol.30, No.5, 2009.

[8] Jones M T.Virtual Linux[EB/OL]. (2006-09-16) .http://www. 128.ibm.com/developerworks.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 283-288

Parameter Optimization of Multi-tank Model with Modified Dynamically Dimensioned Search Algorithm

Xiao-Lan Huang [1], Jun Xiong [2]
[1] Institute of Porous Media Mechanics,Wuhan Polytechnic University, Wuhan 430023, China
E-mail: hxl_huang0226@sina.com
[2] Social and Environmental Engineering Department,JP Business Service Corporation, Tokyo 135-8541, Japan
E-mail: xiongjun79@gmail.com

Abstract—**Multi-tank model is proposed for heavy rainfall infiltration simulation since it can represent the nonlinear transport behavior and give predictions very quickly. On the other hand, because its parameter space is high dimensional, it is difficult to obtain optimal parameters using existing methods. A new optimization approach called modified dynamically dimensioned search (MDDS) is developed for parameters calibration of this new model. It is based on heuristic global search and its adjustment is achieved by dynamically and randomly reducing the number of searching dimensions. Multi-tank model with 32 parameters is applied to an actual case. According to the results, better agreements between observations and calculation results are obtained compared with genetic algorithm. It is clarified that this model is a helpful tool in prediction of water table and stability factor of the slope.**

***Index Terms*—multi-tank model; parameter optimization; modified dynamically dimensioned search; groundwater table prediction; slope stability factor**

I. INTRODUCTION

It is noted that heavy rainfall may cause landslide and incur significant damage to the local area. In order to estimate the slope stability before collapse happens, the first step is to provide accurate water table forecasting. Sugawara and Takahashi [1,2] proposed a simple tank model to describe the long-term stream flow of specified drainage area. Because it can represent the non-linear stream flow behavior, it is widely used for long-term runoff analysis. Recently it is gradually developed to estimate the groundwater fluctuation of slope during rainstorm. In real cases complicate tank distributions are required, which results in high dimensional parameter space. Because of many parameters, it is difficult to properly identify those based on observed data. In previous literatures some researches tried to find solutions for tank model with 4 tanks (16 parameters) using many methods: Kobayashi and Maruyama[3] applied Powell's conjugate direction method to the problem. Watanabe[4] suggested Newton's method. Yasunaga[5] and Hino[6] tried sequential estimation using Kalman filter. Tanakamaru[7], Suzuki[8] tried to use genetic algorithm (GA) as an efficient search procedure. In this paper, multi- tank model is more complicate: four series of tanks are introduced, and 32 parameters have to be retrieved from observations by use of optimization functions. For such high dimensional parameter space, GA is very time-consuming and also the solutions are not so good. Therefore, in order to facilitate calibration process, the development of a new method is needed.

Basically, during the studies of estimating parameters, the calibration function of multi-tank model is replaced with a non-linear optimization function, for example, the most popular way is to minimize the errors between calculations and measurements. In this article, water table observations are adopted as criterion function. Using genetic algorithm to find the solution of tank model's parameters, if it is regarded as an inverse problem, it is an ill-posed problem without uniqueness of the solution especially when parameters are over 30. At the same time because too many parameters will lead to distribution order increase with geometric series, it also presents a significant computational burden. Actually if we consider obtaining results in the limited number of model evaluations, the idea of achieving global optimality will become unreasonable in most automatic calibration process. Therefore for high dimension optimization problems, a better multipoint random optimization method is necessary, and in this article a new approach called modified dynamically dimensioned search is provided as one of such good algorithm that focused on identifying good calibration results when calculation time is limited.

II. MULTI-TANK MODEL AND ITS PARAMETER DETERMINATION

Tank model is composed of one or several series of tanks with some outlets on the side and bottom in each tank. The basic concept is illustrated as Fig.1. From this figure it is obvious that rainwater in the tank will accumulate until water level becomes over than 'Z', then lateral flow happens. Outflow through the side outlets represents components of the total discharge due to the immediate or delayed response to the rainfall. Flow through the bottom holes means the portion of infiltrating flow and does not contribute to the surface flow directly.

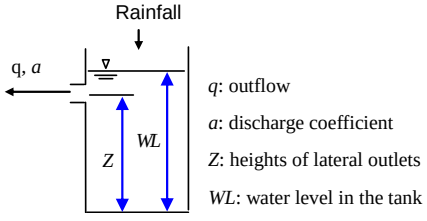


Figure 1. Basic concept of one tank

A. Schematic figure of multi-tank model in the slope

In this study, in order to monitor the water table of the slope and estimate the slope stability factor during rainfall, four series of multi-tanks are distributed as shown in Fig-2, which is designed to simulate rainfall infiltration process and runoff responses. One of these inter-connected tanks is set at the highest position to serve as the base point, which is followed by another two at the medium position and the last one at the lowest position. In this way, they can be designed to evaluate the major groundwater behavior of the slope.

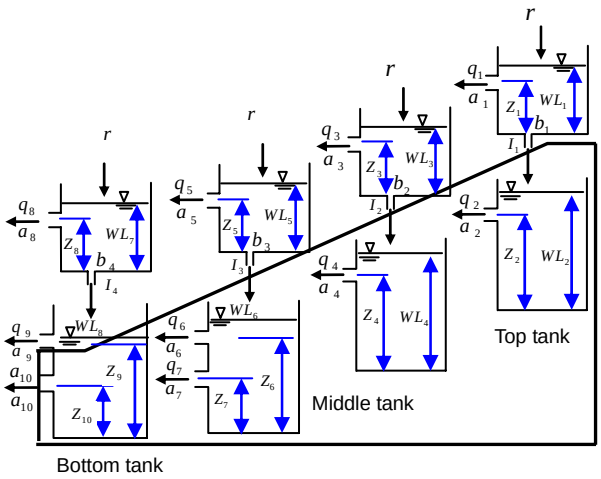


Figure 2. Parameter distribution of multi-tank model

The new model's flow patterns are shown as in Fig.2. According to water balance theory at the specific time, its discrete flow equation can be evaluated:

$$\begin{aligned}
WL_1(t) &= WL_1(t-1) + r(t) - q_1(t) - I_1(t) \\
WL_2(t) &= WL_2(t-1) + I_1(t) - q_2(t) \\
WL_3(t) &= WL_3(t-1) + r(t) + q_1(t) - q_3(t) - I_2(t) \\
WL_4(t) &= WL_4(t-1) + I_2(t) + q_3(t) - q_4(t) \\
WL_5(t) &= WL_5(t-1) + r(t) + q_3(t) - q_5(t) - I_3(t) \\
WL_6(t) &= WL_6(t-1) + I_3(t) + q_4(t) - q_6(t) - q_7(t) \\
WL_7(t) &= WL_7(t-1) + r(t) + q_5(t) - q_8(t) - I_4(t) \\
WL_8(t) &= WL_8(t-1) + I_4(t) + q_6(t) + q_7(t) - q_9(t) - q_{10}(t)
\end{aligned} \tag{1}$$

Where: r(t) is rain intensity (mm/day), a_i is the coefficients of runoff from the side hole of the tank; b_i means coefficients of seepage from the bottom hole; Z_i represents the height of the runoff on the side of tanks; and q_i is seepage runoff volume from the side of the tank; $WL_i(t)$ represents water level in i^{th} tank at time t;. Here the lateral flow discharge $q_i(t)$ and vertical seepage volume $I_i(t)$ at one specific time assumed to be proportional to its corresponding tank water level, which can be evaluated by equation (2):

$$\begin{aligned}
I_i &= b_i \cdot \mathrm{WL}_i^{top} \\
q_i &= a_i \cdot (WL_i - Z_i), q_i \geq 0, (i = 1 \sim 6) \\
q_7 &= a_7 \cdot (WL_6 - Z_7), q_7 \geq 0 \\
q_8 &= a_8 \cdot (WL_7 - Z_8), q_8 \geq 0 \\
q_9 &= a_9 \cdot (WL_8 - Z_9), q_9 \geq 0 \\
q_{10} &= a_{10} \cdot (WL_8 - Z_{10}), q_{10} \geq 0
\end{aligned} \tag{2}$$

Water levels in four lower tanks: WL_2, WL_4, WL_6 and WL_8 are considered to be related to groundwater table of the slope. After obtaining water levels of all the four lower tanks, then groundwater table GWT_i at a specific time t is calculated by the following equation:

$$GWT_i(t) = GWT_i(0) + WL_i^{bot}(t)/i \tag{3}$$

Where: v is effective porosity of the soil where tanks are set. GWTi(0) is the reference groundwater level at initial state, which must be small enough based on available observations. Generally the reference groundwater levels in four lower tanks are defined as smaller than the lowest water table in the history.

B. Definition of optimization function

In order to find appropriate solution, optimization function is necessary. Here this evaluation functions is to minimize the errors between calculations and measurements, which is defined as the following equation.

$$J_{XS} = \frac{1}{M}\sum_{i=1}^{M}\frac{(Q_c(i) - Q_o(i))^2}{Q_c(i)} \tag{4}$$

Where: $Q_o(i)$ is water table observation of four lower tanks; $Q_c(i)$ means calculation results; M is the number of observed data. Generally there are 32 parameters that needed to be estimated by use of the developed new optimization method.

C. Definition of modified dynamically dimensioned search

The dynamically dimensioned search algorithm[9] (DDS) is a novel and simple stochastic single-solution method, and it is based on heuristic global search algorithm that was developed for the purpose of finding good global solutions within the specified maximum function evaluation limit. In short, the algorithm searches globally at the start of the search and becomes more and more local as the number of iterations approaches the maximum allowable number of function evaluations. The adjustment from global to local search is achieved by dynamically and randomly reducing the number of dimensions in the neighborhood. The decision variables in automatic calibration are the model parameters, and the dimension being varied is the number of model parameter, which is changed to generate a new search neighborhood. Candidate solutions are created by perturbing the current solution values randomly selected dimensions only. Its perturbations magnitudes are sampled from a normal distribution N(0,1).

The DDS algorithm is unique relative to current other random optimization approaches because of the way that neighborhood is dynamically adjusted by changing the dimension of the search. For example if GA is adopted, the final solutions will be different for every calculation process. The only algorithm parameter to set in DDS is the scalar neighborhood size perturbation parameter (r) that defines the random perturbation size as a fraction of the decision variable range. An initial value of the r parameter is set as 0.3. Because multi-tank model's parameter space dimension is over 30, in order to improve searching efficiency, with the calculation process going on, r will reduce step by step, the minimal value

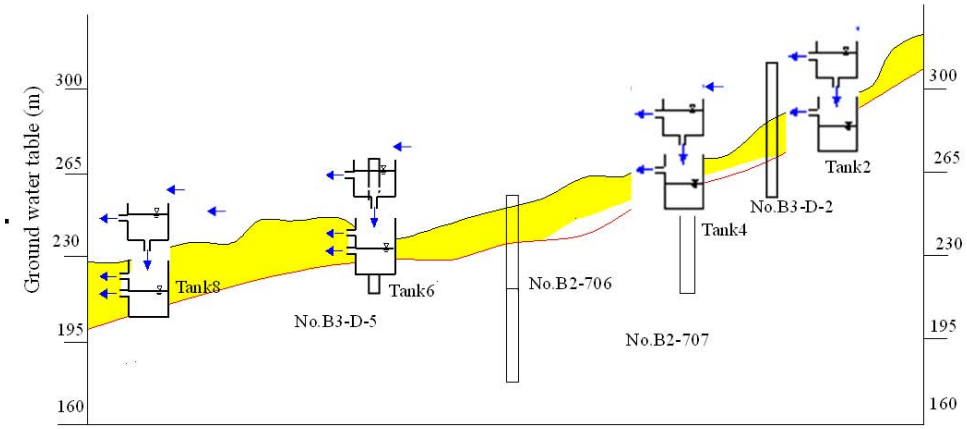


Figure 3. Configuration of multi-tank model in the slope

of r is 0.05, which is different from Tolson[9]. This initial sampling region size is designed to allow the algorithm to escape regions around poor local minima. In the final stage because the current solution is close to final results, in order to avoid big perturbation, the value of r must decrease. And also in order to accelerate the speed of convergence, its update algorithm (STEP5) is also modified: if the change in objective function value is positive, the new solution is accepted with a certain probability; whereas in previous literatures it is abandoned simply. The calculation process of modified dynamically dimensional search (MDDS) algorithm is provided as the follows:

STEP1. Define inputs:

- Neighborhood perturbation size parameter: r(the default initial value is 0.3);
- Vectors of lower, $\mathbf{x}^{\min}$, and upper, $\mathbf{x}^{\max}$, and initial solution, $\mathbf{x}^0 = [x_1, x_2, \ldots, x_m]$.

STEP2. Set counter to 1, i=1, and evaluate objective function **F** at initial solution, $F(\mathbf{x}^0)$:

- $F_{best} = F(\mathbf{x}^0)$, and $\mathbf{x}^{best} = \mathbf{x}^0$

STEP3. Randomly select J of the *m* parameters for inclusion in neighborhood, {N}.

STEP4. For j=1,…,J decision variables in {N}, perturb x_j^{best} using a standard normal random variable: N(0,1), reflecting at decision variable bounds if necessary:

STEP5. Evaluate $F(x^{new})$ and update current best solution if necessary:

- If $F(x^{new}) \leqq F^{best}$, update new best solution:
  $F^{best} = F(x^{new})$ and $x^{best} = x^{new}$
- If $F(x^{new}) > F^{best}$ and $\exp(-(F^{new} - F^{best})/f(j)) > \text{random}(Pn)$
  $F^{best} = F(x^{new})$ and $x^{best} = x^{new}$

STEP6. Update iteration count, i=i+1, and check stopping criterion:

- If i= Maxiter, STOP, print output (e.g: F^{best} and x^{best})
- Else go to STEP3

The only parameter 'r' is defined as the following lines: P_n decreases with the increase of the number of function evaluations (NF is maximum number of function evaluation; 'i' is the current calculation step):

```
Pn=1.0-dlog(dfloat(i))/dlog(dfloat(NF)
if(0.3<Pn.)                 r_val=0.30
if (0.2<Pn.and.Pn<0.3)      r_val=Pn
if (0.1<Pn.and.Pn<0.2)      r_val=Pn
if (0.05<Pn.and.Pn<0.1)     r_val=Pn
if (Pn<0.05)  Pn=0.05;      r_val=0.05
```

III. CASE STUDIES ON THE ACTUAL SLOPE USING MULTI-TANK MODEL

Based on previously suggested procedures, multi-tank model is applied to the slope along Japanese national road No.12 in Yamagata Prefecture to simulate fluctuations of groundwater table induced by during rainfall period.

A. *Outline of the slope*

From the boring survey results, it is revealed that in the slope, weathered rock is about 3 to 10 meter thick. With the history of collapses, it was regarded that it is urgent to determine its water table fluctuations and evaluate its stability. As illustrated in Fig-3, it is the configuration of multi-tank model in the slope. Top tank (tank2) is assumed on the top hill(x=325m); bottom tank (tank8) lies on the lowest part of the slope (x=7m). It is found that the surface layer is almost homogeneous, therefore middle tank distribution is relatively easy: totally there are two series. Because a mound exists near the foot of slope, it has big influence on rainfall infiltration, Tank6 is set at the boundary of the mound (x=98m). Tank4 is set at Boring No.B2-707 (x=225m) to serve as slope lateral flow simulation and help to provide more accurate water table forecasting. The porosities near the four parts are 0.09, 0.12, 0.16 and 0.15 respectively.

As aforementioned, tank model's parameters are difficult to be directly measured by experiments, therefore, in order to evaluate model parameters, historical data of rainfall and ground water table are required. In this case, there are four observation borings along this cut-slope: boring No.B3-D-2(x=310m), B2-707 (x=225m), B2-706(x=165m), and B3-D-5(x=98m) are drilled to monitor the ground water table at the four locations. Rainfall intensities and ground water tables of 100 days have been recorded for parameters optimization.

B. *Analytical conditions*

The parameters of multi-tank model can be identified by the reproduction of observed hydrographs assuming that basic watershed characteristics remain unchanged during the observation of events. In the runoff analysis, because the studied slope area is fairly small in size, the travel time is considered to be relatively short. The data used for the analysis are the 100 daily measurements of precipitation (shown as Fig. 4 between 1994-5-27 and 1994-9-4) and water tables averaged

over catchment of the slope. By using multi-tank model, water tables can be obtained; and using optimization method of modified dynamically dimensioned search algorithm, parameters are retrieved. Here there are totally 32 parameters; their lower and upper bounds of the search for parameters are listed in Table1. The bounds of search are set based on the result of an application of the three-series tank model. The maximum number of function evaluation (Maxiter) is 4000, and it is considered large enough for practical purposes.

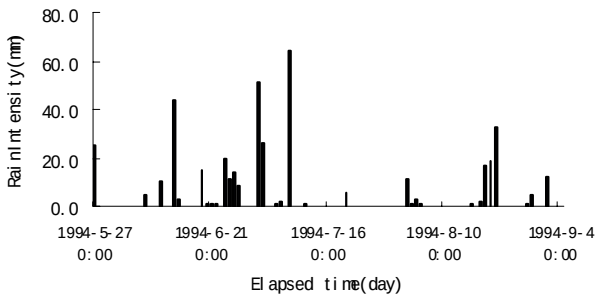


Figure 4. Daily rainfall intensities

TABLE1. PARAMETER BOUNDS

0.0<a(1)<0.6	0.005<Z(1)< 0.4	0.0< WL0 (1)<0.5
0.0<a(2)<0.6	0.005<Z(2)< 0.4	0.0< WL0 (2)<0.5
0.0<a(3)<0.6	0.005<Z(3)< 0.4	0.0< WL0 (3)<0.5
0.0<a(4)<0.6	0.005<Z(4)< 0.4	0.0< WL0 (4)<0.5
0.0<a(5)<0.6	0.005<Z(5)< 0.4	0.0< WL0 (5)<0.5
0.0<a(6)<0.6	0.005<Z(6)< 0.4	0.0< WL0 (6)<0.5
0.0<a(7)<0.6	0.005<Z(7)< 0.4	
0.0<a(8)<0.6	0.005<Z(8)< 0.4	0.0< WL0 (7)<0.5
0.0<a(9)<0.6	0.005<Z(9)< 0.4	0.0< WL0 (8)<0.5
0.0<a(10)<0.6	0.005<Z(10)< 0.4	
0.0<b(1)<0.6	Unit/ m	Unit/ m
0.0<b(2)<0.6		
0.0<b(3)<0.6		
0.0<b(4)<0.6		

C. *Analytical results*

The analysis results of by dynamically dimensioned search are shown in Table 2. It is found that lateral coefficients of upper tanks become bigger and bigger from top to bottom (a(1)= 0.126, a(3)= 0.360, a(5)= 0.318, a(8)= 0.386); the exception is a(5), which lies in the existence of the mound. At the same time, corresponding height of lateral outflow hole become smaller and smaller (Z(1)= 0.397, Z(3)= 0.064, Z(5)= 0.016, Z(8)= 0.011), which means the top part is dry, and it needs more infiltrating water until surface runoff happens. On other hand, at the bottom, surface flow comes into being easily (Z(1) is much bigger than Z(8)). All these phenomenon are consistent with reality. The initial water levels in four upper tanks also have the same order: water level in lower positions has bigger values. From this table, it is demonstrated that multi-tank model can still provide good forecasting of surface stream flow and modified dynamically dimensioned search is very good method that calibrate reasonable parameters.

The final calculation error J_{XS} is represented by 0.05812. Objective function values plotted against the number of function evaluation is shown in Fig 5. From the figures, it is clear that at the beginning objective function values are over 50, but with the calculation process going on, they go down quickly. At the same time, because of its stochastic nature of modified dynamically dimensioned search, the perturbation is very big at the initial stage, which means the algorithm can escape from regions around poor local minima. Eventually, because the current solution is close to final results, in order to avoid big perturbation, perturbation amplitude becomes smaller and smaller gradually, the current solution is close to the final results.

TABLE2 OPTIMIZATION SOLUTIONS BY MDDS

a(1)=0.126	Z(1)=0.397	WL0 (1)=0.012
a(2)=0.107	Z(2)=0.033	WL0 (2)=0.173
a(3)=0.36	Z(3)=0.064	WL0 (3)=0.013
a(4)=0.113	Z(4)=0.080	WL0 (4)=0.207
a(5)=0.318	Z(5)=0.016	WL0 (5)=0.015
a(6)=0.122	Z(6)=0.213	WL0 (6)<0.069
a(7)=0.119	Z(7)= 0.011	
a(8)=0.386	Z(8)=0.011	WL0 (7)=0.015
a(9)=0.099	Z(9)=0.302	WL0 (8)=0.017
a(10)=0.092	Z(10)= 0.011	
b(1)=0.05	Unit/ m	Unit/ m
b(2)=0.05		
b(3)=0.478		
b(4)=0.468		

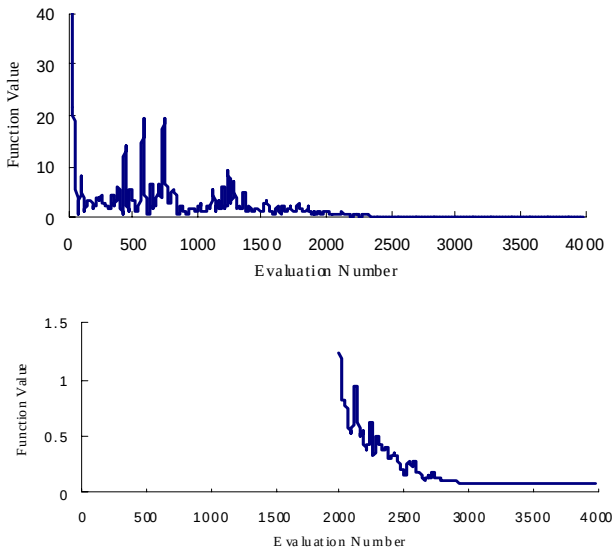


Figure 5. Objective function value against number of function evaluations

Differences of water levels at tank2 and 8 between observations and calculation results are shown as Fig.6 and Fig.7 in the 100 days (from May 27th, 1994 to Sept. 4th, 1994), from which it can be concluded that in the two sites, water levels of observations and calculation results are similar, especially for tank8, they are almost the same. From the two figures, it is illustrated that the default settings of the neighborhood perturbation test functions (*r*) produced good results across the multi-tank model calibration problem. Therefore, the strategy for *r* seems reasonable and suggests its validity for most future application.

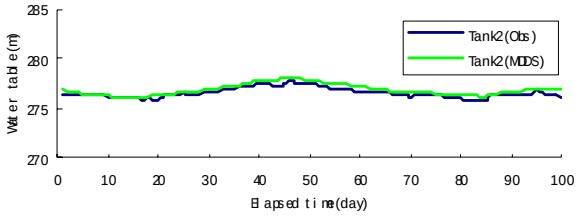


Figure 6. Comparison between optimal results and observations at tank2

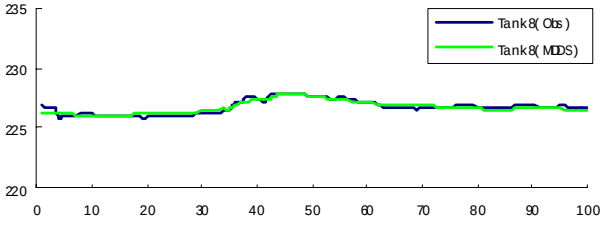


Figure 7. Comparison between optimal results and observations at tank8

In the aforementioned parameters calibration procedures of tank model, Tanakamaru (1993), Suzuki(1999) and others has tried to use genetic algorithm (GA) as an efficient search procedure, where totally there are 16 parameters. Here, in order to compare the effectiveness of MDDS and GA for parameters estimation of multi-tank model, GA will also be tried for parameter estimation of multi-tank model. For multi-tank model, totally there are 32 parameters, in order to improve the quality of the results, multiple populations has been employed, which is known as the migration, or island model. Each subpopulation is evolved over generations by a traditional GA, within each subpopulation it is necessary to perform functions such as selection, crossover and reinsertion. And from time to time individuals migrate from one subpopulation to another.

Comparisons of water levels between observations and calculation results (both MDDS and GA) at two positions are illustrated in Fig.8 and Fig.9 during the 100 days (from May 27th, 1994 to Sept. 4th, 1994), from which it can be concluded that in the two sites, results of MDDS is very good; while water tables obtained from GA is bad: obviously it is trapped in the local optimal solution.

Reproductions of water levels at four locations are shown from Fig.6 to Fig.9), from which it can be concluded that in all four sites, water levels of observations and calculation results from our new strategy are similar, especially for tank6 and tank8, they are almost the same.

After getting the water tables of the several locations during heavy rainfall, the instantaneous groundwater lines can be estimated quickly by spline interpolation method. Then the slope stability factor is calculated with the division methods commonly. In the division method, there are many methods such as Fellenius, Bishop, Janbu and Spencer. Fig.10 is the results of stability analysis of this slope using Bishop method during the analysis period: it is clear that when rain is big, its stability factor decreases sharply, which also demonstrates that rainfall has great influence on slope stability.

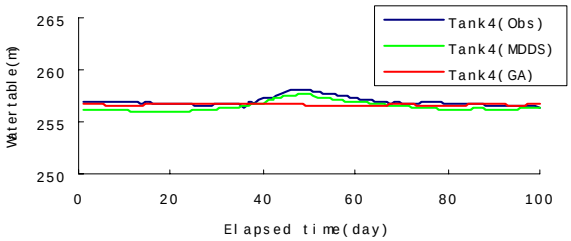


Fig.8 Comparison between MDDS and GA at tank4

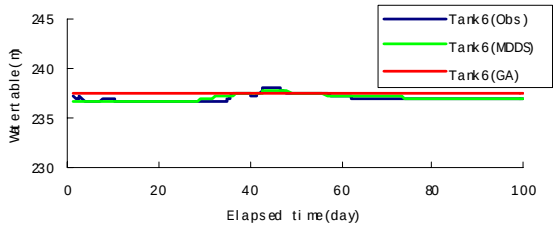


Figure 9. Comparison between MDDS and GA at tank6

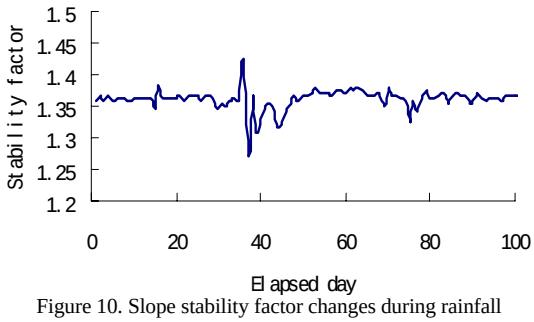


Figure 10. Slope stability factor changes during rainfall

IV. CONCLUSIONS

The paper focused on the behavior of rainfall infiltration process and aimed to develop a simple and quick analytical tool to evaluate underground water table and slope stability factor during heavy rainfall. The insights gained through this study are summarized as the following:

According to water balance (tracking flows of water into and out of the particular hydrologic system of interest), a multi-connected tank model that can reproduce the rainwater movement behavior was developed. A new stochastic single-solution method called modified dynamically dimensioned search was adopted to identify optimal solutions. This method is robust for parameter calibration with high space, since it generated relatively good solutions without requiring any algorithm parameter adjustments. Compared with genetic algorithm, the new method could find relatively good solutions in a shorter time. Multi-tank model was applied to the actual slope. Its consistency with field data was confirmed, and its practicability was proved. Meanwhile, its algorithm is very simple and thus it can be easily coded in any programming language. Although this study focused on multi-tank model, the results are just as relevant to all

environmental simulation modelers who is calibrating parameters of a computationally demanding model.

Although multi-tank model has the advantage of predicting water table fluctuations quickly, it can not give information of infiltration process in the unsaturated zone. The authors are planning to work on the development of unsaturated tank model; therefore the stability for shallow landslide can be assessed during by rainfall. Combined with a new accurate rain gauge, the methodology will be evolved further into an assessment system for correctly predicting the hazards of rainfall that may lead to slope failure. In addition, it is necessary to test the modification and improve the ability of MDDS to locate the exact global optimum or implement a parallelized version of this algorithm.

REFERENCES

[1] M. Sugawara, E. Ozaki, I. Watanabe, etc, "Method of automatic calibration of tank model(second report)," Research notes of the National Research Center for Disaster Prevention, Japan, Vo17,43-89,1978.(in Japanese)

[2] K. Takahashi, Y. Ohnish, J. XIONG and T. Koyama, "Tank Model and its Application to Groundwater Table Prediction of Slope," Chinese Journal of Rock Mechanics and Engineering, Vol.27(12): 2501-2508, 2008.

[3] S. Kobayshi and T. Maruyama, "Search for the coefficients of the reservoir model with the Power's conjugate direction method," Trans. JSIDRE, No.65, pp.42-47, 1976.

[4] K. Watanabe, "Refinements to parameter optimization in the tank model, Proceedings of the 33rd Japanese Conference on Hydraulic," pp.55-60, 1989. (in Japanese)

[5] T. Yasunaga, K. Jinno and A. Kawamura, "Change in the runoff process into an irrigation pond due to land alteration," Proceedings of the 36th Japanese Conference on Hydraulics, Japan, 1992: 629-634(in Japanese).

[6] M. Hino, "Prediction of hydrologic system by Kalman Filter," Proc. of Japan Society of Civil Engineers, 1974, No.221: 39-47. (in Japanese)

[7] H. Tanakamaru, "Parameter identification of tank model with the genetic algorithm," Annuals of Disaster Prevention Research Institute, No.36B-2, 1993:231-239 (in Japanese).

[8] K. Suzuki, H. Momota, H. Takahashi, "Statistical study of tank model identification by genetic algorithm," Japanese Journal of Hydroscience and Hydraulic Engineering, 1999, VOL.17, No.1 May, 11-19.

[9] B.A.Tolson, C.A. Shemaker, "Dynamically dimensioned search algorithm for computationally efficient watershed model calibration,"Water Resources Research, VOL.43, W01413, 2007.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 289-293

Research of High Performance Computing With Clouds

Ye Xiaotao[1], Lv Aili[1], Zhao Lin[2]
[1]Modern Education Technology Center,Henan Polytechnic University, Jiaozuo, China
[2]Neusoft Institute of Information, Dalian, China

***Abstract*—HPC is most commonly associated with computing used for scientific research nowadays, which always uses supercomputers and computer clusters. Cloud computing – a relatively recent, builds on decades of research in virtualization, distributed computing, utility computing and more recently networking, web and software service. Cloud computing includes 3 services: SaaS, PaaS and IaaS. The popular general cloud service like EC2 allow users to provision compute clusters fairly and quickly by paying a monetary value only for the duration of the resources. Recently, HPC give rises of the cloud computing for cheaper economic solutions and more enterprises announced their HPC on-demand service. In this paper, some HPC applications (mostly recently) that have been deployed with clouds are also summarized. The possibility of using cloud computing for HPC is illustrated by experiments and more and more application types are well-suited to use cloud.**

***Index Terms* -HPC, Cloud Computing, EC2, SaaS**

I. INTRODUCTION

High-performance computing (HPC) is the use of parallel processing for running advanced application programs efficiently, reliably and quickly. HPC uses supercomputers and computer clusters to solve advanced computation problems. The application and the data both need to be moved to the available computational resource in order for them to be executed [1]. These infrastructures are highly efficient in performing compute intensive data movement. Today, computer systems approaching the teraflops-region are counted as HPC-computers.

Cloud computing is the latest and perhaps the most dramatic trend in advanced computing paradigms since the introduction of commodity clusters, which have dominated HPC for more than a decade. Clouds offer an amorphous distributed environment of computing resources and services to a dynamic distributed user base. Like clusters, cloud computing exploits economies of scale to deliver advanced capabilities. Unlike clusters, cloud resources are nonspecific and provide basic capabilities but guarantee neither identical properties from run to run nor high availability of specialized system types.

At present, the use of cloud computing in computation science is still limited, but the first step towards this goal have been already done. Last year, the Department of Energy (DOE) National Laboratories started exploring the use of cloud services for scientific computing. On April 2009, Yahoo Inc. announced that it has extended its partnership with the major top universities in United States of America to advance cloud computing research and applications to computational science and engineering.

II. CLOUD COMPUTING

A. Cloud Definition

Cloud computing is Internet-based computing, whereby shared resources, software and information are provided to computers and other devices on-demand, like a public utility. A technical definition [2] is "a computing capability that provides an abstraction between the computing resource and its underlying technical architecture (e.g., servers, storage, networks), enabling convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction." This definition states that clouds have five essential characteristics: on-demand self-service, broad network access, resource pooling, rapid elasticity, and measured service.

B. Cloud Technologies

The cloud technologies such as MapReduce and Dryad have created new trends in parallel programming [3]. The support for handing large data sets, the concept of moving computation to data, and the better quality of services provided by the cloud technologies make them favorable choice of technologies to solve large scale data/compute intensive problems.

Cloud technologies such as Google MapReduce, Google File System (GFS), Hadoop and Hadoop Distrubuted File System (HDFS), Microsoft Dryad, and CGL-MapReduce adopt a more data-centered approach to parallel runtimes[4][5]. In these frameworks, the data is staged in data/compute nodes of clusters or large-scale data centers, such as in the case of Google. The computations move to the data in order to perform the data processing. Distributing file systems such as GFS and HDFS allow Google MapReduce and Hadoop to access data via distributed storage systems built on heterogeneous compute nodes, while Dryad and CGL-MapReduce support reading data from local disks. The simplicity in the programming model enables better support for quality of services such as fault tolerance and monitoring.

Ye Xiaotao (1980-), male, Han, Qinyang Henan, master, lecturer, research area: High Performance Computing

Supported by The Henan project of Higher Education informationization. project number: 506062

Table Ⅰ summarizes the different characteristics of Hadoop, Dryad, CGL-MapReduce, and MPI.

C. Cloud Computing Services Offering

Cloud computing is typically divided into three levels of service offerings: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Serivice (SaaS) [6]. Figure 1 provides such categorization.

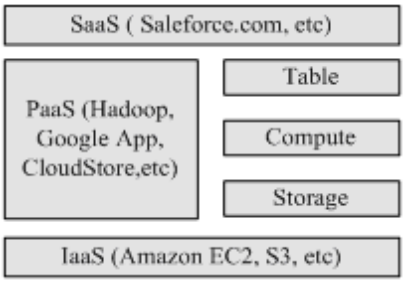


Figure 1. Cloud computing offerings by services.

Infrastructure as a Service - Traditional computing resources such as servers, storage, and other forms of low level network and hardware resources offered in a virtual, on demand fashion over the Internet. IaaS in a general sense, provides the ability to 'summon' resources in specific configurations at will and delivers value similar to what one might find in a traditional datacenter. IaaS' power lies in its massive on-the-fly flexibility and configurability. It can be equated to owning a magic wand that could conjure up a variety of network and server resources in zero time and occupying zero space. Examples include services like GoGrid, Amazon's EC2 [7] and even S3 [8] (as a storage infrastructure play)

Platform as a service implementation provides users with an application framework and a set of API that can be used by developers to program or compose applications for the Cloud. In some cases, PaaS solutions are generally delivered as an integrated system offering both a development platform and an IT infrastructure on top of which applications will be executed. The two major players adopting this strategy are Google and Microsoft.

Software as a Service - Specialized software functionality delivered over the Internet to users who intend to use the set of delivered functionality to augment or replace real world processes. Generally speaking, users within the SaaS space are aggregated into 'tenants', or bodies of 1 or more categorically related users. Think Salesforce.com CRM, or SugarCRM.

Table Ⅱ gives a feature comparison of some of the most representative players in delivering IaaS/PaaS solution for cloud computing.

III. HIGH PERFORMANCE COMPUTING WITH CLOUDS

Cloud computing presents a unique opportunity for batch processing and analytics jobs that analyze terabytes of data and can take hours to finish. If there is enough data parallelism in the application, users can take advantage of the cloud's new "cost associativity": using hundreds of computers for a short time costs the same as using a few computer for a long time. Programming abstractions such as Google's MapReduce and its open-source counterpart Hadoop allow programmers to express such tasks while hiding the operational complexity of choreographing parallel execution across hundreds of cloud computing servers. Some works with MapReduce has already been done and tested over the clouds. Again, the cost/benefit analysis must weigh the cost of moving large datasets into the cloud against the benefit of potential speedup in the data analysis. When we return to economic models later, we speculate that part of Amazon's motivation to host large public datasets for free may be to mitigate the cost side of this analysis and thereby attract users to purchase cloud computing cycles near this data.

TABLE II CLOUD COMPUTING SOLUTION FEATURE COMPARISON

Properties	Amazon EC2	Google AppEngine	Microsoft Azure
Service Type	IaaS	IaaS-Paas	IaaS-PaaS
Support for(value offer)	compute/ storage	compute(web application)	compute/ storage
Value Added Provider	Yes	Yes	Yes
User access Interface	Web APIs and Command Line Tools	Web APIs and Command Line Tools	Azure Web Portal
Virtualization	OS on Xen Hyperiview	Application Container	Service Comtainer
Platform(OS & runtime)	Linux, Windows	Linux	.NET on Windows
Deployment Model If PaaS, ability to deploy on 3rd party IaaS	N.A.	No	No

Some commercial HPC applications that have been deployed with clouds have been described by focusing the nature of the application and the commercial benefits of the deployment with the clouds. For example, the Server Labs, Pathwork Diagnostics, Cycle computing and Atbrox and Lingit.

Nonetheless, the cloud computing model, in spite of its promise, either imposes constraints in conflict with some HPC requirements or simply fails to adequately support them [9]. Among these constraints is the underlying hardware architecture virtualization, which is valuable for generic usage of diverse cloud resources. Such resources generally provide portability but obstruct targeting algorithm optimizations to specific hardware structures, as is typical of HPC applications. The time-critical overhead that virtualization layers add further degrades the performance efficiency and scalability of some HPC workloads. Another performance issue related to clouds is that users share resources among multiple tasks for both computational and networking functionality. The resulting resource contention inserts sporadic and unpredictable delays, further degrading performance and making optimizations more difficult.

Networking is critical to HPC facility operations. The availability of network infrastructure enables-and potentially limits-collaboration among geographically

distributed groups. This is also true for computing systems that support execution of distributed tasks. Because the Internet is a key component of the cloud computing model, this new computing regime will exacerbate any pre-existing limitations in the network infrastructure. The ability to manage costs and acceptable application performance response times will determine operational effectiveness. The network will determine the distance between the data and the computation, which means that in the cloud model, if the bandwidth is low, the user must procure additional data storage near the computation. This increased reliance on data communication will likely be the first deciding criterion for whether an organization will adopt cloud computing.

External I/O can become a serious bottleneck to application performance if not balanced with application needs, buffering, and contention for these resources with other concurrent demands. Checkpoint and restart requirements for purpose of long-term reliability can impose further demands on I/O bandwidth, which, if not available, might seriously degrade overall delivered performance. Thus, I/O could further reduce the value of cloud computing to HPC users.

Beyond performance are the critical issues of security and reliability. Much data is highly sensitive, such as intellectual property, competitive planning information, or highly classified intelligence from mission-critical agencies with strong national security responsibilities. In these cases, users won't trust remote networking, storage, and processing resources, no matter how well-intentioned they assume the encryption and other implemented measures to be. Therefore, such organizations are unlikely to employ clouds for these purposes, which comprise a significant portion of HPC activity. Similarly, clouds might not provide sufficient reliability to adequately minimize risk-a particularly sensitive issue in time-bounded applications. Again, dedicated systems are more likely the preferred platform in these cases.

Edward Walker, a research scientist with the Texas Advanced Computing Center at the University of Texas at Austin, has done performance analysis of Amazon EC2 for high performance scientific applications. His results show a significant performance gap in the examined clusters that system builders, computational scientists, and commercial cloud computing vendors need to be aware of.

IV. EXPERIMENTS AND EVALUATION

HPC as a Service [10] is a computing model where users have on-demand access to including the expertise needed to set up, optimize and run their applications over the Internet. The traditional barriers associated with high-performance computing such as the initial capital outlay, time to procure the system, effort to optimize the software environment, engineering their system for peak demand and continuing operating costs have been removed. Instead, HPC as a Service user has a scalable cluster available on demand that operates and has the same performance characteristics as a physical HPC cluster located in their data room. There are different definitions of cloud computing, but at the core "Cloud computing is a style of computing in which dynamically scalable and often virtualized resources are provided as a service over the Internet." HPC as a Service extends this model by making concentrated, non-virtualized high-performance computing resources available in the cloud.

A. Benefits

HPC as a Service provides users with a number of key benefits as follows.

- HPC resources scale with demand and are available with no capital outlay-only the resources used are actually paid for.
- Experts in high-performance computing help setup and optimize the software environment and can help trouble-shoot issues that might occur.
- Faster time-to-results especially for computational requirements that greatly exceed the existing computing capacity.
- Accounts are provided on an individual user basis, and users are billed for the time they use service.
- A HPC platform for you and your applications: Support for ANSYS, OpenFOAM, LSTC, etc ... and third party support.
- Access from anywhere in the worlds with high-speed data transfer in and out.

B. HPC On-demand Service

More enterprises announced to offer a computing on demand solution aimed specifically at the HPC market, e.g. Penguin Computing's Penguin on Demand (POD), newservers' Bare Metal Cloud, Gompute, SGI's Cyclone and all kinds of middleware like Platform ISF.

Linux cluster maker Penguin Computing hopped on the HPC-in-a-cloud bandwagon with the announcement of its HPC on-demand service August 2009. POD provides a computing infrastructure of highly optimized Linux clusters with specialized hardware interconnects and software configurations tuned specifically for HPC. Rather than utilizing machine virtualization, as is typical in traditional cloud computing, POD allows users to access a server’s full resources at one time for maximum performance and I/O for massive HPC workloads.

Comprising high-density Xeon-based compute nodes coupled with high- speed storage, POD provides a persistent compute environment that runs on a head node and executes directly on the compute nodes’ physical cores. Both GigE and DDR high-performance Infiniband network fabrics are available. POD customers also get access to state-of-the-art GPU supercomputing with NVIDIA Tesla processor technology. Jobs typically run over a localized network topology to maximize inter-process communication, to maximize bandwidth and minimize latency.

Penguin has also been working with a new biomedical startup to understand the performance characteristics of their application on the POD system. Results on an 8 nodes configuration (using Amazon's High-CPU instance) show a runtime of 31.2 minutes on the POD and 18.5 hours on EC2 and shown in Table Ⅲ, putting the

POD about 32x faster than EC2 for this particular application. An infrastructure comparison between POD and EC2 in this test is given in the Table Ⅲ too.

TABLE III INFRASTRUCTURE COMPARISON AND PERFORMANCE FOR APPLICATION FROM, POD VS. EC2

	POD	EC2
Network	1 GbE and DDR InfiniBand	Shared Memory, 300-400MB/s X-transfer rate
Computing Unit	Xeon 5400	1.0-1.2 GHz, 2007 Opteron or 2007 Xeon processor
OS	Linux	Linux, Open Solaris, Windows Server and others
Run Time	31.2min	18.5hours
Latency	47ms	185ms
Throughput	20MB/s	5MB/s

Another example, Gompute provides on demand HPC for technical and scientific computing. Gompute's services allow users to exploit HPC resources over the Internet by paying for what they actually use. Gompute also provides its users with high quality training for the applications supported at Gompute's on demand service. Consultants and independent software vendors can sell their services and software licenses using Gompute.

HPC in the Cloud is a mixed bag. Unless you use a specially designed HPC cloud the I/O resources critical to HPC performance can be quite variable. This may be changing, however, as individual servers contain more cores. Recently IDC has reported that 57% of all HPC applications/users surveyed use 32 processors (cores) or less. When the clouds start forming around 48-core servers using the imminent Magny Cours processor from AMD, many applications may fit on one server and thus eliminate the variability of server-to-server communication. HPC may start to take a very different form as dense multi-core servers enter the cloud. A user may sit at her desk submitting jobs to their own SGE desktop. The resource scheduler will then reach out to local resources or Cloud resources that can run virtualized or bare metal versions of her applications.

V. CONCLUSION AND FUTURE WORK

Cloud computing's potential for the particularly challenging domain of HPC is promising. In fact, many application types in the overall HPC workflow are well-suited to the near-term exploitation of cloud services. Furthermore, institutions that take advantage of clouds might benefit substantially in operational and cost–effectiveness as well as in flexibility and responsiveness to internal workload demands. But don't assume that clouds will easily replace the HPC systems that organizations currently deploy to provide the most extremes in capability; rather, the two world views must coexist, seeking benefits from clouds while achieving HPC's mission–critical requirements.

However, many anticipated properties of distributed cloud environments strongly suggest that clouds can only partly address HPC user needs and that some workload subdomains will remain beyond the capabilities of cloud services. Virtualization, uncertainty of hardware structural details, lack of network control and memory

TABLE I COMPARISON OF FEATURES SUPPORTED BY DIFFERENT PARALLEL PROGRAMMING RUNTIMES.

Feature	Hadoop	Dryad	CGL-MapReduce	MPI
Programming Model	MapReduce	DAG based exectution flows	MapReduce with a Combine phase	Variety of topologies constructed using the rich set of parallel constructs
Data Handing	HDFS	Shared directories/local disks	Shared directories/local disks	Shared directories
Intermediate Data Communication	HDFS/Point-to-point via HTTP	Files/TCP pipes/Shared memory FIFO	Content Distribution Network (NaradaBrokering (Pallickara and Fox 2003))	Low latency communication channels
Scheduling	Data locality/Rack aware	Data locality/Network topology based run time graph optimizations	Data locality	Available processing capabilities
Failure Handing	Persistence via HDFS Re-execution of map and reduce tasks	Re-execution of vertices	Currently not implemented (Re-executing map tasks, redundant reduce tasks)	Program level Check pointing OpenMPI(Gabriel, E.,G.E.Fagg, etal.2004),FT MPI
Monitoring	Monitoring support of HDFS, Monitoring MapReduce computations	Monitoring support for execution graphs	Programming interface to monitor the progress of jobs	Minimal support for task level monitoring
Language Support	Implemented using Java.Other languages are supported via Hadoop Streaming	Programmable via C# DayadLINQ provides LINQ programming API for Dryad	Implemented using Java Other languages are supported via Java wrappers	C, C++, Fortran, Java, C#

access contention, repeatability, and protection and security all inhibit cloud paradigm adoption for certain critical uses. Also, it's unlikely that a general business model, implicit with clouds, will provide the extreme computing and peak performance. Finally, protected access to such facilities is a potential source of competitive edge for science, market, and national security, and the agencies that employ them will therefore limit or entirely preclude offering such systems to a cloud-covered processing world.

VI. ACKNOWLEDGMENTS

I would like to thank my colleagues on the HPU-HPC team for their contributions, insights, and support.

This paper is supported by the high-performance computing platform of Henan Polytechnic University.

REFERENCES

[1] WIKIPEDIA, "High-performance computing", http://en.wikipedia.org/wiki/High-performance_computing

[2] Cloud Computing Denition, National Insitute of Standards and Technology, Version 15, http://csrc.nist.gov/groups/SNS/cloud-computing/index.html

[3] Jaliya Ekanayake and Geoffrey Fox, “High Performance Parallel Computing with Clouds and Cloud Technologies”, 1st International Conference on Cloud Computing, Oct 19-21, 2009.

[4] ASF. 2009. Apache Hadoop Core. http://hadoop.apache.org/core.

[5] ASF. 2009. Apache Hadoop Pig. http://hadoop.apache.org/pig/.

[6] C. Vecchiola, S. Pandey, and R. Buyya, “High-performance cloud computing: A view of scientific applications,” CoRR, vol. abs/0910.1979, 2009.

[7] Amazon Elastic Compute Cloud (EC2), http://aws.amazon.com/ec2/

[8] Amazon.com, Inc. 2009. Simple Storage Service (S3). http://aws.amazon.com/s3.

[9] Thomas Sterling, Dylan Stark, A High-Performance Computing Forecast: Partly Cloudy, Computing in Science & Engineering, July/August 2009, pp.42-49.

[10] HPC as a Service, http://www.penguincomputing.com/POD/HPC_as_a_service.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 294-296

Based on Ant Colony Algorithm the Improved Service Composition method

Xu Hui[1], Huangfu Caihong[2]
Henan Polytechnic University, school of computer science and technology , Henan Jiaozuo, China
Email: xuhui@hpu.edu.cn
Henan Polytechnic University, school of Electrical Engineering and Automation, Henan Jiaozuo, China
Email: hfcaihong@hpu.edu.cn

Abstract—Pheromone is the core of ant colony algorithm. The distribution and the initial value setting of pheromone are in relation to the goodness and badness of an algorithm. Through the analysis and research of pheromone, this paper will give out Web Service model based on ant colony algorithm and transform service combination into coalition generation. It improves the efficiency of the optimal service through the establishment of pheromone and the improvement of updating pheromone strategy.

***Index Terms*—Web service, service composition, ant colony algorithm, coalition, pheromone**

I. INTRODUCTION

The rapid development of Web service technology brings about the increasing demand of Web service shared in the network. But a single Web service can't meet the complicated demand in real life. Therefore, how to make full use of these services and finish the appointed tasks with these well combined services become the focus of recent research. The concept of service combination arises at the historic moment. In fact, service composition is aimed at the value-added services with the principle of high efficiency, flexibility and priority. That's to say each task completed will accompany the help of multiple Web services [1].

Ant colony algorithm is a new simulated evolutionary algorithm and has been widely applied to solve combinatorial optimization. In the ant colony algorithm, pheromone through which the ants choose their road is a communication tool for transferring information among various ants. The lack of pheromone or large deviation of pheromone setting will have serious influence on the final result of algorithm. Pheromone plays a particularly important role on ant colony algorithm [2] [3]. Therefore, this paper applies ant colony algorithm into service composition and improves the pheromone updating strategy through the transformation of service combination into coalition generation.

A. Description of the combinatorial service problems

Now suppose there is a service combination W_i S S_m }, and W_i means i service among service combination（$1 \le i \le m$）, and each service consists of n_i candidate services[5] (as shown in figure 1).

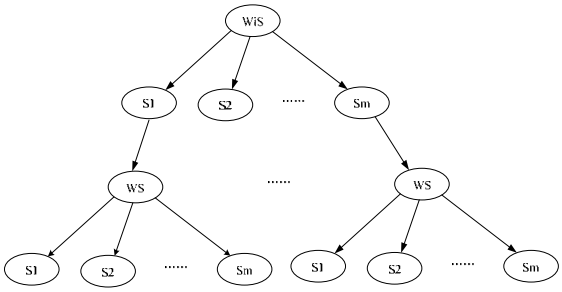


Figure 1. Service composition

According to the analysis above, we can extend more complex combination service model further into an coalition, as shown in figure 2:

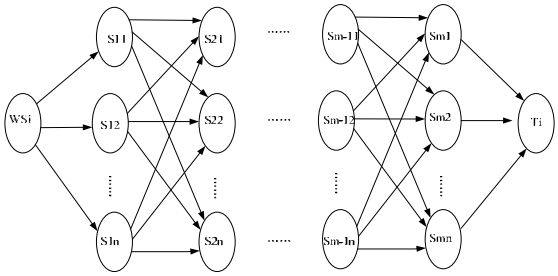


Figure 2. Service composition Model Picture

We can be regarded as an coalition service as it shown in the above figure. Suppose an coalition service WSi consists of m service and each service consists of n Web services. If the user asks for a service, we need choose a way from figure 2; that is to say, we need make an optimal coalition to achieve the best combination of all the services so as to complete the service request. We constitute the composition service based on quality of service (i.e QoS).

II. WEB SERVICE MODEL BASED ON ANT COLONY ALGORITHM

In order to better reflect the ant colony algorithm into the service combination, we introduce several concepts as below[5]:

Author Introduction: Xu Hui (1978 -), male, graduate. Research direction: GIS . Huangfu Caihong(1983-), female, graduate. Research direction: intelligent networks, grid computing.

Service Providers: receiving assigned tasks from agent center and fully updating pheromone it serves after finishing the task. Updated pheromone is housed in the Service functions.

User Demand: submitting its demand to agent center, receiving returned mission situation from agent center, communicating with service providers and finally cancelling communication after task is finished.

Agent Center: responsible for receiving and sorting user request , searching for service that meets the requirements, choosing an optimal solution according to local updating strategy of pheromone and feedback information to the user. Agent center has two tasks:

a) responsible for periodic inspection and reception of the registration of service providers and real-time update service request. When receiving the user's requests, it immediately updates information list.

b) responsible for receiving the user's service request. When finding out the service that fits for the user's requirement of the service, it elects an optimal solution according to local updating strategy of pheromone and sends the solution to the service provider according to the distribution plan.

Its architecture is as follows:

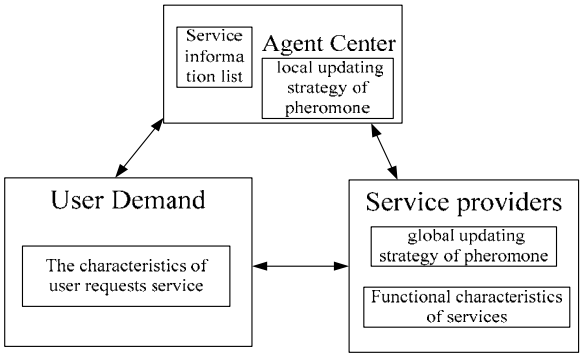


Figure 3. model of The Web Service Based on ACA

A. Algorithm Thought

According to the analysis above, we can apply the way that many ants of ant colony algorithm look for food from a starting point (i.e., the task) into method. It can be understood that the user brings about needs requirement (i.e. starting point, WSi). Then according to the needs of users, multiple relevant basic Web services (i.e, an coalition) are combined to commonly complete tasks (namely the target, Ti). In this way, we can change service composition into seeking for solutions that are appropriate from the starting point WSi to the designated target of Ti and QoS is better. Therefore, service composition can be turned into basic Web services coalition generation.

B. Establishment of information

First we define a Web service: the wsi (i.e., function , QoS, cost). The WS_i is the name of the Web service. Function means the Web service's abilities. QoS means the quality of services. Cost means the cost for service (i.e., response time, Cost, etc.).

When a new service WS_i joins the coalition service (WS), we need add new Web services that provide the above defined information. In 4.1.3 section, the inherent ability of service (WS_i) itself has been defined. We use B_{ij} to impress the WS_i's own ability. When the new service joins the coalition, pheromone needs to be initialized: $\Gamma_i(t)$ =B_{ij}, thus, the information is established.

C. Pheromones improvement

The updating of pheromones in ant colony algorithm plays a very important role. If the pheromone is very slow in renewal, it will increase the cost of total capacity of the coalition and add additional costs in the process of solving coalition cooperation t_j, (such as cost, , response time etc). Meanwhile, the rapid updating might lead to the inappropriate WS chosen by ants and the less better QoS ,which aren't what the users want to see. Therefore, the pheromone refreshing strategy means a lot to the efficiency of ant colony algorithm and the optimal solution.

Pheromones distribution mainly refers to updating rules of the dynamic local pheromone of the center and updating rules of global pheromone of service provider. Among them, the global information is updated on service provider, which makes the service providers become active participants , actively carrying out their information updating and adjustment. In addition, the local updating rules of pheromone is realized in the entity of center agency.

According to the analysis above, information updating strategies are as follows:

The pheromone strength at t time point(i,j) is impressed by $\Gamma_{ij}(t)$,that means residual amount of information of ants based on matching and flabby between task (Tj) and Web services(WS_i) . So pheromone strength of local updates for:

$$\Gamma_{ij}(t+1) = (1-\rho)\Gamma_{ij}(t) + \sum_{k=1}^{m} \Gamma_{ij}^{k}(t) \quad (1)$$

And pheromone strength of global updates for:

$$\Gamma_{ij}(t+1) = (1-\theta)\Gamma_{ij}(t) + \sum_{k=1}^{m} \Gamma_{ij}^{k}(t) \quad (2)$$

Among them,

$$\Delta\Gamma_{ij}(t) = \sum_{k=1}^{m} \Gamma_{ij}^{k}(t) \quad (3)$$

Type: ρ is the constant between 0 and 1, meaning the relative importance of local volatile pheromone , θ means global volatile coefficient, $\Delta\Gamma_{ij}^{k}$ is increment of pheromone strength of the ant from t to t + 1 between path (i,j). Pheromones strength is related to Web

service $V(c)$ on the relevant path , actually it's proportional to the relationship.

According to the analysis above, the global information updating for the service provider can use formula (2) to realize the pheromone updating. Local information updates of agent center can use formula (1). When the service provider finishes tasks assigned by agent center, according to the formula (3) the updated pheromone is housed into service function in order to increase its inherent abilities and improve the efficiency in finishing the next task. When the agent center receive the user's demand, according to the current information service list and the corresponding service capabilities B_{ij}, it selects an optimal solution (coalition) in accordance with pheromone local updating rules (1).

III. ALGORITHM

According to the ant colony algorithm based on the basic thoughts of the combination service, the basic steps of the algorithm are as follows:

a) The initialization parameter. All the services in the initial state need to provide parameters on capacity, QoS, costs and so on, which is the basis of the establishment of the initialization pheromone (B_k) of each resource. Set the initialization of information $\tau_{ij}(0)= B_k$ of digraph (i, j) and initial time $\Delta\tau_{ij}(0)=0$. According to the previous study, set time t = 0 and cyclic number N_c =0; set maximum cycle number $N_{c_{max}}$ and put m. ants on n service.

*b)*Collecting service information of new resources at beginning of each cycle, setting initialization pheromone for new resources, updating information according to the formula (1), marking service restoring (disconnection due to fault reasons and so on) , modifying pheromones based on task completion condition -- reward successes, punish failure, making stop marks for disconnection service..

c) cycle times $N_c \leftarrow N_c + 1$.

d) Index k = 1 on taboo list of ants.

e) Setting ants number k $\leftarrow$ k+1.

f) Ant individual chooses service and moves forward according to the probability calculated through state transition probability formula.

g) Modifying pointer of taboo table, namely moving ants after selecting a new service and adding the service to the ant individual's taboo list.

h) Checking whether the tabulist is already full or not, executing the (8th) step if full, repeating the step 6 if not.

i) If the services in digital collection N has not been traversed , jumping to the step 4 or carrying out step 8.

j) Updating pheromone on each path according to the formula (2).

k) Setting ants, choosing WS services which fit for and has better QoS (coalition has the larger values) as the end condition . If the service meet the requirements, namely if cycle number $N_c \geq N_{c_{max}}$, end circulation and output calculation result; otherwise empty taboo list and jump to step2.

CONCLUSION

This paper analyzed the importance of pheromone in ant colony algorithm and the influence of the pheromone updating strategy on the algorithm. Therefore, the web service model of ant colony algorithm of pheromone based on ant colony algorithm arises according to combination service. Through application of updating strategy of pheromone in ant colony algorithm to web service model, this paper proposes two updating strategies: pheromone global strategy and local updated pheromone strategy . Finally, there are the implementation steps of this algorithm.

REFERENCE

[1] Liao Jun, Tan Hao, Liu JingDe. Describing and Verifying Web Service Using Pi - Calculus. Chinese Journal of computers, Vol. (28),No.(4),apr,2005 ,635-643.

[2] Papazoglou MP, Georgakopoulos D. Service oriented of computing, Communications, ACM 2003,46 (10) : 25.28.

[3] Duan HaiBin. The principle and application of ant colony algorithm. Beijing: science press, 2005.

[4] Jiang ChangYuan. The ant colony algorithm theory and its application. Computers, June,2004,1-3.

[5] Luo JunGong Han HongJiang, etc. Multi-tier distributive system framework. Hefei university (natural science edition) based on Web Service , Vol. (27),No.(1),Jan.2004,18-22.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 297-300

A Fast Geometry Figure Recognition Algorithm Based on Edge Pixel Point Eigenvalues

Wenqing Chen[1,2], Leibo Yao[2], Jianzhong Zhou[1], Hongzheng Dong[2]
[1]College of Hydropower & Information Engineering HuaZhong University of Science and Technology, Wuhan, China
Email: cwq@lit.edu.cn
[2]Dep. of Electrical Engineering and Automation Luoyang Institute of Science and Technology, Luoyang, China
Email: leibo3008@126.com, prof.zhou.hust@263.net, lylgdhz@lit.edu.cn

***Abstract*—In view of some shortcomings about frequently-used currently shape recognition algorithms such as large amount of calculation, long processing time, single figure recognition or demanding to pre-set templates, a fast geometry figure recognition algorithm based on edge pixel point eigenvalues is presented in this paper, which polygon apexes and its rank orders are quickly recognized firstly based on the different variation laws of the eigenvalues of polygon apexes and other edge pixel point and the exact shape recognition of the polygon is finished as well, then the figure center and radius, the length of major and minor axle be can worked out by eigendistance and the equation of a circle or ellipse is constructed to make a fast recognition for a circle or ellipse be done. The simulation result shows the algorithm merits such as recognizing rich kinds of figure, lower computational complexity, higher processing speed, no pre-setting template.**

***Index Terms*—eigenedge-distance, eigenvalue, eigendistance, eigenvalue follow-pixel, figure recognition**

I. INTRODUCTION

Shape recognition is one of the most important research aspects in pattern recognition field and has been deeply pervasive in image analysis, machine vision, object recognition and other application fields. At present, the frequently-used shape recognition algorithms for the closed geometry figure such as ellipse, circle, polygon and so on, are based on Hough transform [1-4], Radon transform [5], neural network [6-7], shape matching [8], clustering [9] and other algorithms [10]. The algorithm based on Hough transform is popular and have advantages when dealing with beeline, curve, circle or ellipse, but it really has difficulties in the detection of other shapes and has large amount of calculation. Although the algorithm based on neural network is able to recognize more shapes, it requires pre-setting similar shape templates and training them and has large amount of calculation and higher time complexity. The algorithm based on shape matching requires the shape description and representation which are used for the comparison between the image to detect and the template. Furthermore, the shape descriptor is of extreme importance, but it is difficult to get a fit descriptor.

In view of some shortcomings about the algorithms above, a new algorithm is presented in this paper, which can recognize the closed geometry figures such as polygon, circle and ellipse and has some advantages such as lower complexity, higher speed and precision, no templates, easy realization, and RST(Rotation, Scaling and Translation) invariability.

II. ALGORITHM KEY ELEMENT DEFINITIONS

Definition 1 Set (x, y) be the coordinate of a certain pixel point on the image edges and (x', y') be the coordinate of another one, the distance between (x, y) and (x', y') along the edges is fixed and is called eigenedge-distance of (x, y), d_{EED}. Given an eigenvalue of (x, y) v_{EV},

$$v_{EV} = \left\| x - x' \right| - \left| y - y' \right\|.$$

The spot (x', y') is called eigenvalue follow-pixel point of the spot (x, y) and all eigenvalue follow-pixel points must be the same outspread direction, either clockwise or anti-clockwise. The straight-line distance between the spot (x, y) and the spot (x', y') is called eigendistance of the spot (x, y), d_{ED}. The value of d_{ED} is fixed and depends on the size of figure recognized.

It is an example about Definition 1 in Fig1. Set (x, y) be the coordinate of the spot P and (x', y') be the coordinate of the eigenvalue follow-pixel point of the spot P. The outspread direction of the eigenvalue follow-pixel point is anti-clockwise in Fig1, and the distance between (x, y) and (x', y') along the edges is d_{EED} of the spot P. The difference $\left\| x - x' \right| - \left| y - y' \right\|$ is the v_{EV} of the spot P. The straight-line distance between (x, y) and (x', y') is d_{ED} of the spot P. The d_{ED} of (x, y) is equal to d_{EED} in Fig 1.

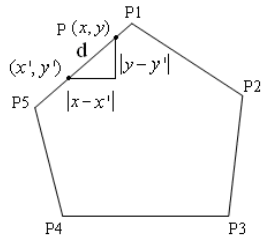


Figure 1. An example of Definition 1

According to Definition 1, the following theorem is established obviously.

Theorem 1 If two pixel points such as P and Q1 on the image edge and their eigenvalue follow-pixel points are situated on the same edge, the v_{EV} of P and Q1 is equal.

Proof: The area near P1 in Fig 1 is enlarged in Fig 2.

Because P, P', Q1, Q'1 are situated on the same edge, two triangles △PO1P' and △Q1O2Q1' are congruent ones and the v_{EV} of the spot P and Q1 is equal according to Definition 1. End.

According to Theorem 1, the following corollary is established obviously.

Corollary 1 The v_{EV} of each pixel points on a curve is different.

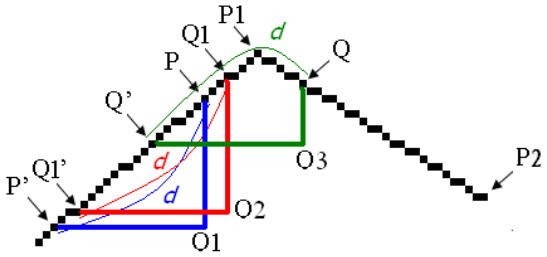


Figure 2. Equal and different v_{EV}

Combining view Fig 1 and Fig 2, obviously, the v_{EV} of each pixel point on PP1 is equal to the v_{EV} of P and P1 approximately, but the v_{EV} of each pixel point on P1P2 is different from the v_{EV} of each one on the PP1, because they are not situated on the same edge. If a spot and its eigenvalue follow-pixel are not on the same edge, its v_{EV} varies continuously with the spot's same search direction movement, either clockwise or anti-clockwise (clockwise in Fig2). Once the spot and its eigenvalue follow-pixel are situated on the same edge P1P2, the v_{EV} of the spot become steady again, and it is equal to the v_{EV} of P2. The pixels on P2P3 have the same law.

The law noted above can be shown in Fig3 as follows.

Given a fixed d_{EED}, each edge of a polygon (pentagon in Fig3) can be divided into two parts and the length of one part is equal to d_{EED}. Given the length of P1A is equal to d_{EED}, then the v_{EV} of each pixel on P1A change gradually and the v_{EV} of the pixel points on AP2 are equal to each other according to Theorem 1. The other edges have the same law.

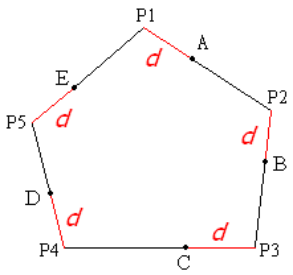


Figure 3. v_{EV} variation law

According to the description above, the following two corollaries are established obviously which are related figure recognition theoretical basis. Corollary 2 and Corollary 3 can be applied to the recognition of polygon and circle or ellipse respectively.

Corollary 2 The apex numbers of a polygon are equal to the times that the v_{EV} of all pixel points on the polygon edges changes from steady to variation along the polygon edges in a certain direction (clockwise or anti-clockwise).

Corollary 3 Given a fixed d_{EED}, the d_{ED} of each pixel point on a circle is the same to each other, but the d_{ED} of each pixel point and its adjacent ones on an ellipse is different. The midpoint between the pixel point with minimum d_{ED} and their eigenvalue follow-pixel point along the edge is the spot where the ellipse and its semi-major-axis cross and there are two such intersections on the ellipse. Similarly, when the midpoint with maximum d_{ED} appears, two intersections can be obtained where the ellipse and its semi-minor-axis cross, then the equation of the ellipse can be gained.

III. Recognition Algorithm Based on Edge Pixel Point Eigenvalues

According to the theorems and corollaries above, the geometry figures recognition algorithm is described detailedly as follows:

Algorithm

Input: closed geometry figure

Output: coordinate of all apexes and the center of circle or ellipse, length of radius or length of semi-major-axis and semi-minor-axis

Begin

1) Construct the set *EDGE* with all pixel points of the figure by edge tracing;

2) Delete the superfluous pixel points in *EDGE* and get the new set *EDGE'* to ensure that there are only two edge pixels adjacent to each pixel point in *EDGE'*;

3) Set a fit d_{EED} according to the circumference of figure;

4) Calculate the v_{EV} of each pixel in set *EDGE'*, and construct the v_{EV} set EV which is ordered and circular;

5) Analyze the times that the elements in *EV* vary from steady to variation and calculate the coordinates of the apexes at which the v_{EV} varies firstly every time;

6) If the elements in *EV* vary n times, the figure is a polygon with n edges and an ordered set of all apexes can be constructed simultaneously;

7) If the elements in *EV* vary continuously, the figure is not a polygon, then calculate the coordinate of the figure center according to *EDGE'* and the d_{ED} of every edge pixel points in *EDGE'* with the given d_{EED} and construct the d_{ED} set *ED* which is ordered and circular;

8) If the elements in *ED* are equal to each other, the figure is a circle. Then calculate the radius of the circle;

9) If the elements in *ED* are not equal to each other, calculate the minimum d_{ED} and maximum d_{ED} to locate four intersections mentioned in Corollary 3;

10) Calculate the length of semi-major-axis and semi-minor-axis and construct the equation of an ellipse according to the center coordinate in step 7);

11) If there are enough pixel points satisfying the equation in *EDGE'*, it is ellipse;

End

IV. ALGORITHM REALIZATION AND SIMULATION RESULTS

The algorithm is simulated by VC++ 6.0 on the computer with CPU 1.4G and memory 512M. The figures in Fig 4 have been recognized by the algorithm, and the resolution is 256*256.

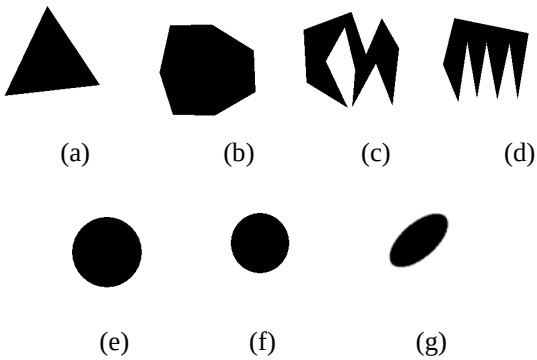


Figure 4. Geometry figure recognized

The recognition results are shown in Table 1.

TABLE I. RECOGNITION RESULTS

No.	C	T	Result	
a	440	63	3 apexes	(2.0,0.6)(3.5,2.9)(0.7,3.2)
b	506	78	7 apexes	(2.5,1.2)(3.7,1.9)(3.7,3.2) (2.5,3.8)(1.3,3.8)(0.9,2.5) (1.2,1.1)
c	912	94	13 apexes	(2.2,0.8)(2.6,2.0)(3.1,1.0) (3.6,1.9)(3.4,3.5)(2.9,2.3) (2.2,3.5)(2.3,2.5)(2.0,1.2) (1.4,2.3)(2.0,3.6)(0.8,2.9) (0.8,1.3)
d	1019	93	10 apexes	(0.9,0.6)(3.3,1.1)(2.9,3.2) (2.7,1.4)(2.3,3.2)(1.9,1.4) (1.6,3.1)(1.3,1.4)(1.0,3.3) (0.5,2.1)
e	334	63	circle	(2.3,2.4) R=1.1
f	228	48	circle	(2.8,2.4) R=0.7
g	240	109	ellipse	(2.0,2.1) A=1.1 B=0.5

Notes: 1. Suppose the coordinate of upper left corner is (0, 0);
2. C is circumference of figure;
3. T is algorithm running time, unit: ms;
4. R is radius of a circle;
5. A is length of semi-major-axis;
6. B is length of semi-minor-axis.

V. DISCUSSIONS

According to the algorithm description and simulation results above, some discussions about the algorithm performance can be done as follows:

1) Suppose that the resolution size of the figure is W*H, according to the description above, the scale of the data processed in every step is directly proportional to the scale of the edge pixel points obviously. So the time and space complexity of the algorithm is O(n) and n(n<<W*H) is the number of the edge pixel points.

2) According to the simulating results in Table 1, the recognizing time is directly proportional to the circumference and the numbers of the figure apex. It is faster to recognize circle than ellipse with the same circumference.

3) Comparing with the recognition algorithms based on Hough transform, the recognition algorithm based on edge pixel point eigenvalues with lower time and space complexity lowers the realization difficulty.

4) The recognition algorithms based on neural network which requires training with pre-setting similar shape templates will cost huge amount of calculation and high time complexity, while all these shortcomings can be overcome by applying the recognition algorithm based on edge pixel point eigenvalues.

5) The recognition algorithm based on edge pixel point eigenvalues need not the shape description and representation and the comparison of the similarity between the image to detect and the template, but it necessary for the algorithm based on shape matching.

6) Abundant information such as the coordinate of all apexes and the center of circle or ellipse, the rank order of the apexes, the length of radius or length of semi-major-axis and semi-minor-axis can be obtained through the procedure of the algorithm programming.

7) Recognition results will not be affected by rotation, scaling and translation of the figure in the recognition algorithm based on edge pixel point eigenvalues.

VI. CONCLUSION AND FUTURE WORKS

The simulation results and discussions above indicate the recognition algorithm based on edge pixel point eigenvalues has some advantages such as lower time and space complexity, higher speed and precision, no templates, easy realization, and RST (Rotation, Scaling and Translation) invariability , but there are still some problems need to be tackled.

1) It will have difficulties for the algorithm in this paper in dealing with the figure with noise.

2) The image recognition with multiple figures still needs to be improved.

3) The figure such as polygon, circle or ellipse suit the algorithm best while others not.

Further studies will be helpful to solve the problems above.

REFERENCES

[1] Teng Jinzhao, Qiu Jie, "Fast and precise detection of straight line with hough transform," J. *Journal of Image and Graphics*, 2008, pp. 234-237.

[2] Chen T C, Chung K L, "An efficient randomized algorithm for detecting circles," J. *Computer Vision and Image Understand*, 2001, pp. 172-191.

[3] Chui S H, "An effective voting method for circle detection," J. *Pattern Recognition Letters*, 2005, pp.121-133.

[4] Zhou Y J, Zheng Y P, "Estimation of muscle fiber orientation in ultrasound images using revoting hough transform (RVHT)," J. *Ultrasound in Medical and Biology*, 2008, pp.1474-1481.

[5] Pan Wei, Zheng Haijiang, "Beeline detection and implement based on ridgelet transform," J. *Journal of Xiamen University (Natural Science)*, 2006, pp. 775-778.

[6] Osowski S, Nghia D D, "Fourier and wavelet descriptors for shape recognition using neural networks-a comparativestudy," J. *Pattern Recognition*, 2002, pp. 1949-1957.

[7] Du J X, Huang D S, Wang X F, “Shape recognition based on neural networks trained by differential evolution algorithm,” J. *Neurocomputing*, 2007, pp. 896-903.

[8] Zhang Xianquan, Guo Mingming, Tang Ying, “New geometric feature shape descriptor,” J. *Computer Engineering and Application*, 2007, pp. 90-92.

[9] Ning Li, Lin Yi, “Point-set clustering based geometry-figure,” J. *Computer Engineering and Design*, 2008, pp. 2613-2615.

[10] Zhang D S, “Review of shape representation and description techniques,” J. *Pattern Recognition*, 2004, pp. 1-19.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 301-303

A new algorithm for service composition model

Huangfu Caihong[1], Xu Hui[2]
[1]Henan Polytechnic University, school of Electrical Engineering and Automation, Henan Jiaozuo, China
Email: hfcaihong@hpu.edu.cn
[2]Henan Polytechnic University, school of computer science and technology, Henan Jiaozuo, China
Email: xuhui@hpu.edu.cn

***Abstract*—Because of the rapid development of Web services, it is a urgent need to solve the problem that how to find the optimal Web services combination quickly and efficiently in the network. This paper gives a new service composition method, builds the corresponding algorithm model, defines the algorithm parameters, and provides the corresponding service matching mechanism and evaluation mechanisms, so it puts forward a new way of thinking to solve the combination of services.**

***Index Terms*—web services, service composition, ant colony algorithm, algorithm model, service matching**

I. INTRODUCTION

With the growing maturity of Web services technology, more and more stable and easy available Web services share in the network. Due to a single service can provide limited functionality, how to combine Web services shared to become a value-added services, how to provide for the more strong service capabilities to better meet the needs of users is the urgent need to address the problem. Services combination have already made a lot of research in a different extent at home and abroad, and these researches have promoted the development of service composition[1]. Ant colony algorithm is another heuristic search algorithm by using swarm intelligence to solve combinatorial optimization problems after tabu search algorithm, artificial neural network algorithm. It has some characteristics, such as intelligent search, global optimization, robustness, positive feedback, distributed computation, and easy integration with other algorithms[2][3]. Therefore, the basic ant colony algorithm model modified can be used for other combinatorial optimization problems. Based on these characteristics, this paper will apply Ant colony algorithm to the service composition method to find the best Web services to meet user needs.

II. NEW SERVICE COMPOSITION MODEL

A. The definition of service composition

Service composition problem[4] can be defined as follows:

Supposing Web Service collection is $WS_i = \begin{pmatrix} ws_{11} & \cdots & ws_{m1} \\ \cdots & & \cdots \\ ws_{1n} & \cdots & ws_{mn} \end{pmatrix}$, each ws_{ij} all have a capability: $B_{ij} = \left\{ b_{ij}^1, b_{ij}^2, \cdots, b_{ij}^r \right\}$, $b_{ij}^k \geq 0$, ($1 \leq i \leq n$, $1 \leq j \leq n$, $1 \leq k \leq r$), among them, b^k_{ij} is used to describe the ability of ws_{ij} to provide. Task collection T= {t_1, t_2, t_i, … , t_m}, $t_k \geq 0$, （$1 \leq i \leq m$）, each t_k has a certain capability demand $B_k = \left\{ b_k^1, b_k^2, \cdots, b_k^l \right\}$, $b_k^j \geq 0, (1 \leq k \leq n, 1 \leq j \leq l)$, when ws_{ij} is selected to complete the task t_j, ws_{ij} will obtain the corresponding interest $P\left(t_j\right)$.

B. Algorithm Description

Defined as follows: suppose m is the number of ant colony, d_{ij}(i= 1, 2, … , n) expresses additional spending of WS_i which completes task T_j, such as cost, response time; Γ_{ij} expresses pheromone quantity which ants remain accordance with the degree of slack match about task T_j and Web service WS_i. When ant k(k=1, 2, … , m) is in the course of the campaign, the direction of ant transfer is decided by WS_i pheromone quantity. When the ant comes to a node, the node is increased as the coalition; P_i^k expresses the probability of ant k choosing WSi, that is the probability of ant k choosing WS_i to join the coalition.

$$p_{ij}^k = \begin{cases} \dfrac{[\Gamma_{ij}]^\alpha [d_{ij}]^\beta}{\sum_{u \in J_k} [\Gamma_{iu}]^\alpha [d_{iu}]^\beta}, i \in J_k \\ 0, else \end{cases} \quad (1)$$

Among them, J_k is WS collection which ant k doesn't visited yet. Parameters α and β are used to control the familiarity degree and the relative importance degree of additional spending.

$\Gamma_{ij}(t)$ expresses pheromone concentration at t（i, j）time, that is pheromone quantity which ants remain accordance with the degree of slack match about task T_j and Web service WS_i. Then the pheromone concentration at t+1 time is:

$$\Gamma_{ij}(t+1) = (1-\rho)\Gamma_{ij}(t) + \sum_{k=1}^{m} \Gamma_{ij}^k(t) \quad (2)$$

Among them: ρ is constant about 0 and 1. It is the relative importance of remaining pheromone; $\Delta\Gamma_{ij}^k$ is the increment of pheromone concentration when the ant k at

Author Introduction: Huangfu Caihong(1983-), female, graduate. Research direction: intelligent networks, grid computing. Xu Hui (1978 -), male, graduate. Research direction: GIS

time t to t +1 between the route (i, j). Pheromone concentration is related with $V(c)$ of Web Service in corresponding path, and the increment of pheromone concentration is proportional to $V(c)$.

In this paper, through ant colony algorithm , the optimal solution will be found among Web service compositions, so the best combination of services can be achieved.

C. Service Matching Mechanism

When users propose a number of Web services to establish coalition C, the initial ant carries the basic information of task t_i. We define task t_i(Id, TaskName, Information, T_{need}), among them: Id is the number of tasks; TaskName is the name of tasks; Information is the basic information which task t_i carries match service; T_{need} is the time limit of completing task. If task t_i doesn't find required services within the prescribed time, task fails. We has been defined the capacity of the task t_i in Section 1.1, so B_k expresses the capacity of the task t_i. According to the value of B_k, we can match the Web service collection from task t_i. When ants arrive to certain WS to find task t_i completed, ants can stop searching. That demands once ants reach nodes, the ability vector of WS is increased. Calculating the current ability vector, and judging whether the ability demands is complete. If yes, stop searching, else continue routing. When the last ant forms task solving coalition and stops searching, a cycle ends.

D. Service Evaluation Mechanisms

According to the definition of service composition in Section Ⅱ, we define a coalition C as a non-empty subset of N. Coalition C has an ability vector $B_c = \left\{b_c^1, b_c^2, \cdots, b_c^r\right\}$, B_c is the sum of all WS ability vector in coalition, $B_c = \sum_{0<i<r} B_i$. The requirement of coalition C finishing task t_j is: $\forall\ 1 \le i \le r, b_j^i \le b_c^i$. The value of each coalition C is given by characteristic function $V(c)$. Suppose $V(c) \ge 0$, $V(c) = P(t_j) - F(c) - C(c)$. Among them, $P(t_j)$ expresses the benefits obtained by the completion of tasks; $F(c)$ is the cost of coalition members the total capacity equivalent ; $C(c)$ is the additional cost of coalition members cooperating to solve t_j, such as expenses, reaction time. If coalition C does not satisfy the mentioned necessary conditions, $V(c)$ is 0, else $V(c)$ is positive number [5]. In this process, $V(c)$ is one of the important standard in choosing and judging Web service QoS. It reflects that the task has the capacity of completing good and bad in process of the user selecting required service. The larger the value of coalition $V(c)$,that more definitely affirmative by the user's, the lower cost, the better Qos of service, the higher efficiency in completing task, the smaller the possibility to refuse to implement tasks of after the coalition receive the user makes the demand again [5] [6] [7]. When users propose requirement, if the costs which Web service completing tasks don't be considered and solution spaces exit in system, then solution can be found, but the quality can't guarantee. If the task of the evaluation mechanism are abandoned, then that will lead to add to communication number of times, decline the whole performance, lower the efficiency of the algorithm, and increase user costs.

III. ALGORITHM REALIZATION

In the initial, m ants are placed in different nodes (services) , and each side is given the pheromone amount τ_{ij} (t) = C, where C is said that constant, that the pheromone amount of each path are equal. The first element of the tabulist of each ant is assigned to in the node (service). When the ants completed a cycle, calculating $\Delta\tau_{ij}$ (t), and updating the amount of information each side, then starting a new round of circulation. When the circular loop reach preset the maximum number of cycles $N_{c\max}$ or when all the ants choose the same path approach, the program terminates.

The pseudo-code of algorithm model program are as follows:

The first step:

initial parameters: Set t=0, N_c =0, on each side τ_{ij} (0)=B_k, and $\Delta\tau_{ij}$ (0)=0

initial all ants $tabu_k$

Repeat until not termination condition

For k=1 to m do

m ants are placed randomly in the n nodes (service), capturing the initial pheromone of services;

The second step:

Set s=1 (s is the subscript of tabulist)

For k=1 to m do

the initial node (service) capacity vector of k-ant are placed onto the $tabu_k$ (s)

pheromone part updates;

The third step:

Repeat until tabulist is full

Set s=s+1

For k=1 to m do

According to the transition probability $p_{ij}^k(t)$, the next step nodes (services) are selected, and k-ant will be transferred to the node j (services), and the capacity of vector j insert s into $tabu_k$ (s) in

The fourth step:

For k=1 to m do

Calculating the current coalition k-ant capacity vector L_k , judging whether the path through the service nodes by the combination of QoS are the best, if yes, set this path to the best path;

Update the pheromone table value;

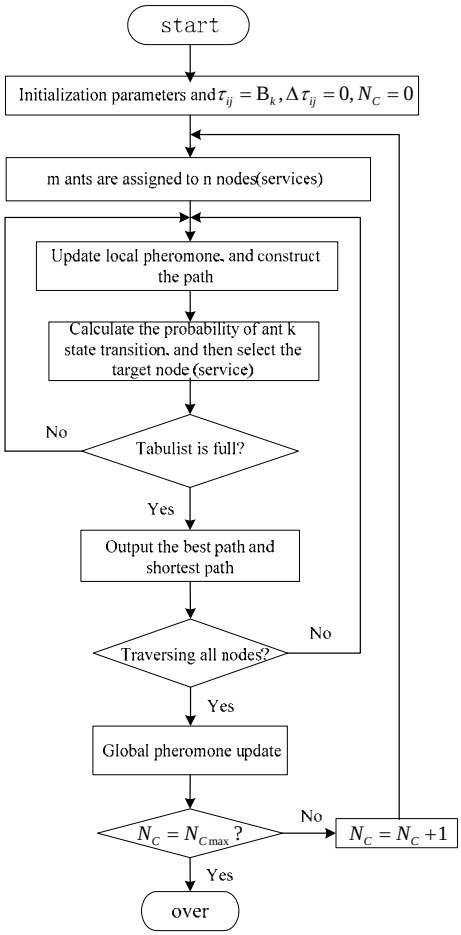


Figure 1. Procedure of algorithm

The fifth step:

Calculating each side τ_{ij} (t+n)

Set t=t+n

Set $N_c = N_c + 1$

The sixth step:

If($N_c < N_{c_{max}}$) and(not all ants choose the same path)

Then empty other service node records from the start node to node traversed circulates, clear all the tabulist

Go to the second step

Output the shortest path

Termination of the program

Algorithm flow chart shown in Figure 1.

IV.CONCLUSION

This paper applies ant colony algorithm to services composition, based on the superiority of ant colony algorithm in solving to the coalition generating problem. The paper establishes the corresponding algorithm model, gives the service matching mechanism, makes ant colony algorithm which is applied to service composition problems possible. While the corresponding evaluation mechanism is given, service quality has also standard

REFERENCE

[1] HU Hai-Tao, LI Gang, HAN Yan—Bo. An Approach to Business-User-Oriented Larger-Granularity Service Composition. Chinese Journal of Computers, 2005, 28(4): 693-703

[2] YE Zhiwei 1 ZHENG Zhaobao. Configuration of Parameters α,β,ρ in Ant Algorithm, Editorial Board of Geomatics and Information Science of Wuhan University, 2004, 29（7）:597-601

[3] M Dorigo, V Maniezzo, A Colorni. The ant system: optimization by a colony of cooperating agents[J]. IEEE Transactions on Systems, Man, and Cybernetics-Part B, 1996, 26(1): 29~41

[4] LUO Jun-hong; HAN Jiang-hong; ZHANG Li; ZHANG Jian-jun. Multi-tier distributive system framework based on Web Service. Journal of Hefei University of Technology(Natural Science), 2004, 27(1), 18-22

[5] JIANG Jian-guo; XIA Na; QI Mei-bin; MU Chun-mei. An Ant Colony Algorithm Based Multi-task Coalition Serial Generation Algorithm [J] . Acta Electronica Sinica, 2005, 33(12A):2178-2182

[6] Li Man, Wang Da-Zhi, Du Xiao-Yong, Wang Shan. Dynamic Composition of Web Services Based on Domain Ontology. Chinese Journal of Computers, 2005,28(4):644-650

[7] LIAO Yuan; TANG Lei; LI Ming Shu. A Method of QoS-Aware Service Components Composition. Chinese Journal of Computers, 2005, 28（4）: 627-634

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 304-307

The Improvement of Replacement Method for Web Caching

Rui Wang, Jing Lu
School of Computer Science and Technology / Henan Polytechnic University, Jiaozuo, China
wangrui@hpu.edu.cn

***Abstract*—Currently, the implementation of WEB caching is mostly based on traditional cache updating algorithms. However, due to the diversity of the WEB traffic pattern, the traditional algorithm s for cache updating can not be used in WEB environment effectively. Literature (F.Bonchi,2001) provides a Web updating algorithms based on Decision Trees, in this paper, we provide a web policy, based on Genetically Evolved Decision Trees, the result of experiment indicates this policy improves the updating efficiency of the Web pages.**

Index Terms—**Wed data mining; Web caching; Decision trees; Replacement algorithms; C4.5; GATree**

I. INTRODUCTION

The rapid expansion of the World Wide Web has resulted in major network traffic and congestion. Web data circulation has been almost doubling every six months, and despite efforts for capacity increases demands aren't always kept up . Improving response times and access latencies for clients became a quite important and challenging issue. Web caching has been proposed as a technique to reduce both the Internet traffic and the access times for (frequently) requested objects. Many of the Web caching aspects are originated from the caching idea implemented in various computer and network systems and web caching introduces new issues in Web objects management and retrieval across the network. The overall process of accessing data is no longer dependent on the client/server interaction. A client requests object(s) residing at a server, but instead of accessing the specified server, its local storage media is checked first. If the requested data resides in local cache is withdrawn from there with no extra network access cost, otherwise the original server needs to be contacted. Web Caches are implemented such that information will reside closer to user(s) since clients retain a local cache for Web objects storage. Therefore, both the load of the origin servers and the network traffic reduces, since upon requesting Web objects the clients can access their local cache instead of fetching the data from their original server. In this paper, the problem of supporting effective Web object caching is addressed and certain evolutionary techniques are proposed.

LRU(least recently used) is the most widely used important replacement algorithm ever developed for main memory and disk caching. LRU exploits temporal locality of reference, keeping the recently used while dropping the least recently used objects. LRU is simple to implement, robust and effective in paging scenarios. However, LRU is not suitable for web caching because it focuses on the recently used and equal size objects. Web caching is on the basis of documents that vary dramatically in size. Moreover, the recently accessed data may be just temporarily referenced.

Another replacement algorithm (named s2) based on C4.5 achieves higher performance than the conventional LRU(F.Bonchi,2001). In this article, we proposed a new web caching policy (named S2.1), which is compared with C4.5 and GATree, to improve the performance for web caching.

II. C4.5 ALGORITHM AND GATREE

ID3 is a well-known machine learning algorithm, representing decision-tree-based method in inductive learning. ID3 and the later algorithm C4.5 are both top-down learning algorithm which come from development of Concept Learning System (CLS) of Hunt by Quinlan. Through learning a group of training data, a decision tree structure knowledge representation was constructed. By Comparing the attribute value of the internal node in decision tree and judging the following down branches of a node according to different attribute value, the conclusion has come out at the leaf node of the decision tree, So a path from the root to leaf nodes corresponds to a decision rule, the whole decision tree corresponding to a group of disjunctive expression rules. The greatest advantage of decision-tree-based learning algorithm is that it does not require the user to understand a lot of background knowledge in the learning process. Such as long as the training data can be found expression in attribute-conclusion, we can use this algorithm to learn.

C4.5 algorithm is an improvement over ID3 algorithm, inheriting all the advantages of ID3 algorithm. For example, C4.5 also adopted the "window" concept. We can construct a decision tree using part of cases at first, then test and adjust it by using the remaining cases. C4.5 algorithm can handle continuous-valued types of attribute, it can classify the attribute set to equivalence classes, and the attribute values in the same class will come on the same branch in judgments. Coupled with simple, efficient, reliable, C4.5 algorithm becomes more significant in the inductive learning. There are also some inadequate about C4.5 algorithm. First, C4.5 uses the divide and conquers strategy, and local optimal algorithms in the internal nodes of the tree. Therefore it received the final results

despite high accuracy, but still can not reach the global optimum results. Secondly, the evaluation of the decision tree in C4.5 is mainly based on the error rate, not consider the depth of the tree, the number of nodes. However, the average depth of the tree directly corresponds with the forecast rate of the decision tree, and the number of nodes of the tree represents the size of the tree. Thirdly, because the evaluation of the decision tree is acquired with the structure at the same time, it would be hard to adjust the structure and content of the tree, when it is constructed. The improvement of the decision tree is very difficult. In addition, the classification of attribute value in C4.5 must test each node, and does not have a mechanism for the use of heuristic search, so is less efficient.

GATree algorithm is an optimization Decision Tree Algorithm on the basis of the genetic algorithm. We know that the groups search strategy and the information exchanging among individuals are the two major characteristics of genetic algorithms, mainly with the performance at the global optimum performance and potential parallelism. As in the process of the tree structure C4.5 does not necessarily get the optimal decision tree, although the results of genetic algorithms and evolutionary theory can not be guaranteed the theoretical optimal decision tree, but it provides a method can be tried. Due to the survival of the fittest, it makes the more adaptable decision tree to retain as far as possible, and it makes the more adaptable decision tree appeared in the process of evolution as a result of offering the adjustment and reorganization of the mechanism of the decision tree.

The shape and the number of nodes of the decision tree may be greatly different in different individuals. Also, because the number of attributes and attribute values are no restrictions on the number of the amount, a fixed-length strings to decision tree is not appropriate.

We use GATree representation to build a population of minimal binary decision trees (trees that consist from one node and two leaves). Every decision node has a random chosen value as its installed test. This is done in two steps. First we choose a random attribute. Then, if that attribute is nominal we randomly choose one of its possible values; if it is continuous we randomly pick an integer value belonging to its min-max range. This approach reduces the size of the search space and it is straightforward. Still, it has problems with real-valued attributes; for this work

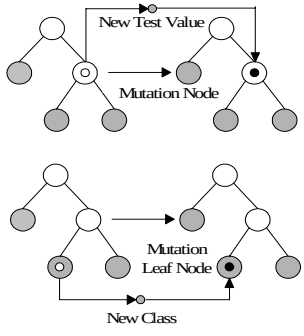


Figure 1. Example of the mutation operation

we concentrated on nominal attributes. Leaves are

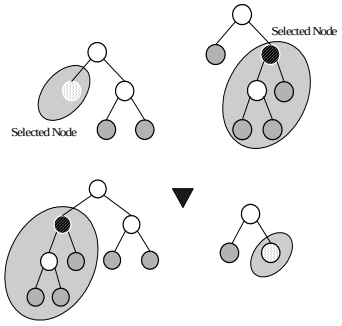


Figure 2. Example of Crossover operation

populated using the same line of thought; we just pick a random class from the ones available.

The basic form of the proposed algorithm introduces minimum changes to the mutation-crossover operators. Mutation chooses a random node of a desired tree and it replaces that node's test-value with a new random chosen value. When the random node is a leaf, it replaces the installed class with a new random chosen class (Figure 1).

The crossover operator chooses two random nodes and just swaps those nodes' sub-trees. Since predicted values rest only on leaves, the crossover operator does not affect tree's coherence (Figure 2).

Payoff function:

$$payoff\ (tree_i) = correctCla\ ssfied_i^2 * \frac{x}{size_i^2 + x}$$

Table 1. C4.5 compared with the GATree algorithm

	Accuracy			Size	
	C4.5	OneR	GATree	C4.5	GATree
Colic	83.84±3.41	81.37±5.36	85.01±4.55	27.4	5.84
Heart-Statlog	74.44±3.56	76.3±3.04	77.48±3.07	39.4	8.28
Diabetes	66.27±3.71	63.27±2.59	63.97±3.71	140.6	6.6
Credit	83.77±2.93	86.81±4.45	86.81±4	57.8	3
Hepatitis	77.42±6.84	84.52±6.2	80.46±5.39	19.8	5.56
Iris	92±2.98	94.67±3.8	93.8±4.02	9.6	7.48
Labor	85.26±7.98	72.73±14.37	87.27±7.24	8.6	8.72
Lymph	65.52±14.63	74.14±7.18	75.24±10.69	28.2	7.96
Breast-Cancer	71.93±5.11	68.17±7.93	71.03±8.34	35.4	6.68
Zoo	90±7.91	43.8±10.47	82.4±4.02	17	10.12
Vote	96.09±3.86	95.63±4.33	53.48±4.33	11	3
Glass	55.24±7.49	43.19±4.33	53.48±4.33	60.2	8.98
Balance-Scale	78.24±4.4	59.68±4.4	71.15±6.47	106.6	8.92
AVERAGES	78.46	72.64	78.75	43.2	7.01

GATree was able to produce the most accurate results (Table 1) even though the difference with C4.5 is not significant. However, those results were accompanied by extremely small decision trees (C4.5 produced six times bigger trees on average).

III. S2.1 REPLACEMENT STRATEGY

As mentioned above, C4.5 produces good accurate results but with unnecessarily big trees. So we take

GATree as the classifier algorithm in S2 to improve the performance for the web caching.

A. log data preprocessing

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

Common log file format standard

216.140.123.22 - - [31/May/2003: 05:54:15 +0400] "GET / img/logo.gif HTTP/1.0" 304 1164

216.140.123.22 is the host;

[31/May/2003: 05:54:15 +0400] is the timestamp;

"GET / img / logo.gif HTTP/1.0" is the requests; Thereinto, "GET" is HTTP method, "/ img / logo.gif" is the requested address (URL), "HTTP/1.0" is the HTTP protocol.

304 is the HTTP response code;

1164 is the response bytes.

According to S2 algorithm, there are defined as follows:

Ndir – the number of URL directory layers; in the cases should be 1;

FirstDir - the first directory layer of URL, in the cases should be "img";

NextAccess - the total request of the same URL before the next visit;

FileExt - the request URL document file name suffix, in the cases should be "gif";

Hour - request moment, in the cases should be 5;

Size – the number of bytes response to client; in the cases should be 1164.

According HTTP1.1 protocol(L.Masinter,et al.,1999), the URL can be used as cache file should be: HTTP method must be "GET", HTTP response code must be 200, and the request URL does not contain any parameters (ie the URL does not contain "?").

B. GATree Algorithm

The definition of classification according to the cache size as follows:

Cache (s) - the Web cache system with the size of s;

AvgDSize (s) - the average file size in the Web cache;

Tertile (t, s), t ∈ {1,2,3} - the file number for the cache storage state of t * 33.3%;

Max (s) – the number of individual when cache is full,;

Class0 - NextAccess ∈ (1, Tertile (1, s));

Class1 - NextAccess ∈ (1, s), Tertile (2, s));

Class2 - NextAccess ∈ (Tertile (2, s), Tertile (3, s));

Class3 - NextAccess ∈ (Tertile (3, s), Max (s));

Ndir, FirstDir, NextAccess, FileExt, Hour, and Size are observation attributes of the GATree algorithm.

C. weight distribution

The weight distribution of LRU replacement strategy as follows:

WLRU (Ei) = j (j is the visit time for the file Ei)

The weight distribution of S2.1 replacement strategy as follows:

WS2.1 (Ei) = j + α (c) * AvgDsize (s) / Ei.size; c ∈) (0,1,2,3;

c is the category which file Ei belong to according to GATree algorithm.

A (3) = Max (s);

A (c +1) = 2 α (c);

Ei.size is the size of file Ei

D. Performance Test

We will use Berkeley Web log (Berkeley,2007) as a test data, simulator test results are shown in Figure 3.

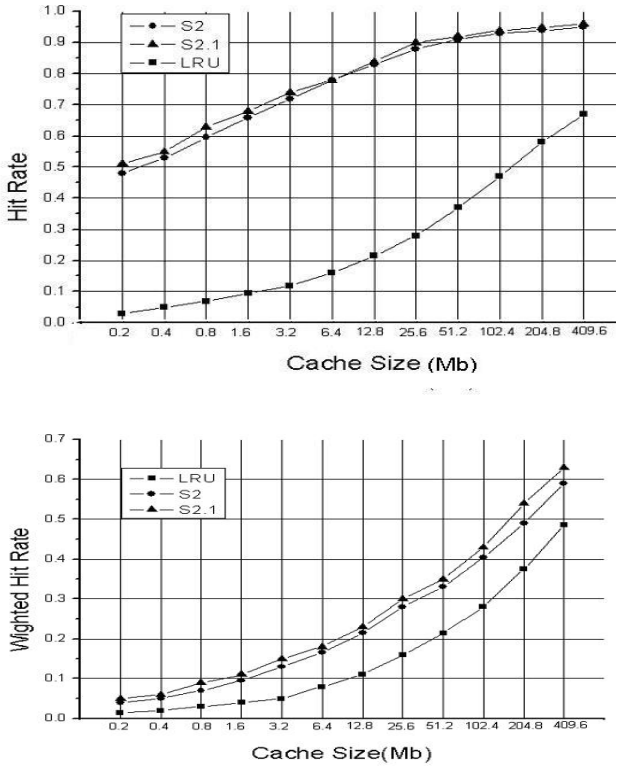


Figure 3. Comparing experiment with Berkeley log

IV. CONCLUSION

According to the results of the experiment: First, with the same size cache, S2.1 algorithm has higher hit rate than S2 algorithm, and even with the same hit rate, S2.1 algorithm is in responses to the client URL request faster than S2 algorithm, because of the decision tree of GATree algorithm is far smaller than C4.5 algorithm.

A basic drawback of GAs, when processing bit logs, GATree is tardiness, as mentioned in the literature (Athanasios Papagelis,2002). So S2.1 replacement strategy can only change decision tree of the system in the short-term, not adapt to online learning, it would be more suited to the Web server cache system, but not proxy server system.

REFERENCES

[1] F.Bonchi.(2001).Web log data warehousing and mining for intelligence web caching[J]. Data & knowledge engineering, 2001,39:165-189.

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[3] Williams,S.,Abrams,M.,Standridge,C.,et al.(1996).Fox removal policies in net work caches for world wide web documents[A].In: Steenstrup, M.,ed. Proceedings of the ACM SIGCOMM '96[C]. ACM Press, 1996, 293-306.

[4] Athanasios Papagelis . (2002).GATree: Genetically Evolved Decision Trees[EB/OL]. http://www.gatree.com. 2002.

[5] L.Masinter,T.Berners-Lee,R. Fielding, (1999).Hypertext transfer protocol- TTP/1.1,Tecchnical Report RFC 2396, The Internet Society[EB/OL], http://www.w3.org/ Protocols,1999.

[6] Temple University,(1999).Computer and Information Sciences Department in the college of science and Technology of Temple University. building Classification Models:ID3 and C4.5[EB/OL] http://www.cis.temple.edu/ ~ingargio/ cis587/ readings/ id3-c45 .html , 1999

[7] Berkeley,(2007).Department of Electrical Engineering and Computer Sciences, University of California, Berkeley CS HTTP logs[EB/OL], http://www.cs.berkeley.edu/logs/http.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 308-311

A Novel Preprocessing Approach for Digital Meter Reading Based on Computer Vision

Lei Haijun[1],Li Lingmin[2], Li Xianyi[3*]
[1]College of Computer and Software,Shenzhen University,Shenzhen,China
leihaijun2002@163.com
[2]College of Information Engineering,Shenzhen University,Shenzhen,China
[3]College of Mathematics and Computational Science,Shenzhen University,Shenzhen,China

***Abstract*—Through analyzing and researching the characteristic of the digital meter image, a preprocessing approach for digital meter reading based on computer vision is presented. Firstly, the digital meter image was filtered by enhanced homomorphic filter. Then, the image was binarized with Otsu's method and rotated with the Cartesian moments. Finally, the image was segmented into several sub-blocks, each of which contains a single character. The experimental results show the effectiveness of our method for digital meter image under uneven illumination.**

***Index Terms*—computer vision; image preprocessing; enhanced homomorphic filtering; digital meter;**

I. INTRODUCTION

In recent years, the digital meter is used in variety industrial measurement and control applications for higher accuracy, easier manipulation and multifunctional[1]. Though the interface for wireless communication is found in some high-grade digital meter, in some circumstances such as scientific experiment, measurement controlling, power meter reading[2] etc, the measurement results still need manual reading. Those results would be inputted into the computer for the post processing from the record sheet. This is a time-consuming, ineffective and low accuracy method, so people try to find a way to read the results from the meter automatically. The automatic meter reading system based on computer vision is one of the solutions.

The digital meter automatic meter reading (AMR) process, show in Fig.1, consists of three main parts: meter image capturing, meter image preprocessing and meter recognition.

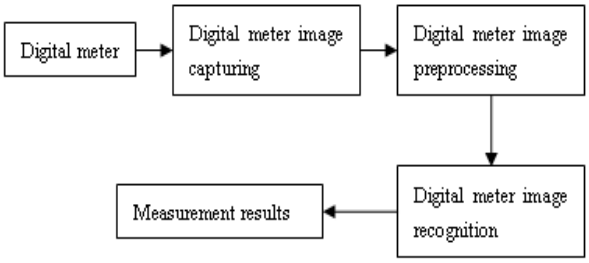


Figure 1. The framework of digital meter reading based on computer vision

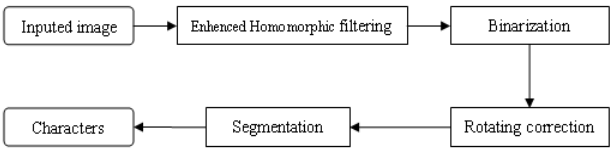


Figure 2. The flowchart of digital meter preprocesing procedure

And the meter image preprocessing procedure, as shown in Fig.2, includes homomorphic filtering, binarization, rotating correction and segmentation. In the AMR system, image preprocessing plays a vital role. The quality of image preprocessing directly affects the overall performance of the recognition method. The digital meter, generally, shows the measurement results (include characters and Arabic numerals, we also call it foreground area) in the liquid crystal display (LCD). The foreground area is showed in black or other dark color and the background area is showed in light color. Because of the unique displaying feature, the background area of LCD is highly sensitive to uneven illumination. So in some meter gray-scale images the gray level of the background area is very close to the foreground area. If we binaries with the simple global binarization method, the low gray level background area would be classified as the foreground area. In this paper, we present a novel preprocessing algorithm for digital meter value recognition. This method could obviously eliminate the interfering of the external uneven illumination to the digital meter image.

The rest of this paper is organized as follows: section Ⅱ introduce the proposed enhanced Homomorphic filtering algorithm. Section Ⅲ briefly describes the binarization method, the rotating correction method and the segmentation algorithm. The experimental results are shown in section Ⅳ and conclusions are given in section Ⅴ.

II. ENHANCED HOMOMORPHIC FILTERING

A. Illumination-reflectance Image Model.

According to the illumination-reflectance image model theory, an image of a certain object is usually formed by the light illuminating the object and the light reflected by

Corresponding author: LI Xianyi(xyli@szu.edu.cn)

the object. These two factors are called illumination component and reflectance component, separately[3].

The nature of illumination component is determined by the properties of light source, the value of illumination component is non-zeros and infinite. The reflectance component is determined by the optical properties of the object, the value of illumination was restricted between 0 and 1. Illumination component $i(x,y)$ and reflectance component $r(x,y)$ are forming the resulting image $f(x,y)$ of an object by a multiplicative relationship[4]:

$$f(x,y) = i(x,y) * r(x,y) \quad \cdots\cdots(1)$$

B. Homomorphic filtering

The homomorphic filter is an approach in the frequency domain based on the illumination-reflectance image model. It is thought that the illumination component has slow spatial variation, which is characterized as the low frequency component, and the reflectance component usually arouse the sudden variation in the spatial domain, which represents the higher frequency component.

$$F\{f(x,y)\} = F\{i(x,y)\} * F\{r(x,y)\} \quad \cdots\cdots(2)$$

As shown in (2), in the two dimensional Fourier transform of the image, the illumination component and the reflectance component could not be separated directly. But if we calculate the natural logarithm of the image before the Fourier transform, this problem could be solved:

$$F\{\ln(f(x,y))\} = F\{\ln(i(x,y))\} + F\{\ln(r(x,y))\} \quad \cdots\cdots(3)$$

After that the image could be operated with different frequency domain treatment by the filter H:

$$Z(u,v) = F_i(u,v)H(u,v) + F_r(u,v)H(u,v) \quad \cdots\cdots(4)$$

Where Z, F_i and F_r are the Fourier transform of $\ln f$, $\ln i$ and $\ln r$, and (u, v) is the coordinates in the frequency domain.

If high-pass filter is used, the reflectance component would be preserved and the illumination component would be eliminated. After inverse Fourier transform and exponential transform, we could get the final result.

$$g(x,y) = e^{z(x,y)} = e^{F^{-1}\{Z(u,v)\}} \quad \cdots\cdots(5)$$

The expression of an improved Gaussian high-pass filter used for homomorphic filtering is shown in (6):

$$H(u,v) = (\gamma_H - \gamma_L)[1 - e^{-(D^2(u,v)/D_0^2)}] + \gamma_L \quad \cdots\cdots(6)$$

Where γ_H and γ_L determine the maximum and minimum value of the filter, respectively. D_0 is the image center in the frequency domain, $D(u,v)$ is the distance between coordinate (u, v) to D_0.

C. Enhenced Homomorphic filtering

As mentioned above, the reflectance component usually describes the detail of the object in image and the illumination component represent the illumination of external light resource. So the homomorphic filtering based on high-pass filter could be used to eliminate the uneven illumination.

But when applying the traditional homomorphic filter to the digital meter image, the attenuation of low-frequency component of the image would greatly decrease the light intensity of the background. The contrast between the background and foreground would be weaken. In order to overcome this problem, we proposed the enhanced homomorphic filtering method. The basic thought of the enhancement is that finding a way to avoid or compensate the decrease of the light intensity which is appeared after the homomorphic filtering.

The enhanced homomorphic filtering approach is described below:

Step 1: Convert the grayscale image $f(x,y)$ from 0~255 to 0~1;

Step 2: Add 1 to every pixels of the image:

$$f(x,y) = f(x,y) + 1 \quad \cdots\cdots(7)$$

Step 3: Implement the homomorphic filtering with (3) and (4) and the parameter γ_H and γ_L, separately, are 4 and 0.5 in this paper;

Step 4: Add a parameter to the inverse Fourier transform result:

$$z(x,y) = F^{-1}\{Z(u,v)\} + k * \mu \quad \cdots\cdots(8)$$

Where μ is the mean value of the imputed image and k is a constant factor which is -0.009 in this paper.

Step 5: Get the final result $g(x,y)$ after calculating the exponential result of $z(x,y)$.

III. BINARIZATION, ROTATING CORRECTION & SEGMENTATION

A. Binarization

After being filtered, the image needs to be binarized. As is known to all, there are two way to get the threshold for binarization: local threshold method and global threshold method. The Otsu method is a popular global threshold method. It classified the image into two classes: the target and the background.

Suppose the gray-scale range of the image is {0,1,……, L-1}, and threshold is T. The mean of whole image is μ. The pixel amount of the target is $\omega_0(T)$, the mean of the target is $\mu_0(T)$. The pixel amount and the

mean of the background are $\omega_1(T)$ and $\sigma_1(T)$. The formula of variance between the two clusters is as follows:

$$\sigma_b(\mathrm{T})=\omega_0(\mathrm{T})(\mu_0(\mathrm{T})-\mu)^2+\omega_0(\mathrm{T})(\mu_0(\mathrm{T})-\mu)^2 \quad \cdots\cdots(9)$$

When $\sigma_b(T)$ is arriving the maximum value, the optimal threshold is gotten.

B. Rotating Correction

The Cartesian moment is a kind of invariant moment. The moment features of the Cartesian moment is unchanged when the object is translated, rotated, scaled[5]. It could calculate the angle of rotation automatically, so the Cartesian moment is selected as our rotating correction method.

Suppose $f(x, y)$ is the gray distribution density function of an image. The two-dimensional (p + q)th order Cartesian moments of a density distribution function $f(x, y)$ are defined in terms of Riemann integrals as:

$$m_{pq} = \sum_{y=0}^{M-1}\sum_{x=0}^{N-1} x^p y^q f(x, y) \quad \cdots\cdots(10)$$

Where M is the height and N is the width of the image. One order Cartesian moment represents the centroid of the image.

$$\bar{x} = \frac{m_{10}}{m_{00}} \quad \bar{y} = \frac{m_{01}}{m_{00}} \quad \cdots\cdots(11)$$

The central moments of an image as follow:

$$\mu_{pq} = \sum_{y=0}^{M-1}\sum_{x=0}^{N-1} (x-\bar{x})^p (y-\bar{y})^q f(x, y) \quad \cdots\cdots(12)$$

Two order central moments represents the direction feature of image. Assuming the maximal and minimal two order moments of the image as the principal axes, and then the direction of the principal axes can denote as follow:

$$\varphi = \frac{1}{2}\tan^{-1}\left(\frac{2\mu_{11}}{\mu_{20}-\mu_{02}}\right) \quad \cdots\cdots(13)$$

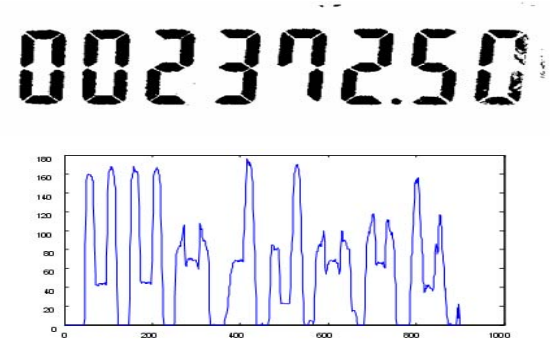


Figure 3. The rotated image and it's VPM

C. Segmentation

Segmentation is the last preprocessing steps. The vertical projection map(VPM) is applied in this step. The VPM could describe the texture distribution of the image. Through counting black pixels of the image in the vertical direction, the VPM, shown in Fig.3, would appear one or two peaks in the character area and appear a valley in the blank area. So according to the distribution of the peaks and the valleys, the digital image could be divided into several sub-blocks and each block only contains a single character.

IV. Experimental Results

In this section, we present a digital meter image as the experimental example and illustrate the entire preprocessing.

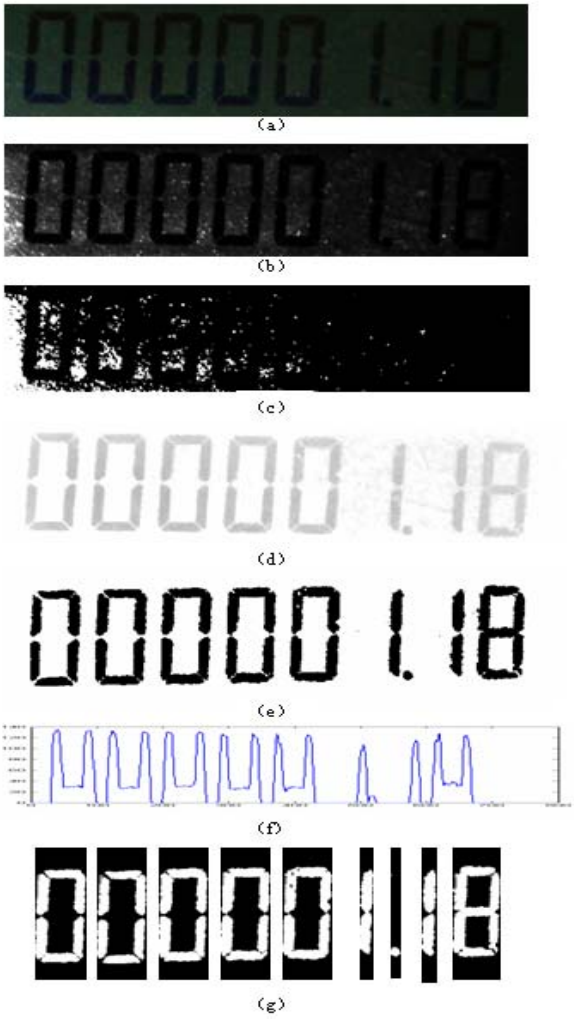


Figure 4. The test image and the results of each step of the prerprocessing

As shown in Fig.4, Fig.4(a) is the original gray-scale image. Because of the external illumination, the left side of the image is darker than the right side; Fig.4(b)and Fig.4(c) are results of traditional homomorphic filtering approach and Otsu binarization method, after homomorphic filtering, the intensity of the background is

so close to the foreground that the binarization method could not provide a reasonable threshold. Fig.4(d) and Fig.4(e) is the result of proposed enhanced homomorphic filtering method and its binarization result. We can see that the proposed method compensate the decrease of the intensity of the background, so the foreground and background are classified correctly. Fig4.(f) is the rotating correction result with the Cartesian moment. Fig.4(f) and Fig.4(g) are the vertical projection map and the segmentation result of the image, respectively.

V. CONCLUSION

In this paper, we present a novel preprocessing approach for digital meter reading based on computer vision. The proposed preprocessing approach includes enhanced homomorphic filtering, binarization, rotating correction and segmentation. The proposed enhanced homomorphic filtering algorithm could eliminate the interference of the uneven illumination and avoid the decrease of intensity of the background area. The Cartesian moments provides a simple and accurate rotating correction method. As a result, we believe that our method is an attractive alternative to currently available methods for digital meter image preprocessing.

ACKNOWLEDGMENT

This work was supported in part by the integration project of production teaching and research by Guangdong Province and ministry of education (No:2009B090300267),
the Foundation for the Innovation Group of Shenzhen University (Grant: 000133) ,the National Natural Science Foundation of China (Grant No.60972037) and the Foundation for Research of Shenzhen University (No: 200736)..

REFERENCE

[1] HaiBo Zhang, HuiChuan Duan, ShuFu Xie etc, “A Preprocessing Algorithm For Meter Display Value Recognition,” *Application Research Of Computers*, 2005, issue.2, pp.240-242.

[2] Shutao Zhao, Baoshu Li, Jinsha Yuan, Guiyan Cui, “Research on Remote Meter Automatic Reading Based on Computer Vision,” Transmission and Distribution Conference and Exhibition: Asia and Pacific, 2005 IEEE/PES, China, 2005, pp.1-4.

[3] Holger G. Adelmann, “Butterworth equations for homomorphic filtering of images,” *Computers in Biology and Medicine*, 1998,Vol.28, issue.2,pp.169-181.

[4] R.C. Gonzales, R.E. Woods, Digital Image Processing, Addison-Wesley, Reading, Massachusetts, 1992.

[5] Ming-Kuei Hu, “Visual Pattern Recognition by Moment Invariants,” *IRE Transactions on Information Theory*, 1962,Vol.8,issue.2,pp.179-

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 312-314

Analyze and model to chirped fiber grating with new apodization function

Yingli Yang[1] and Guodong Wang[2]
[1]School of Computer Science and Technology,Henan Polytechnic University, Jiaozuo , Henan, China
Yangyingli@hpu.edu.cn
[2] School of Electrical Engineering and Automation Henan Polytechnic University, Jiaozuo, Henan, China
wgd@hpu.edu.cn

Abstract—**The chirped fiber grating was modeled with new apodization function. The reflection characteristics of the grating was analyzed. When the new apodization is applied to it, the fiber grating exhibits a flattop spectrum with steep edges and high reflectivity. The bandwidth utilization defined as the ratio of -1:-30 could be achieved a larger value(>0.90).**

***Index terms*—fiber grating; sinusoidal apodization function;linearly chirped**

I. INTRODUCTION

Fiber Bragg grating (FBG) is a passive component which is easy to produce, low-cost and of superior performance. It has developed into a critical component for many applications in optical communication systems and optical sensors systems [1-6]. FBG can be used as optical filters [7-8], gain-flattening filters [9], feedback mirrors in fiber lasers [10] and dispersion compensator [11-12].

Fiber gratings with ideal box spectra are rapidly becoming critical apparatus in dense wavelength division multiplexed communications system. In order to achieve high bandwidth utilization, several methods are presented. Through a periodic sinusoidal modulation of the refractive index profile in fiber Bragg gratings, Ibsen[13] reported gratings with multiple equally spaced and identical wavelength channels. Based the design of a grating period variation adapted to apodization function, Carballar[14] obtained the ideal box spectrum. Sinusoidal chirps of grating periods are introduced by Zhang to improve their performance as dispersion compensators and multi-channel filters [15]. Based on the outer cladding being etched as hyperbolic function, the reflection spectra of fiber gratings will be steep edges, flattop, high reflectivity and low side lobe when the grating is held under the tension [16].

In this paper the linearly chirped fiber Bragg grating with a new apodization function is proposed and numerically characterized. The new apodization function various along the z axis is not humdrum and the reflection spectra of grating is steep edges, flattop, high reflectivity and low side lobe..

II. DISCUSS AND RESULTS

The new apodization function various along the z axis can be expressed as

$$f(z) = 1 - g + g\sin\left(\frac{2\pi z}{L}\right) \tag{1}$$

Where L is the grating length and g is apodization factor, $0 \le g \le 0.5$.

The refractive index profile along the propagation direction (z) that originates the fiber grating perturbation can be described by

$$n(z) = n_0 + \delta n(z)\left[1 + v\cos\left(\frac{2\pi}{\Lambda} z\right)\right] \tag{2}$$

where n_0 is the refractive index of the fiber core, v is the fringe visibility, $\delta n(z) = \delta n f(z)$ and Λ is the grating period

$$\Lambda = \Lambda_0(1 + c_0 z) \tag{3}$$

Where Λ_0 is the initial grating period and c_0 is the linearly chirped modulus.

In this paper we take the method of piecewise uniform approach to calculate the reflection spectra. This method is simple to implement, almost always sufficiently, accurate, and generally the fastest.

Firstly, we calculated the reflection spectrum of the grating with the grating parameters are: $L = 9.8\text{cm}$, $\delta n = 0.003$, $\Lambda_0 = 0.530\mu m$, $c_0 = 2.91nm/cm$ and the apodization factor $g = 0.3$. The reflection spectrum is illustrated in Fig.1. The

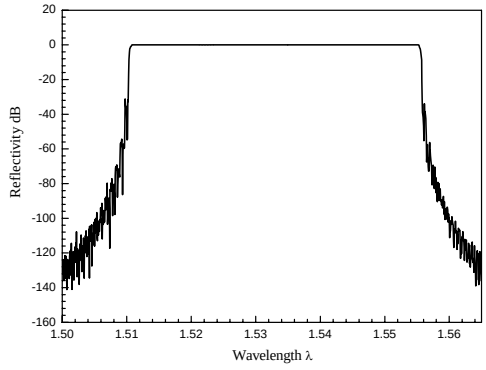


Fig. 1 Reflection spectrum of the fiber grating with new apodization

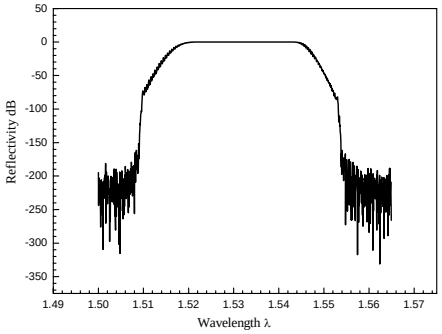


Fig. 2 Reflection spectrum of the fiber grating with Gauss apodization

reflection spectrum of this grating approximates an ideal box spectrum, whose side lobe is less than -30dB. The parameter that evaluates the spectrum's steepness at the band edges is BWU, which was defined as the ratio of the bandwidths at -1 and -30dB [15]. The larger the BWU is, the steeper the reflection spectrum is. For comparison we also calculate the reflection spectrum of grating with Gauss apodization function, which is illustrated in Fig.2. By comparing we note that the BWU of the grating with new apodization is larger than that with Gauss apodization. The BWU of the former is 0.9907 and the BWU of the latteris only 0.6307.

Secondly, we computed the value of BWU when the apodization factor takes different value, which is illustrated in Fig.3. The parameters of the fiber Bragg grating are $\Lambda = 0.5357$, $L = 8.1cm$, $\delta n = 0.0002$ and $c_0 = 1.6nm/cm$. This figure shows that the value of BWU will decreased smaller with the apodization factor g creasing when g is changed near the zero. However, when g is changed near the value 0.5, the changed value of BWU is great with the value of g creasing. This is can be explained that: there is a phase shift in the fiber grating when the g is near the value of 0.5. So, a transmission extent will appear in the reflection spectrum and the value of BWU will decreased rapidly, which is illustrated in Fig.4.

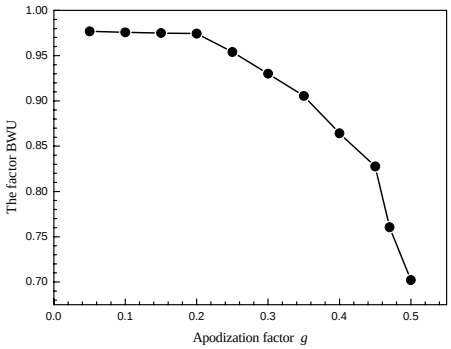


Fig. 3 The BWU factor as a function of the apodizatoin factor *g*

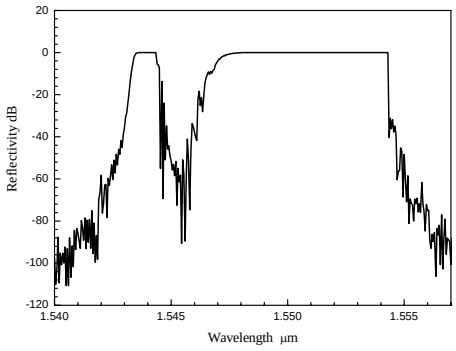


Fig. 4 Reflection spectrum of the fiber grating with *g*=0.48

III CONCLUSION

As a conclusion, a new apodization function is introduced to improve the performance of chirped fiber Bragg gratings. When the new apodization is applied to it, the fiber grating exhibits a flattop spectrum with steep edges and high reflectivity. The bandwidth utilization defined as the ratio of -1:-30 could be achieved a larger value(>0.90).

ACKNOWLEDGMENTS

This work is supported by the Open Foundation for Henan Provincial Open laboratory for Control Engineering Key Disciplines, China (No. KG2009-16) and by the Doctor Foundation for Henan Polytechnic University, China (No. 648393).

REFERENCES

[1] Oiwa Masaki, Minami Shunsuke and Tsuji Kenichiro, et al. Influence of nonideal chirped fiber Bragg grating characteristics on all-optical clock recovery based on the temporal Talbot effect. Applied Optics, 2009, 48(4):679-690.

[2] C. L. Lee, Y. Lai, Optimal narrowband dispersion-less fiber Bragg grating filters with short grating length and smooth dispersion profile, Opt. Commun. 235(2004):99-106.

[3] C. Caucheteur, F. Lhomme, K. Chah, M. Blondel, P. Megret, Fiber Bragg grating sensor demodulation technique by synthesis of grating parameters from its reflection spectrum, Opt. Commun. 240(2004):329-336.

[4] Y. W. Lee, I. Yoon, B. Lee, A simple fiber-optic current sensor using a long period fiber grating inscribed on a polarization-maintaining fiber as a sensor demodulator, Sensors Actuators A 112(2004): 308-312.

[5] Z. C. Li, H. Y. Tam, L. X. Xu, Q. J. Zhang, Fabrication of long-period gratings in poly(methyl methacrylate-co-methyl vinyl ketone-cobenzyl methacrylate)-core polymetoptical fiber by use of a mercury lamp, Opt. Lett. 30(2005): 1117-1119.

[6] M. Kulishov, J. M. Laniel, N. Belanger, D. V. Plant, Trapping light in a ring resonator using a grating-assisted

coupler with asymmetric transmission, Opt Express 13 (2005):3567-3568.

[7] T. Erdogan, Fiber grating spectra, J. Lightwave Technol. 15(1997)1277-1294.

[8] T. Erdogan, Cladding-mode resonances in short- and long-period fiber grating filters, Opt. Soc. Am. A 14(1997) 1760-1773.

[9] A. E. Lobo, C. M. de Sterke, J. A. Besley, N-fold symmetric grating as gain-flattening filters, J. lightwave Technology. 23 (2005): 1441-1448.

[10] H. Zhou, G. Xia, Y. Fan, T. Deng, Z. Wu, Output characteristics of weak-coupling fiber grating external cavity semiconductor laser, Opto-Electron. Rev. 13(2005):27-30.

[11] P. Li, T. G. Ning, T. J. Li, X. W. Dong, S. S. Jian, Studies on the dispersion compensation of fiber Bragg grating in high-speed optical communication system, Acta Phys. Sin. 54(2005):1630-1635.

[12] J. Kwon, Y. Jeon, B. Lee, Tunable dispersion compensation with fixed center wavelength and bandwidth using a side-polished linearly chirped fiber Bragg grating, Opt. Fiber Technol. 11(2005): 159-166.

[13] M. Ibsen, M. K. Durkin, M. J. cole, R. I. Laming, Sinc-sampled fiber Bragg gratings for identical multiple wavelength operation, IEEE Photon. Technol. Lett. 10(1998):842-844.

[14] A. Carballar, M. A. Muriel, J. Azana, Fiber grating filter for WDM systems: an improved design, IEEE Photon. Technol. Lett. 11(1999): 694-696.

[15] L. Zhang, C. X. Yang, Improving the performance of fiber gratings with sinusoidal chirps, Appl. Opt. 42(2003).

[16] Guo-dong Wang, Cai-Xia Liu, Dong-Ming Sun, Wen-Bing Guo and Wei-You Chen. Improving the performance of fiber gratings with cladding being etched as hyperbolic function. Optik, 117(2006): 477-480.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 315-318

The research of Wireless UWB Intrusion Detection

Yang Bo[1], Shen yu-bin[2]
[1]School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, China
Email: jzyangbo@hpu.edu.cn
[2]Institute of Information Engineering, Jiaozuo University, Jiaozuo, China
Email:shenyb616@163.com

***Abstract*—Based on former related studies of intrusion detection, constructing intrusion detection system according to UWB area network denial-of-service attack, and use network nodes to build intelligent network according to UWB network denial-of-service attack, at the MAC layer; and design UWB network denial-of-service attack defense system. Use the collaboration between them to complete detection of intrusion behavior. Use multi-agent technology as intrusion detection engine, which is based on network, and forecast the likelihood of attack, and finally complete the work of testing safety protection.**

***Index Terms*—UWB; agen; area network; Channel attacks**

I. INTRODUCTION

The birth of Ultra-bandwidth (UWB) technology realizes the UWB within short distances, and high-speed data transmission. It will bring low power consumption and Ultra-bandwidth with relatively simple wireless communication technology, to wireless LAN and personal area network PAN interface and access technology.

UWB can transmit information with very high data rate (such as 480Mbit/s) and very low power(such as200μW), within limit, and that is much better than the bluetooth. The data rate of Bluetooth is 1 Mbit/s, and its power is 1mW. In wireless area network application, a manipulator can recognize and communicate with each other with limited space, computer, printer, PDA and camer, which will eliminate mixed wiring in office. In personal space, all sorts of intelligent wireless devices with optional increase or decrease, communicate in the air. [1] UWB can provide fast wireless peripherals visit to transmit photos, documents, and videos. At the same time, through UWB, you may conveniently download, at home and office, the content in video cameras to PC for edit, and then send them toTV for browse, wirelessly; and easily realize personal digital assistant (PDA), phones and PC data synchronization, load game and audio/video files to PDA; and delivery of audio files between MP3 player and multimedia PC, etc. [2]

In UWB network, there are no fixed infrastructures like base station or mobile exchange center. Those mobile nodes can realize mutual communication through wireless connection, and for those distant nodes they can rely on other nodes as routing to forward messages. Because of its open properties, wireless network is very easily to be attacked [3]. The attacker using hardware and software of the wireless network to access network information, which is a big threat to information security of the units and departments; and preventing UWB area network intrusion behavior is a new field of network security.

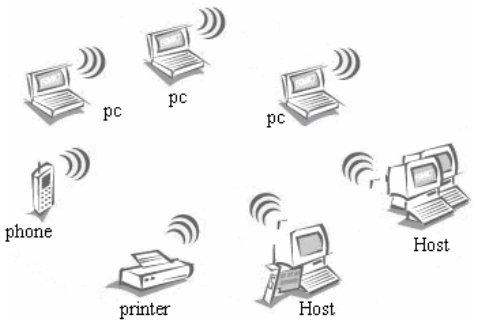


Figure 1. UWB domain net

II. SECURITY OF UWB PERSONAL AREA NETWORK

A. *The network security questiones of UWB mainly include:*

1) The information in UWB personal area network can transport through the open radio channel of limited bandwidth, and the intruder can directly monitor information without visit to physical link; UWB network has no control center, and malicious nodes can join the network easily and know the position of the node in the target routing by sending routing and location information to unauthorized node in network; mobile nodes that lack safety protection mechanism are easy to be captured [4].

2) Mobile node can move freely without being restricted, and in the UWB network routing protocol, mobile nodes establish communication between the source node and destination node through exchanging network topology information, and there is no distinct difference between normal nodes and malicious nodes, and malicious node can enter the network. through false routing information.

3) UWB network has no fixed infrastructure,and malicious nodes can eavesdrop and modify business in wireless channel by disguised as a normal node.

The inherent characteristics of UWB network and security vulnerabilities will cause more internal network attack. And the attacks can be divided into initiative attacks and negative attacks. Negative attacks do not interfere with any service, its purpose is to steal information. initiative attack is to change communication data actively, aimed at increasing burden to network, damage the operation or make this node lose contact with its neighbouring nodes to make it cannot use network service effectively.

B. Channel attacks

Channel attack is a serious attack against UWB network routing protocols, especially for those defensive routing protocols, and it build a private channel between two malicious nodes; the attacker records data or position information and transfer steal information to another location through this private channel.

Since malicious nodes connect through a private network rather than a normal one, it is also called as channel attack. Because the distance of private channel is longer than single wireless transmission range, so the packet can arrive goal node earlier by private channel than by normal multi-path transmission. Now it seems that channel is beneficial rather than harmful because it can make packet arrive goal node as soon as possible. However, if the channel attacker not transfer all packets faithfully, but intentionally transfer parts of them, such as control information packets, or doctor content in packets, then it will cause the packet loss or destruction of the packet. At the same time as channel can cause false path shorter than the actual path, so will disturb routing mechanism rely on distance information between the nodes, resulting failure of routing discovery.

Channel is very difficult to test because the path it used to convey information is not usually part of the actual network; and it is also very dangerous, for they can make damage without knowing used agreement or service provided by network .

Figure 2 shows how channel attacks. Among them, S means the source node, W means destination node, M, N means malicious node, and A, B say intermediate node. When node M receives RREQ , channel will deliver it to node N. When the node N receives RREQ, it deliver RREQ node W.It seems that the delivery of packets through node s, M and N. Node N deliver RREP the same way to node M through channel. So, node M, N false claim that there exist a path between them, thus deceive node S choose path M, N (because it is the shortest path),. The channel speed between attackers should be faster than that of the reasonable nodes, so the speed of deliver packets through channel is faster than other paths. If the attacker using this channel fairiy and reliable it will not cause harm, the attacker is actually provide a more effective way of Internet connection. However, if the channel attacker does not faithfully deliver all packets, but intentionally deliver parts, such as only transfer control information packets or doctor content of the packets, then will cause loss or destruction of the packets. At the same time as channel can cause false path which is shorter than actual path, so it will disturb routing mechanism which rely on distance information between nodes, and cause failure of discovery process of routing.

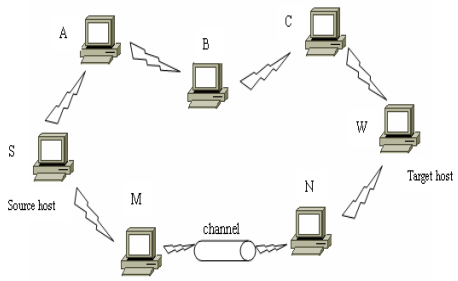


Figure 2. Channel attacks

III. UWB NETWORK MAC LAYER DENIAL-OF-SERVICE ATTACK

Against features of UWB network MAC layer denial-of-service attack, construct a intrusion detection system based on neural network and agent, which build nodes by each network node, and use the collaboration between them to complete intrusion detection. According to abnormal behavior or abnormal resources situation of UWB equipment judge whether attack activities occur. The key of anomaly detection lies in how agents exist as complete intrusion detection system.

In UWB network there are two types of denial of service attacks: UWBMAC layer attack and UWB network layer attack. There are mainly two methods to implement denial of service attacks in UWBMAC layer:

1) Block wireless UWB channel used by goal node equipment in UWB network,and cause goal node equipment in UWB network out of use.

2) Use target nodes in UWB network as a bridge and let it continuously relay invalid data frames to exhaust available resources of goal node equipment in UWB network.

In one UWB area network, lots of agents can be constructed, and agents can get data directly or through filter; Each agent in physical equipment will deliver the cases they find and related data to transceiver. Each physical equipment has only one transceiver in operation, and it monitor all operation conditions of agents in the machine, including start and stop, and it can convey configuration orders to agents, and can also process data from agent; The transceiver will report data summaried by itself to one or more monitors. Each monitor manage operations of lots of transceivers, and monitor can see data within the whole network, so it can carry out high-level related checks, and then detect intrusions related to several machines; it also can organize monitors according to level, which means some monitors will report to upper monitor or one transceiver report to lots of monitors , so can provide data redundancy and avoid single point fault. But no matter how to divide the levels,finally,there must be a monitor to provide information to the end user by using user

interface, and provide users with a control interface through thisinterface.

IV. DESCRIPTION OF RESOLUTION ALGORITHM OF ATTACK DETECTION

Restricted detection should be made in connection stage, and avoid the occurrence of attack from the source. This paper designs the following detecting steps against UWBMAC layer denial-of-service attack in connection stage.

Step1: In initialization phase of the system, mainly execute a series of routine examination, reading system security logs, etc, and then begin to monitor each node in UWB network, and start registration and analysis of security log.

Step2: When a new connection requests occur, the system will check whether the current system resources are available. System resources refers to indexes to ensurethe system operating safely, including identity of system user, competence, Local space, memory space condition, and the protocol configuration, etc. If the current system resources are available, then make the next step, otherwise, to step 13.

Step3: Detect if users are authorized. If the user is not authorized, continue to the next step, otherwise, jump to step 15.

Step4: Detect if it is first connection requests. If it is the first connection requests, continue to the next step, otherwise, jump to step15.

Step5: Setting address signs for the target address

Step6: Acquiring related connection information, Including the identity of the nodes in the UWB network, ask for connection type, use agreement, and UWBMAC address, all key information related to the connection.

Step7: Judge whether this MAC address is in agent information sheet.

Step8: Check whether warnings have been sent, if it have been sent then update threshold of this address in the address sheet, and removed attack number increase. At the same time, jump to step 4.

Step9: Read feature model and parameter thresholds from the eigenvalue model and parameter threshold database. Feature model and threshold parameter library store thresholds of each parameter that connected with connection (such as limit number of connection requests from or arrive mobile node of MAC address, etc), and summary feature model of denial of service attack according to former transfer mode. This feature model has learning function, can absorb recent transfer mode s occurred in system into feature model, according to certain algorithms, so as to promote ferfection and reason of feature model.

Step10: With time interval, monitoring neighboring data packets of visit node.

Step11: Detect whether operation condition of current system match the mode. If match, then update agent data, otherwise, proceed to the next step.

Step12: Detect whether parameters reach threshold of system resources. If reach threshold, then update eigenvalue model and data in parameter threshold database and agent data, at the same time, jump to step 13. Otherwise, allow connection to new request.

Step13: Refuse new connection requests or send warning. Refuse link or send warning can be handled differently according to the severity of the threat, and also can give the final decision to the user.

Step14: Use the new trust modify neighbor trust list and normalization

Step15: Finding out in all agents whether having records according to user and the previous transfer mode. If there is records, then step to 16, and if it is a bad record jump to step 13. If there is no records, jump to step 6.

Step16: Detect feature model and schema matching of threshold parameter library, if do not match, then jump to step 6. Otherwise, send alarm.

UWBMAC layer's performance of denial-of-service attack defense system be measured by the following aspects:

3) The foresight ability: because the operating environment of old network U field is bad,plus bandwidth resource limitation,etc,which make this network structure relatively weaker than other fixed network. Therefore, stop denial of service-attack behavior of consumption channel information resources, as soon as possible. This requires testing system has certain predictability, identify potential danger of denial of service attacks and make decisions timely.

4) Effectiveness:it contains two aspects: give timely warning about denial of service attack,and reduce misstatement about normal service behavior as far as possible.In view of the former's harm to the entire network,high requirements are needed. Efficiency is an important index in measuring the performance of testing system,and is also an important basis in measuring whether the Algorithm of the testing system is scientific.

5) Requirements to the system: due to the defense system is running online, and equipments in UWBnetwork is mostly portable equipments, and it has only limited resources, thus need the detection system reduce requirements of system resources (mainly memory) ,and don't make it a burden to system operation, after all, denial of service attack behavior doesn't happen often.

V. CONCLUSION

Simulation results show that the memory occupancy rate of UWBsystem significantly reduces as the time interval increases, until finally stable. This is because the defense system has not enough historical data at beginning, and always in study phase, and the defense system need to analyze and calculate each link behavior, in a large amount, and continuously amend characteristic mode value, which require a lot CPU memory; with the increase of time interval, defense system can reduce or stop observations of normal service behavior, and data accumulated by the system will become more and more. At this time, for the vast

majority of connection, judgment can be made rely on previous knowledge, and the use of memory by defense system will stabilize.

REFERENCES

[1] Moe2Win, Robert A ScholtZ.Impulse Radio: How It Works.IEEECommunication Letters, 1998, 2(2):36–38P[2] XiaominChen, ayfe kiaei.Monocycle Shapes for Ultra Wideband System. IEEE ISCAS}2002, (l):597–600P

[2] XiaominChen, Sayfe kiaei.Monocycle Shapes for Ultra Wideband Systetn.IEEE ISCAS}2002, l(l):597–600P

[3] Stubble field A, Ioannidis J, Rubin A D. Using the Fluhrer, Mantin, and Shamir attack to break WEP[R]. AT&T Labs Technical Report TD-4ZCPZZ, Revision 2, 2001.

[4] Xiaojing Huang, Yunxin Li.Generating Near-WhiteUltra-Wideband Signais with Period PN Sequenees.Conference Proceesings of the IEEEVTC–2001, Rhodes, Greece.2001(5):1184–1188P

[5] LinManQi, QianHuaLin.Distributed denial of service attack: principles and strategies.Computer science[J], 2000, 27(12):41~45.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 319-321

Optimization and Simulation of Wireless Sensor Networks Routing Algorithm Based on ZigBee

Lu Yongfang [1], Li Haitao [2]
[1] Department of Mechanical and Electrical Engineering, Jiaozuo university, Jiaozuo, China
Email: zlxlyf@163.com
[2] School of Electrical Engineering and Automation, Henan Polytechnic University, Jiaozuo, China
[3]Department of Basic science ,Jiaozuo university, Jiaozuo, China
Email: dwxba2002@sina.com

***Abstract*—The reliability and sustainability of data transmission are key indicators in wireless sensor networks (WSNs) . ZigBee wireless sensor networks are composed of many sensor to the organization form of Ad Hoc - wireless networks, Integrated sensor technology, embedded technology, distributed information processing technology and so on. Widely applied in automatic control and remote control.This paper introduces ZigBee wireless sensor network, including the network configuration, data structure and performance characteristics of the router, It detailed analysis the common tree routing algorithms of ZigBee, which is common used,as well as the advantages and disadvantages.Then a routing optimization algorithm which is based on the routing table and neighbors table used alternately is proposed.It is achieved by the way of the improvement of ZigBee wireless sensor network routing protocols as well as add related command frame and other channels. Simulation results show that, compared with the original algorithm routing algorithms, the improved routing algorithm can reduce by 20% -30%, effectively reduce the cost, and greatly improve the reliability and sustainability of the routing ZigBee wireless sensor networks.**

***Index Terms*—Zigbee; wireless sensor networks(WSNs); Routing alorthms; Data structure**

I. INTRODUCTION

With the rapid development of information technology and the continuous improvement of industrial automation, industrial, home automation and industrial telemetry remote areas of the growing demand for wireless data communication strong, particularly in the industrial field of wireless data transmission reliability, stability , power consumption, performance requirements also increase. Wireless sensor network is in the industrial control, medical care, transportation monitoring, smart home and so on, by the deployment of a large number of micro-region in the monitoring of sensor nodes through wireless communication form a multi-hop's self-organizing network system to carry out cooperative sensing, acquisition and processing of network coverage area where the object being monitored information, send observers. ZigBee wireless sensor network is composed of many sensors to form self-organizing form Ad-Hoc wireless network, which combines sensor technology, embedded technology, distributed information processing technology and ZigBee technology. Zigbee IEEE 802.15.4 protocol is synonymous with the technology under this agreement is a short distance, low-power wireless communication technology, is a range of wireless tag technology and emerging communication between Bluetooth technology, Can be embedded in a variety of devices, with high communication efficiency, low complexity, low power, low rate, low capital, high-quiet nature and all-digital, and many other

ZigBee-based wireless devices WSNs includes two, full function device (FFD) and reduced functionality device (RFD). FFD can FFD, RFD Communications, FFD can not only send and receive data, but also with routing functions; the RFD and FFD can communicate, RFD is not directly between the communicating. with a perfect combination, are widely used in the field of automatic control and remote control. advantages. It is because of these advantages to promote ZigBee wireless sensor networks.

II. ZIGBEE NET WORK OF PHYSICAL DEVICES

A. ZigBee wireless sensor networks logic devices

There are usually in ZigBee networks are 3 types of logic devices (network nodes): Coordinator (coordinate points), router (routing nodes) and terminals (terminal node). Coordinator (focal point) is the main controller of the entire network must be a FFD, responsible for initiating the establishment of the new network, beacons sent network management network nodes and storage nodes in the network information; router (routing nodes) are usually involved in route discovery, message forwarding, by connecting to other nodes to extend network coverage, etc., must also be FFD. Terminal equipment (terminal node) through the focal point or zigBee ZigBee routing nodes connected to the network, but does not allow any other node through which joined the network, can be FFD or RFD.

B. ZigBee network topology

IEEE802.15.4/ZigBee agreements usually three topologies: joint topology (Star), cluster structure (Cluster-tree), and network structure (Mesh), which cluster node. Structure (Clustertree), and mesh structure (Mesh) belong to the point to point topology.

III. ZIGBEE WIRELESS SENSOR NETWORKS DATA STRUCTURE

A. Routing Table

ZigBee coordinator and router nodes are stored with a routing table to forward packets for other nodes in the network to save a routing table entry. ZigBee coordinator and routers can maintain the routing table.

B. Routing table

If the router or ZigBee coordinator maintains a routing table, it should also maintain a routing table, routing table entry is the long-standing and unchanged, while the routing table entry in the routing process only exists and can be regenerated.

C. Neighbor Table

ZigBee network, each node save a neighbor table to store the other nodes within the transmission range of node information, see Table 2. Equipment each received equipment from the corresponding neighbors of any frame, its corresponding entry should be updated.

IV. ZIGBEE WIRELESS SENSOR NETWORK ROUTING

A. ZigBee wireless sensor network routing algorithm commonly used in the idea

The core algorithm is used to forward the data to find the destination address whether it is itself. If yes, it is no longer transmitted; If not, see if it came up from a valid path. The so-called effective path is from the father node or nodes come from the child, if a child node from the data, according to its destination to be reached, if it is their children node, then forwarded to the child node to the end, if not their own child node, the next to reach the address is the father node, and thus a level, along the tree until you find the destination address。

B. ZigBee wireless sensor network routing algorithm commonly used in analysis

a) In the ZigBee wireless sensor network, when the coordinator to establish a new network, it will give its own distribution network address ox000o, network depth DePth == 0. , If node (i) you want to join the network, and the node (k) connection, then the node (k) will be referred to as node (i) the parent node. Ak according to its address and network depth Depthk, node (k) for the node (i) distribution networks and network address Ai depth Depthi = Depthk + l. Network depth that only the father-son relationship with the network, a transmission frame transmitted to the ZigBee coordinator by passing the minimum number of hops. zigBee Coordinator own depth of 0, and its depth is a child device.

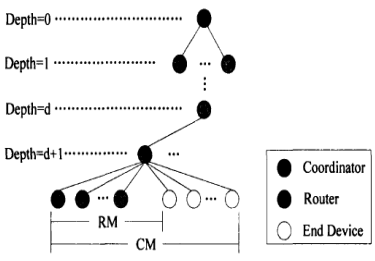


Figure 1: ZigBee tree structure

b) Figure 1 for the ZigBee tree. Parameters nwkMaxChildren (Cm), said router or coordinator in the network to allow the maximum number of devices have sub. Parameters nwkMaxRouterS (Rm), said sub-node maximum number of routers, and the remaining number of terminal equipment number.

c) A new RFD node (i), it does not have routing capabilities, it is connected with the co-ordination as the coordinator of the first n nodes. According to its depth d, the parent node (k) for sub-node (i) distribution network address:

$$A_i=A_k+Cskip（d）\cdot R_m+n \quad (1)$$

d) If a new child node FFD, it has routing capabilities, the parent node (k) would give it distribution network address:

$$Ai=Ak+l+Cskip(d)\cdot(n — l) \quad (2)$$

e) Suppose a router to the network address of the destination address for the D to send data packets, the router's network address is A, the network depth d. Router will be the first by the expression:

$$A<D<A+Cskir(d — l) \quad (3)$$

f) Determine whether the destination node as its child nodes. If the destination node is its child nodes, and the purpose of Node is a terminal device, the next hop node address D; If the destination node is not terminal, the next hop Node address:

$$N = A+1+\left[\frac{D-(A+1)}{Cskip(d)}\right]\times Cskip(d) \quad (4)$$

If the destination node is not its own child nodes, then the next hop node is the parent node of the router.

D. ZigBee wireless sensor network routing algorithm commonly used in the inadequacies

Advantages of this algorithm is: clever use of the distribution of each network node address obtained was characteristic of tree structure, to select the routing and equipment do not keep in memory a routing table, nor spent Wancheng find the path to the operating result network traffic significantly lower. Tree routing algorithm, but there are many deficiencies, as according to an address on the routing tree can not take the shortest path, the path than it actually take a long, easy to generate additional traffic, more prone to failure.

V. ZIGBEE WIRELESS SENSOR NETWORK ROUTING ALGORITHM OPTIMIZATION

E. ZigBee wireless sensor network routing algorithm optimization ideas

Optimization of the routing algorithm used in the ZigBee routing algorithm based on the alternate routing table and neighbor table, if the destination node in the source (relay) node in the neighbor table, then send it directly; fruit destination node for the source (relay) node descendants of the node, then in accordance with the original Cluster-Tree routing algorithm to select a child node sends; if you do not meet these two conditions, then

compare the source (relay) node and its neighbors, in addition to the table other than the parent and offspring routing node to the sink expenses, in accordance with the principle of least cost routing next hop selection. First, we calculated the source (relay) node to the destination node routing overhead and recorded as MinHop, next hop is recorded as its parent node. Then one by one parent and offspring, calculated outside the neighbor table to the sink node in routing overhead Hop-Coun, t if HopCount +1 <MinHop, makes MinHop = HopCount +1, and note the next hop for the current node, until neighbor nodes of all eligible calculation is completed, the last data packet sent to the selected next hop node.

F. ZigBee wireless sensor network routing optimization algorithm

a) Improved Routing Protocol.From the above analysis, ZigBee network layer routing algorithm, the default maximum transmit power and data routing discovery packet routing. Here we improve the routing protocol. In the improved ZigBee network, RN-node to other nodes in the network to send information, the optimal transmit power used to send data packets to the parent node, and then forward the data packet from the parent node; when the RN + node to other nodes in the network to send information , the first visit to the routing table to obtain the corresponding next hop address and the optimal transmit power, then the optimal transmit power down hop address to send packets.

b) 4.2.2 Improved data structures.Can improve the routing table data structure, design and LQI the command frame and add them to the protocol stack, and then nextHopAddress node received LQI asked the command frame, back to the source node QI response command frame, which contains the forward link the LQI value.

VI. MATLAB SIMULATION AND ANALYSIS

The experiment used a network simulator NS-2 platform [8], and provided the IEEE 802. 15. 4MAC layer and physical layer modules based on the simulation environment, the network size set to 100 m × 100m, the node transmission distance 20 m, the packet is 40 B, the rate of sending data packets 2 packets / s, business source use CBR. ZigBee network parameters set Cm = 4, Rm = 4, Lm = 5, neighbor size of the table set 5. As the network nodes randomly placed, topology changes may occur some node can not join the network, where only consider the existence of more than 75% of the nodes in the network situation. The simulation results shown in Figure 2:

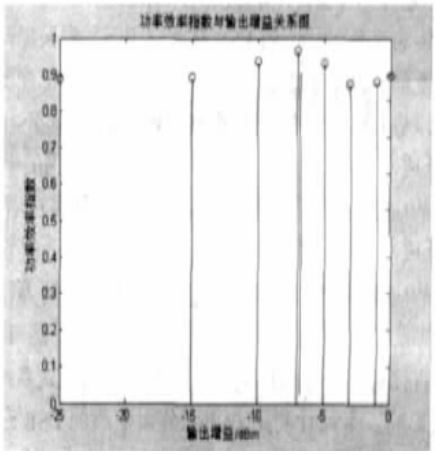


Figure 2: MATLAB simulation results

VII. CONCLUSION

Simulation results show that the improved routing algorithm than the original algorithm to reduce large routing overhead. When the destination node is the coordinator, the algorithm and the algorithm performance opportunities rather routing table, especially when the destination node randomly selected, using the improved routing algorithm can reduce the routing cost, to achieve the purpose of reducing power consumption, achieve storage efficiency and routing balance of performance, very suitable for storage space is limited and there is a high performance requirements of routing applications.

REFERENCES

[1] Li-Li Tong, an agreement based on zigBee technology development and platform design: (Master thesis). Wuhan: Shenyang Industrial University

[2] 0Emerging Teehoologiesthat Will Change the World.Teehnology Review,April 2003,pp33-49(in Chinese)

[3] Wu Jinrong, “On Time-Table Probelem for Arranging Courses in Universities,Operations Research and Management Science”, Operations Research and Management Science, No.6, 2006, p. 66-71

[4] a modified ZigBee Cluster-Tree network routing algorithm, measurement and control technology, 2009,28 (9) :52-55

[5] D. Kornack and P. Rakic, “Cell Proliferation without Neurogenesis in Adult Primate Neocortex,” Science, vol. 294, Dec. 2001, pp. 2127-2130, doi:10.1126/science.1065467

[6] Ma Jiangqing,SAID:A self-adaptive intrusion detection system in wireless sensor networks[c]//InformaitonSecurity Applications,7th International workshop,June2007,pp125-168

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 322-325

Study on the Partial Systematic Resampling Algorithm of Particle Filter

Jinxia Yu [1,2], Wenjing Liu [1], Yongli Tang [1,3]
[1.] College of Computer Science and Technology, Henan Polytechnic University, Henan Jiaozuo 454003, China;
Email:melissa2002@163.com
[2.] Jiangsu Provincial Key Lab of Image Processing and Image Communication, Nanjing University of Posts and Communication, Jiangsu Nanjing 210003, China;
[3.] Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China

***Abstract*—The partial systematic resampling algorithm (PSR) classifies the particles according to the weight threshold value established in advance before the resampling. Systematic resampling (SR) is carries on the minority particles, so it increases the particle diversity and reduces the computation time. In this paper, it firstly presents PSR algorithm, then, analyzes several kinds of the weight threshold values .At last, the conclusion are drawn by comparison the performance of. partial systematic resampling particle filters (PSPF) when take the different threshold values under simulation.**

***Index Terms*—resampling, weight threshold value, partial systematic resampling**

I. INTRODUCTION

Particle filter (PF) [1] uses sequential Monte Carlo methods to solve nonlinear non-Gaussian state estimation of recursive dynamic system. So, PF is widely used in the data analysis of financial field, economic statistics, information management and other fields [1-3]. Its key idea is to represent the posterior probability density ($p(x)$) [4, 5] by a set of random samples (particles) with associated weights $\{x_k^i, w_k^i\}$. The importance sampling (IS) algorithm and sequential importance sampling (SIS) algorithm are the basis of PF. However, the potential problem of the PF algorithm bases on SIS is the sample degradation [6, 7], that is, all but one particle will have negligible weight after a few iterations. Hence, particle sets can not express the actual distribution of posterior probability. To address the sample degradation, sampling importance resampling (SIR) [8, 9] is introduced in PF.

The basic idea of resampling is to eliminate particles with small weights and to concentrate on particles with large weights by adding resampling step between the importance sampling. The most representative resampling algorithms are multinomial resampling [10, 12], stratified resampling [11, 13], systematic resampling [11, 12] and the residual resampling [12, 13]. Although it can solve the sample degradation to a certain extent, resampling operation reduces the variety of samples, increases the computational complexity, and introduces additional resampling variance [14]. Therefore, the new resampling algorithm to overcome the issues is studied.

The partial systematic resampling algorithm, before the resampling, classify the particles according to the weight threshold value established in advance, systematic resampling is carries on the minority particles, it increase the particle diversity and reduces the computation time . In this paper, it firstly presents the partial systematic resampling algorithm, then, analyzes several kinds of the weight threshold values .At last, the conclusion are drawn by comparison the performance of partial systematic resampling particle filters when take the different threshold values under simulation.

II. BASIC PF ALGORITHM

The state equation and measurement equation of the dynamic system are described as follows:

$$\begin{cases} x_k = f(x_{k-1}) + u_{k-1} & (1) \\ y_k = g(x_k) + v_k & (2) \end{cases}$$

Where, x_k is the system state at time k ; y_k is the system measurement at time k ; u_k and v_k are the process noise and measurement noise at time k respectively (they obeys the independent and identical distribution). The state model $f(\ )$ and observation model $g(\ )$ are known and at least one non-linear .The state equation (1) characterizes the state transition probability of the system $p(x_k \mid x_{k-1})$, and measurement equation (2) characterizes the likelihood probability $p(y_k \mid x_k)$.

From the perspective of Bayesian filter, given that the initial state x_0 is $p(x_0 \mid z_0) \equiv p(x_0)$, the state transition probability $p(x_k \mid x_{k-1})$ and likelihood probability $p(y_k \mid x_k)$ the problem-solving core is to estimate the posterior probability density function (PDF) $p(x_k \mid y_{1:k})$.

The particle filter is the Bayesian filter's variety. It uses a set of weighted samples to approximate the posterior probability density function

$$p(x_k \mid y_{1:k}) = \sum_{i=1}^{N} w_k^i \delta(x_k - x_k^i)$$

The particle filter algorithm has three important steps: particle production (important sampling), weight computation and resampling.

Step 1 Produce particle (important sampling)

$$x_k^i \sim q\left(x_k / x_{k-1}^i, y_{0:k}\right) i = 1...N$$

Step 2 Compute weight and normalize weight

$$w_k^i \propto w_{k-1}^i p\left(y_k / x_k^i\right) p\left(x_k^i / x_{k-1}^i\right) / q\left(x_k^i / x_{0:k-1}^i, y_{1:k}\right)$$

$$\widehat{w}_k^i = w_k^i / [\sum_{j=1}^N w_k^j]^{-1}$$

Step 3 State estimate

$$\overline{x}_k = \sum_{i=1}^N x_k^i \widehat{w}_k^i$$

Steps 4 Resample

Duplicate the high weight particle and get rid of the low weight one from the particle set $\left\{x_k^i, w_k^i\right\}_{i=1}^N$, obtain the new particle set $\left\{x_k^j, w_k^j\right\}_{i=1}^N$.

III. PARTIAL SYSTEMATIC RESAMPLING ALGORITHM

A. Partial resampling

The key idea of partial resampling [15] is to perform resampling only on particles with larger or smaller weights Particles with moderate weights are not resampled.

Firstly we establish two weight threshold values w_h, w_l where, $0 < w_l < w_h$ and divide particle set into two groups according to the weight threshold values w_h, w_l.

Group A $\{x_k^j, w_k^j\}_{j=1}^{N_h} w_k^j > w_h \ or \ \ w_k^j < w_l$

Group B $\{x_k^j, w_k^j\}_{j=1}^{N-N_{hl}}, w_h > w_k^j > w_l$

Where N_{hl} is the number of particles in group A, Particles of group A with larger or smaller weight are not stable and needed resampling. Particles of group B with relatively modest weight are more stable and not resampling. N_{hl} Particles are resampled from the group A and particles of B constitute the new particle set.

Resampling is done faster because it is done on a much smaller number of particles, and communication is shorter since fewer particles are replicated and replaced. Moreover, the PR can control the thresholds either for keeping a degree of particle diversity or for reducing the degeneracy.

B. Partial systematic resmapling

Systematic resampling is carried on group A

.ALGORITHM1: PARTIAL SYSTEMATIC RESAMPLING

Step1 to initialize relative parameter: the size of particle j=1…N, time step t =1…T, weight thresholds w_h and w_l

For each time step t to do Step 2-3

Step2 Group particles into two group :

For j=1:N do

If $(w_k^j > w_h \ or \ \ w_k^j < w_l)$

$(x_k^j, w_k^j) \in A$

Else $(x_k^j, w_k^j) \in B$

$A = \{x_k^j, w_k^j\}_{j=1}^{N_{hl}}$,

$B = \{x_k^j, w_k^j\}_{j=0}^{N-N_{hl}}$

Step3 Resample:

$\{x_k^{j^*}, w_k^{j*}\}_{j=1}^{N_{hl}}$ =systematic resample $\{x_k^j, w_k^j\}_{j=1}^{N_{hl}}$

Particle set after resmapling:

$\{x_k^j, w_k^j\}_{j=1}^N = \{x_k^{j^*}, w_k^{j*}\}_{j=1}^{N_{hl}} \cup \{x_k^j, w_k^j\}_{j=0}^{N-N_{hl}}$

C. The weight threshold value

The size of weight threshold value $w_{l,} w_h$ is essential. for computing time, the diversity of particles and particle filter performance .If threshold is too large, the number of particles selected for resampling will be bigger, the computing time will increase, if too small, will reduce the number of particles resampling and reduce the performance of particle filter . Literature [15] lists several kind of weight threshold value as follows:

$w_h = [2/N, 5/N, 10/N]$

$w_l = [1/2N, 1/5N, 1/10N]$

IV. EXPERIMENT ANALYSIS

In order to evaluate the performances of partial systematic resmapling algorithms under different weight threshold values. We designed simulation program using mat lab 7.0. Partial systematic resampling particle filter is recorded as PSPF-2; when w_h *is* $2/N$ *and* w_l *is* $1/2N$, as PSPF-5; when w_h *is* $5/N$ *and* w_l *is* $1/5N$ and as PSPF-10 when w_h *is* $10/N$ *and* w_l *is* $1/10N$.

A. Experiment 1

We designed simulation program using matlab 7.0 to track a single target motion from a fixed visual observation points. Target tracking model using CV model , the state vector is $X(t) = [x(t) \quad v_x(t) \quad y(t) \quad v_y(t)]'$ where parameters are x coordinate and x direction velocity , y coordinate and y direction velocity in two-dimensional plane at t moment; T is sample time interval.; $m(t) = \left[\alpha_x(t) \quad \alpha_y(t)\right]'$ is target random acceleration as the result of random noise, here for Gaussian white noise

distribution. Target tracking system state equation as follows:

$$X(t)=\begin{bmatrix}1 & T & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & T\\ 0 & 0 & 0 & 1\end{bmatrix}X(t-1)+\begin{bmatrix}0.5T^2 & 0\\ T & 0\\ 0 & 0.5T^2\\ 0 & T\end{bmatrix}m(t-1)$$

The target tracking system which obtained measurement equation

$Z(t)=\arctan(x(t)/y(t))+n(t)$ from a fixed visual observation points, where $Z(t)$ is the observed target heading angle on the polar coordinates; $n(t)$ that obeys the Gaussian distribution white noise.

Figure 1 shows comparison of positioning error in the direction of y among the partial systematic resampling particle filters PSPF-2, PSPF-5 and PSPF-10. In 100 sampling period, the particle number N is1000, the mean square errors is 0.6465, 0.2377 and 0.2056, the performance of PSPF-10 is the best.

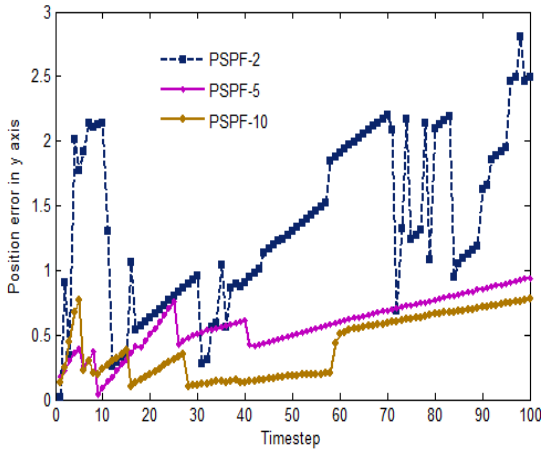


Figure 1. Comparison of positioning error in the direction of y

Table 1 shows comparison of computation time among the partial systematic resampling particle filters PSPF-2, PSPF-5 and PSPF-10. In 100 sampling period, the particle number N is1000, 2000and3000, with the increase of the number of particles computation time are increasing, but the computing time of PSPF-10 is always the smallest, PSPF-2 is the maximum.

B. Experiment 2

The state vector is $X(t)=[x(t)\ \ y(t)\ \ \theta(t)]$ where parameters are x coordinate y coordinate and angle θ in two-dimensional plane at t moment.

The state equation as follows:

$$\begin{cases}x(t+1)=x(t)+action(1)\cos[\theta(t)]-action(2)\sin[\theta(t)]\\ y(t+1)=y(t)+action(1)\sin[\theta(t)]+action(2)\cos[\theta(t)]\\ \theta(t+1)=\theta(t)+action(3)\end{cases}$$

The measurement equation is

$$Z(t+1)=[x(t+1)\ \ y(t+1)]+randn(1,2)senorNoise$$

TABLE I.
COMPARISON OF COMPUTATION TIME

Particles number	PSPF-10 average running time (s)	PSPF-5 average running time (s)	PSPF-2 average running time (s)
1000	0.1990	0. 2210	0. 2790
2000	0.3100	0.3360	0.3600
3000	0.4700	0.4940	0.5850

Where

$$action=action+[0.02\ \ 0.02\ \ 0.01]rand(1,3)$$

$randn(1,2)$ is an array with one by two rows, elements of the array are normal distributed random numbers, $randn(1,3)$ is an array with one by three rows and elements of the array are normal distributed random numbers, the noise of the sensor, senorNoise is 0.05 .the initial value X is [0. 0. 0], the initial value action is [1. 0. $\pi/3$], the number of particles is 500.

Figure 2 shows the number of distinct particles .In all of the partial systematic resampling algorithms PSPF-2, PSPF-5 and PSPF-10, the number of distinct particles exponentially decreased although there was a little difference among them. Especially the pspf-2 showed the fastest convergence, and the pspf-10 showed the slowest convergence as shown in Figure.2 better particle diversity in the PSPF-10 than that in others.

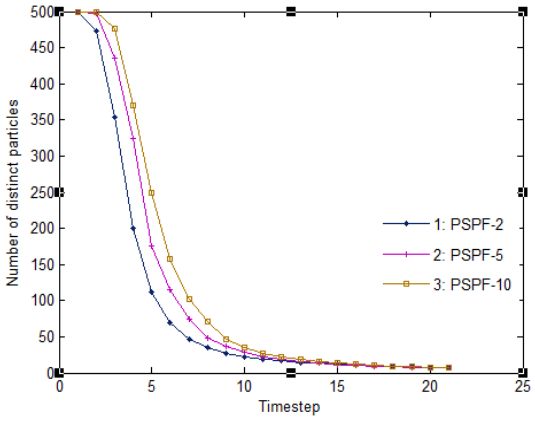


Figure 2. Comparison of number of distinct particles

V. CONCLUSIONS

The partial systematic resampling algorithm classifies the particles according to the weight threshold value established in advance before the resampling, systematic resampling is carries on the minority particles, it increase the particle diversity and reduces the computation time. In this paper, it firstly presents the partial systematic resampling algorithm, Then, analyzes several kinds of the weight threshold values .At last, the conclusion are

drawn by comparison the performance of partial systematic resampling particle filters when take the different threshold values under simulation. PSPF-10 is the best whether performance or particle diversity, while the larger calculation time.

ACKNOWLEDGMENT

This work is supported by Natural Science Foundation of Henan Educational Committee of China (2008B520015, 2009B520013), Doctoral Foundation of Henan Polytechnic University of China (B2008-61, B2009-91), Open Foundation of Jiangsu Province Key Laboratory for Image Processing and Image Communication (ZK208002).

REFERENCES

[1] N. J. Gordon, D. J. Salmond and A. F. M. Smith, “Novel approach to nonlinear and non-Gaussian Bayesian state estertimation”, IEE Proceedings on Radar and Signal Processing, vol 140, no. 2, pp.107-113, Apr. 1993.

[2] S. Y. CHENG and J. Y. ZHAO, “Review on Particle Filters”, Journal of Astronautics, vol 29, no. 4, pp.1099-1111, July .2008.

[3] S. Q Hu and Z. L. Jing, “Overview of particle filter algorithm”, Control and Decision, vol 20, no. 4, pp. 361-365, Apr. 2005.

[4] E. Belviken and P J. Acklam, “Monte Carlo filters for nonlinear state estimation” , Automatica, vol 37, no. 2, pp.177-183,2001.

[5] G. Casella. “Statistical inference and Monte Carlo algorithms”, Test, vol 5, no. 2, pp. 249-344, 1977.

[6] M. S. Arulampalam, S. Maskell and N. J. Gordon, “A tutorial on particle filters for online non-1inear/non-gaussian Bayesian tracking”, IEEE Transaction on Signal Processing, vol 50, no.20, pp.174-188, Feb. 2002.

[7] A. Doucet, S. J. Godsill and C. Andrieu, “On sequential Monte Carlo sampling methods for Bayesian filtering”, Statistics and Computing, vol 10, no.3, pp.197-208, 2002.

[8] J. S. Liu and R. Chen, “Blind deconvolution via sequential imputation”, Journal of the American Statistical Association, vol 90, no.2, pp.567-576, 1995.

[9] D. B. Rubin, “Comment on the calculation of posterior distributions by data augmentation”, Journal of the American Statistical Association, vol 82, no.398, pp.543-546, 1987.

[10] A. F. M. Smith and A. E. Gelfand, “Bayesian statistics without tears: A sampling-resampling respective”, American Statistician, vol 46, no. 2, pp.84-88, 1992.

[11] Z. H. Du, “Research on Particle and its Application in MIMO wireless communication”. Chengdu, Electronics Science and Technology University, 2008.

[12] C. Feng, M. Wang and Q. B. Ji “Analysis and Comparison of Resampling Algorithms in Particle Filter”, Journal of System Simulation, vol 21, no 4, pp.1101-1105, Feb, 2009.

[13] J. X. Yu, Z. X. Cai and Z. H. Duan，“Survey on Some Key Technologies of Mobile Robot Localization Based on Particle Filter”, Journal of Application Research of computers, Vol 24, No.11, pp.9-4, Nov.2007.

[14] K. H. Xia and Z. L. Xu, “Critical technologies and application of particle filter”, ELECTRONICS OPTICS & CONTROL, vol 12, no.6, PP.14-19, Dec.2005

[15] M. Bolic, P. M. Djuric, and S. Hong, "Resampling algorithms for particle filters: a computational complexity perspective," Journal on Applied Signal Processing, no. 15, pp. 2267-2277, 2004.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 326-328

Comprehensive Information Based Pornographic Image Recognition Model

Hairu Guo, Peiqian Liu, and Jiyu An
College of Computer Science and Technology,Henan Polytechnic University,Jiaozuo, China
guohr@163.com

***Abstract*—We present a proto model to recognize the pornographic image from benign image. Although different models for this application were presented in the past, most of them are based on the syntactic information of the image and have rather poor performance. Recent advances in the information science theory in particular Comprehensive Information Theory (CIT) has shown that not only syntactic information, but also semantic information and pragmatic information should be used for many information processing problems such as pornographic image recognition. Based on CIT, a novel model is discussed in this paper. The methodology reported in the paper may lay certain foundation to solve the bottleneck of the pornographic image recognition and a new, and a promising approach to the research of other image understanding problems may also hopefully be initiated.**

***Index Terms*—comprehensive information, pornographic image recognition, syntactic information, semantic information, pragmatic information**

I. INTRODUCTION

With the rapid growth of the Internet, any user can access and browse a large volume of contents on the web. Internet is a double edged sword as it brings us great convenience, while there are some objectionable contents, such as pornographic images, which are very harmful to people's bodies and mind, especially to teenagers. It is a meaningful and urgent task to protect people from accessing unexpected pornographic images. There are a number of research efforts can be found in the recent literatures which are to develop an effective detection and filtering technology to prevent the access to unexpected pornographic images [1-10]. The key is how to identify an image is pornographic or not. This is an intelligent process in which the information contained in the image to be examined is used to generate the knowledge, and produce intelligent strategy [11]. It is obvious that most detection technologies of pornographic images are based on syntactic information (low-level visual features such as color, texture and shape), but generally these information will fail to distinguish benign images with large skin regions from pornographic images [1,3,8,9,12]. Instead of the syntactic information is used, all components of the comprehensive information include syntactic information, semantic information and pragmatic information are fully utilized for pornographic image detecting.

The rest of this paper is organized as follows. In Section 2, Comprehensive Information Theory is introduced which is the basis theory to be used to describe the comprehensive information based pornographic image recognition model (CIBPIRM) in Section 3. In Section 4, we made a conclusion in methodology.

II. INTRODUCTION TO CI

In 1948, C. E. Shannon has published a famous paper titled by "A Mathematical Theory of Communication", and pioneered the Shannon Information Theory (SIT). Because the task of communications is to duplicate the waveforms of the signals, sent from the source, at the destination and need not care about the meaning and value of the signals. Although it brought the world into the information times, there are many new problems have shown that SIT has some limitations. Today people have recognized that the full usage of information resources, more precisely the understanding of information, is crucial for dealing with all kinds of intelligent systems, but SIT can't provide sufficient support to this end.

Information science theory [11] points out: information is in multi-level, include ontology level and epistemology level. Any information in the epistemology level is the state of motion and its change form about the object which is perceived by the subject, including three essential factors, which is the form, meaning and the utility, called the grammar, the semantic and the pragmatic separately. The "Comprehensive Information"(CI) is defined as a trinity - the form, the meaning, and the value all related to the object's states and the manner of the states varying.

Syntactic information reflecting its structure and grammar, semantic information describing its meaning related with objects, and pragmatic information expressing its utility related with subjects. The CI is shown in Figure 1 below.

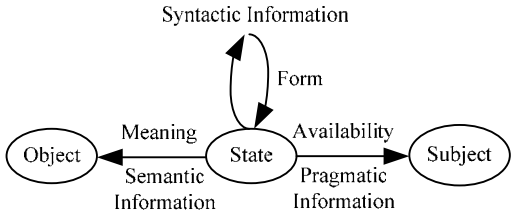


Figure 1. Comprehensive Information

The characteristic of CI makes it especially fit for describing pornographic image. It has provided us an original method to face the problem of pornographic image recognition.

III. CI-BASED PORNOGRAPHIC IMAGE RECOGNITION MODEL

This section describes how to define the syntactic information, semantic information and pragmatic information in the pornographic image recognition model. Then we combine these kinds of information into comprehensive information of pornographic image, which can be used to recognize whether an image is pornographic or not.

A. Syntactic Information

In CIT, syntactic information refers to the form of objects. For pornographic image recognition, syntactic information can be expressed in skin feature, texture feature and shape feature.

In principle, syntactic information can be obtained through the following steps.

Step 1: Observe a sample pornographic image S_1, extract the features (include skin feature, texture feature and shape feature), denoted as f_1 (f_1 may be a vector);

Step 2: Establish the similarity criteria of the features, then observe another sample image S_2, extract the features, denoted as f_2, compare f_2 with f_1, if f_1 is similar with f_2 under the similarity criteria, S_2 is accepted, otherwise abandoned.

Step 3: Repeat the step 2 for N times (here N is a sufficiently large positive integer), $\{f_k\}$ is obtained to depict a pornographic image, k=1, 2, ..., K, where K is a positive integer and is smaller than N.

Step 4: When K is steady with the increasing N or N is no increase, frozen the K samples, then the $\{f_k\}$ is a pornographic image's feathers set.

Step 5: Given a new image, $\{C_k\}$ can be obtained by calculating the value of f. Here, $\{C_k\}$ is the syntactic information can be used to recognize a pornographic image.

B. Pragmatic Information

In CIT, pragmatic information refers to usefulness of the objects. For pornographic image recognition model, it can be described what the harmful level of an image is.

In order to describe or measure pragmatic information of the images, first we define a set of categories, closely following [13], where the images are grouped into five different categories according to the image on the harmful levels:

Class 1: inoffensive images,

Class 2: lightly dressed persons, might be offensive in very strict environments,

Class 3: partly nude persons, might be objectionable in school environments,

Class 4: nude persons, likely objectionable in many environments, and

Class 5: porn images, probably offensive in most environments.

Then we can calculate the syntactic information through the following steps.

Step 1: Defined the general goal for the subject clearly, marked with $G=\{G_n\}$,n=1, …, 5;

Step 2: Input X, calculate D(X) and the corresponding U(X) which is related with the value of G according to the description of X;

Step 3: Repeat the step 2 for all the training images;

Step 4: Study the relationship between D(X) and U(X), generate a rule;

Step 5: When a new image X is input, U(X) can be obtained by the rule in step 4;

Step 6: A vector $\{U_k\}$ can be produced to measure the pragmatic information of the images, k=1, 2, ..., N.

C. Semantic Information

Until now, little literature is found to make use of the image's semantic information besides the corresponding text meaning for the image [14].

In CIT, semantic information refers to meaning of the objects. So, the semantic information of the image is not easy to obtain directly compared with syntactic information and pragmatic information. But it can be measured with the help of the syntactic information and pragmatic information indirectly.

The steps to calculate the semantic information are follows:

Step 1: Input X, calculate C(X) based the algorithm in subsection A;

Step 2: Input X, calculate U(X) based the algorithm in subsection B;

Step 3: Make an operation between C(X) and U(X) with implication operation, CONT: KC |→KU, calculate the logical realism of X, marked as T(X);

Step 4: Repeat from step 1 to 3 for all images, a vector $\{T_k\}$ can be produced to measure the semantic information of the images, k=1, 2, ..., N.

D. Comprehensive Information

After defining syntactic information, semantic information and pragmatic information of an image, we can combine these components into comprehensive information. The comprehensive information vector is defined as:

$$\begin{bmatrix} X_1 & \cdots & X_k & \cdots & X_N \\ C_1 & \cdots & C_k & \cdots & C_N \\ T_1 & \cdots & T_k & \cdots & T_N \\ U_1 & \cdots & U_k & \cdots & U_N \end{bmatrix} \tag{1}$$

In this paper we use a linear regression model to predict comprehensive information of the image as in:

$$\eta_k = \alpha \cdot C_k + \beta \cdot T_k + \gamma \cdot U_k \tag{2}$$

In the linear regression model, coefficient α, β and γ can be trained by the training set. The integrated effect score of each image is measured by comprehensive information marked as η_k in formula (2). The image that has the highest amount of comprehensive information is then recognized as pornographic image.

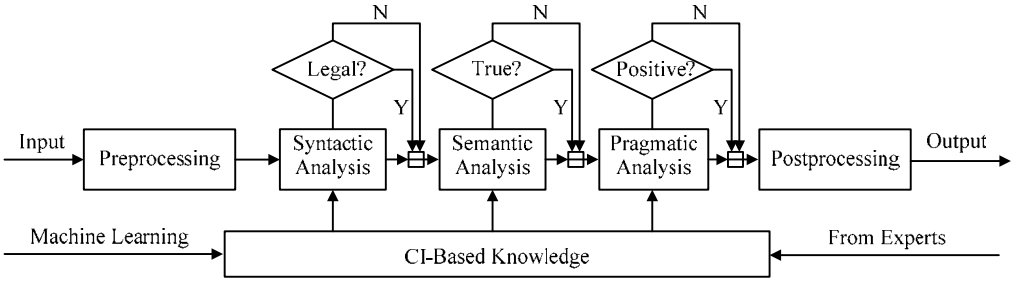


Figure 2. CI-Based Pornographic Image Recognition Model

E. CI-Based Pornographic Image Recognition Model

A CI-Based Pornographic Image Recognition Model is given in Figure 2. The input of model is a variety of images including pornographic and inoffensive. The output is whether the input image is pornographic or not. In the model, there are five processing parts followed by preprocessing, syntactic analysis, semantic analysis, pragmatic analysis and postprocessing. The preprocessing of the image is mainly responsible for extracting image features, such as color, texture, shape. The syntactic information, semantic information, pragmatic information and comprehensive information of an image can be obtained by the algorithms in subsection A, B, C and D. The CI based knowledge is the kernel part of the model which can be build by experts or machine learning technology. In syntactic analysis process, it can be determined whether the form of the input image is legal to pornographic images or not based on the CI Based Knowledge. Similarly, whether the meaning of the image is true or not and whether the availability of the image is positive or not compared with pornographic images can be respectively obtained based on the CI Based Knowledge in the semantic analysis and pragmatic analysis process.

IV. CONCLUSIONS

The study in this paper is expected to have a theoretical and methodological breakthrough in the pornographic image recognition and detection. Comprehensive Information is the inevitable choice to deal with the future application of the information problem. In this regard, this paper is a very useful exploration. The methodology of this paper can be applied to spam filtering, search engine filtering, blocking of undesirable web information, and so on.

REFERENCES

[1] Wang J, Wiederhold G, Firschein O. System for screening objectionable images using daubechies' wavelets and color histograms, Computer Communications, Vol.21, no.25, pp.1355-1360, 1998.

[2] Jiao F, Gao W, Duan L, et al. Detecting adult image using multiple features, in Proc. of Int. Conf. on Infotech and Info-net ICII'01, pp.378-383, 2001.

[3] A Abadpour, S Kasaei. Comprehensive Evaluation of the Pixel-Based Skin Detection Approach for Pornography Filtering in the Internet Resources, in Int. Symposium on Telecommunications, Shiraz, Iran, pp.829-834, 2005.

[4] C.-Y. Jeong, J.-S. Kim, and K.-S. Hong, Appearance-based nude image detection, in Proc. of 17th Int. Conf. on Pattern Recognition (ICPR'04), Cambridge, UK, Vol.4, pp.467-470, 2004.

[5] Schettini R, Brambilla C, Cusano C, et al. On the detection of pornographic digital images, Visual Communications and Image Processing, Vol.5150, pp.2105-2113, 2003.

[6] Lin Y, Tseng H, Fuh C. Pornography Detection Using Support Vector Machine, 16th IPPR Conf. on Computer Vision, Graphics and Image Processing, 2003.

[7] Jeong C, Han S, Choi S, et al. An Objectionable Image Detection System Based on Region of Interest, in Proc. of 2006 IEEE Int. Conf. on Image, pp.1477-1480, 2006.

[8] Zhu H, Zhou S, Wang J, et al. An algorithm of pornographic image detection, in Fourth Int. Conf. on Image and Graphics, pp.801-804, 2007.

[9] SUN Y, RUAN Q. A new fast system for objectionable image identification based on shape features, in 9th Int. Conf. on Signal Processing, pp.1087-1090, 2008.

[10] Yang J, Fu Z, Tan T, et al. A Novel Approach to Detecting Adult Images, in Proc. of the 17th Int. Conf. on Pattern Recognition, Vol.4, pp.479-482, 2004.

[11] Y.X. Zhong, Principles of Information Science, BUPT Press, Beijing, 2002.

[12] Wen Zhiqiang, Zhu Yanhui, Peng Zhaoyi, Survey on Web Image Content-Based Filtering Technology, 1st International Conference on Information Science and Engineering (ICISE 2009), IEEE Press, 2009, pp. 1463-1466, doi: 10.1109/ICISE.2009.1152.

[13] Deselaers. T., Pimenidis. L., Ney. H., Bag-of-visual-words models for adult image classification and filtering, 19th International Conference on Pattern Recognition (ICPR 2008), IEEE Press, 2008, pp. 1-4, doi: 10.1109/ICPR.2008.4761366.

[14] Liu Yizhi, Lin Shouxun, Well-Defined Semantic Templates for Pornographic Images Identification, 1st International Conference on Information Science and Engineering (ICISE 2009), IEEE Press, 2009, pp. 1507-1510, doi: 10.1109/ICISE.2009. 1350.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 329-332

Road Traffic Freight Volume Forecasting Using Support Vector Machine

Shang Gao[1], Zaiyue Zhang[2] and Cungen Cao[2]
[1]School of Computer Science and Technology, Jiangsu University of Science and Technology, Zhenjiang 212003, China
Email: gao_shang@hotmail.com
[2]Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China
Email: yzzjzzy@sina.com, cgcao@ict.ac.cn

***Abstract*—The grey system forecasting model, neural network forecasting model and support vector machine forecasting model are proposed in this paper. Taking the road goods traffic volume from year of 1996 to 2003 in the whole country as a study case, the forecasting results are got by three methods. Compared with grey system forecasting model and neural network forecasting model, the accuracy of the support vector machine combining forecasting method is higher.**

***Index Terms*—grey system, neural network, support vector machine, combining forecasting, traffic volume**

I. Introduction

With the development of society, transportation plays a more and more important role in social economical progress, at the same time transportation is rapidly progressing too. As far as we know, the accurate and objective prediction of the future road transportation demand is the just foundation of the scientific transportation planning. It turns out that the excepted effect gains verification, effectively improves the model accuracy, and makes more exact forecasting, which expects to be helpful to concerned departments and personnel for them to grasp the traffic market trend or make decision. Combining forecasting has brought great attention to by the forecasting circles since 1969 when J . M. Bates and C. W. J. Granger proposed its theory and method. The theory and methods of combining forecasting have been developed widely in recent years [1]. For practical cases of various forecasting problems, combining forecasting models may have different forms. Among them proportional mean combining forecasting models are widely used , such as simple weighted arithmetic proportional mean combining forecasting model , simple weighted square root proportional mean combining forecasting model , simple weighted harmonic proportional mean combining forecasting model , generalized weighted arithmetic proportional mean combining forecasting model and generalized weighted logarithmic proportional mean combining forecasting model , etc. In this paper, the grey system forecasting model, neural network forecasting model and support vector machine forecasting model are proposed. Based on grey system forecasting model, neural network forecasting model and support vector machine forecasting model, the linear combining forecasting model, combining support vector machine forecasting model are set up.

II. Grey Predicting Model

Since 1982, there has been a quick development in grey systems theory in China, and it is also very successful in the application of the theory to many real projects, such as agriculture, society, economics, engineering, IT, data mining, management, biological protection, robot, ecology, image processing, environmental studies, etc.. Grey model GM(1,1) due to whole distinguishing features: modeling by less data (suiting the data as few as 4), thus underlay grey modeling and grey forecasting. Because sometimes the precision of grey method by means of AGO (accumulated generation operation) and IAGO (inverse accumulated generation operation) can not meet the requirement of actual forecasting, much research in theory and application has been done.

The GM (1,1) model means a single differential equation model with a single variation. The modeling process is as follows: First of all, observed data are converted into new data series by a preliminary transformation called AGO (accumulated generating operation). Then a GM model based on the generated sequence is built, and then the prediction values are obtained by returning an AGO' s level to the original level using IAGO (inverse accumulated generating operation).

Now we introduce the grey predicting model GM(1,1). Let $X^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(n)\}$ .By defining

$$x^{(1)}(k) = \sum_{i=0}^{k} x^{(0)}(i),$$

We get a new series $X^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \cdots, x^{(1)}(n)\}$.

To some processes, $X^{(1)}$ is the solution of the following grey ordinary differential equation [2]

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b \qquad (1)$$

where a and b are grey numbers. The equation (1) is called GM (1, 1).

By taking average $z^{(1)}(k)=\frac{1}{2}[x^{(1)}(k)+x^{(1)}(k-1)]$ and other transformations, we get that

$$\begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \\ \\ x^{(0)}(n) \end{bmatrix} = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} \tag{2}$$

which can be simplified as $y_N = B\hat{a}$, where $y_N=[x^{(0)}(2),x^{(0)}(3),\cdots,x^{(0)}(n)]^T$, $\hat{a}=\begin{bmatrix} a & b \end{bmatrix}^T$ and $B=\begin{pmatrix} -z^{(1)}(2) & -z^{(1)}(3) & \cdots & -z^{(1)}(n) \\ 1 & 1 & \cdots & 1 \end{pmatrix}^T$.

TABLE I.
THE ACTUAL TRAFFIC VOLUME FROM 1996 TO 2003 AND THE FORECASTING VALUES OF GM

Time	Actual Data	forecast value	relative error
1996	984	984.0	0.00%
1997	977	950.2	2.75%
1998	976	980.1	0.42%
1999	990	1010.9	2.11%
2000	1039	1042.7	0.36%
2001	1056	1075.5	1.85%
2002	1116	1109.4	0.59%
2003	1160	1144.3	1.36%
2004		1180.3	
2005		1217.4	
2006		1255.7	
2007		1295.2	
2008		1336.0	

If rank(B)=2, the equation (2) has a unique solution: $\hat{a}=(B^TB)^{-1}B^Ty_N$. Therefore, from (1) we obtain the generating model:

$$\hat{x}^{(1)}(k+1)=\left(x^{(0)}(1)-\frac{b}{a}\right)e^{-ak}+\frac{b}{a} \tag{3}$$

From (3), each value of $\hat{x}^{(0)}(k)$ can be computed. Thus, we compute the feedback values $\hat{x}^{(0)}(k+1)=\hat{x}^{(1)}(k+1)-\hat{x}^{(1)}(k)$:

$$\begin{cases} \hat{x}^{(1)}(1)=x^{(0)}(1) \\ \hat{x}^{(0)}(k)=(x^{(0)}(1)-\frac{b}{a})(1-e^{a})e^{-a(k-1)} \qquad (k=2,3,\cdots) \end{cases} \tag{4}$$

The road goods traffic volumes from year of 1996 to 2003 in the whole country are listed in Table 1 in detail. We can get the forecasting values of grey GM(1,1) listed in Table 1.

III. NEURAL NETWORK PREDICTING MODEL

An Artificial Neural Network (ANN) is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems. ANNs, like people, learn by example. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurons.

A neural network consists of simple processing units and each of the processing units has natural inclination for storing experimental knowledge and making it available for use. These simple processing units, called neurons or perceptions, form distributed network. An artificial neural network is an abstract simulation of a real nervous system that contains a collection of neuron units communication with each other via axon connections. Due to its self-organizing and adaptive nature, the model potentially offers a new parallel processing paradigm that could be more robust and user-friendly than the traditional approaches. As in nature, the network function is determined largely by the connections between elements. We can train a neural network to perform a particular function by adjusting the values of the connections (weights) between elements.

A neuron is a processing unit, which has n inputs and m outputs. $x_1,x_2,\cdots,x_n$ are outputs of previous layers. w_{ij} is the weight by which neuron i contribute to neuron j . b_j is the threshold of neuron j .The net input net_j is defined by [3]

$$net_j=\sum_{i=1}^{n}x_iw_{ij}-b_j$$

where O_j is the output of the neuron j .Then $O_j=f(net_j)$.

f is a transfer function, which takes the argument input and produces the output. The transfer function is very often a sigmoid function, in part because it is differentiable. The sigmoid transfer function is

$$f(net)=\frac{1}{1+e^{-net}}$$

The back-propagation network represents one of the most classical examples of an ANN, being also one of the most simple in terms of the overall design. The network is a straight feedforward network: each neuron receives as input the outputs of all neurons from the previous layer. We adopt a three-layer back-propagation network (see Figure 2). The pretreatment life data are fed to the inputs. The output of network is life distribution. The network has some hidden. The objective is to train the weights and the thresholds, so as to minimize the least-

squares-error between the teacher and the actual response.

In this paper, A standard three-layer multi-layer perceptron trained using the back propagation (BP) algorithm is used. The back-propagation network has one input, tree hidden neurons and one output. The value of time is input, and the forecasting value is output. The ANN was trained with the following parameters: learning parameter=0.5, momentum=0.2, error=0.01. The forecasting value data are inputs of trained network. The actual output of network can be calculated by using these weights and the thresholds. We can get the forecasting values of ANN listed in Table 2.

TABLE II.
THE ACTUAL TRAFFIC VOLUME FROM 1996 TO 2003 AND THE FORECASTING VALUES OF ANN

Time	Actual Data	forecast value	relative error
1996	984	1012.0	2.85%
1997	977	1026.2	5.03%
1998	976	1040.6	6.61%
1999	990	1055.2	6.59%
2000	1039	1070.2	3.00%
2001	1056	1085.3	2.78%
2002	1116	1100.7	1.37%
2003	1160	1116.4	3.76%
2004		1129.5	
2005		1145.6	
2006		1161.9	
2007		1178.4	
2008		1195.1	

IV. SUPPORT VECTOR MACHINE PREDICTING MODEL

Support vector machine(SVM) proposed by Vapnik in 1992 is a new machine 1earning method, which is developed based on Vapnikcher vonenkis (VC) dimension theory and the principle of structural risk minimization(SRM) from statistical learning theory. Originally, SVM were developed for pattern recognition problems. Recently, with the introduction of e-insensitive loss function, SVM have been extended to solve non-linear regression problems. SVM has been tested on a lot of application fields including classification, time, serial estimation, function approximation, text recognition, etc.. SVM has the comprehensive theory foundation such as the universal convergence, speed of convergence, controllability of generalization ability.

Consider the problem of approximating the set of data, $D=\{(x_i,y_i)\,|\,i=1,2,\cdots,l\}$, $x_i\in R^n$, $y_i\in R$, with a linear function[4],

$$f(x)=\langle w,x\rangle+b \tag{5}$$

Using ε -insensitive loss function,

$$L_\varepsilon(y)=\begin{cases}0 & for \quad |f(x)-y|<\varepsilon\\ |f(x)-y|-\varepsilon & otherwise\end{cases} \tag{6}$$

The optimal regression function is given by the minimum of the functional,

$$\begin{aligned}&\min_{w,b,\xi_i,\xi_i^*}\Phi=\frac{1}{2}\|w\|^2+C\sum_{i=1}^{l}(\xi_i+\xi_i^*)\\ &s.t.\quad ((w\cdot x_i)+b)-y_i\le\varepsilon+\xi_i \qquad i=1,2,\cdots l\\ &\qquad y_i-((w\cdot x_i)+b)\le\varepsilon+\xi_i^* \qquad i=1,2,\cdots l\\ &\qquad \xi_i,\xi_i^*\ge 0, i=1,2,\cdots l\end{aligned} \tag{7}$$

where C is a pre-specified value, and ξ_i , ξ_i^* are slack variables representing upper and lower constraints on the outputs of the system.

The optimal regression function is given by the minimum of the functional,

$$\begin{aligned}&\min_{w,b,\xi_i,\xi_i^*}\Phi=\frac{1}{2}\|w\|^2+C\sum_{i=1}^{l}(\xi_i+\xi_i^*)\\ &s.t.\quad ((w\cdot x_i)+b)-y_i\le\varepsilon+\xi_i \qquad i=1,2,\cdots l\\ &\qquad y_i-((w\cdot x_i)+b)\le\varepsilon+\xi_i^* \qquad i=1,2,\cdots l\\ &\qquad \xi_i,\xi_i^*\ge 0, i=1,2,\cdots l\end{aligned} \tag{8}$$

where C is a pre-specified value, and ξ_i , ξ_i^* are slack variables representing upper and lower constraints on the outputs of the system.

Equivalently one can solve the dual formulation of the optimization problem:

$$\begin{aligned}&\max_{\alpha,\alpha^*}W=-\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)\langle x_i,x_j\rangle\\ &+\sum_{i=1}^{l}[\alpha_i(y_i-\varepsilon)-\alpha_i^*(y_i+\varepsilon)]\\ &s.t.\quad \sum_{i=1}^{l}(\alpha_i-\alpha_i^*)=0\\ &\qquad 0\le\alpha_i,\alpha_i^*\le C, i=1,2,\cdots l\end{aligned} \tag{9}$$

Solving Equation (9) determines the Lagrange multipliers α_i , α_i^*, and the regression function is given by

$$w=\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)x_i \tag{10}$$

The non-linear mapping can be used to map the data into a high dimensional feature space where linear regression is performed. The kernel approach is again employed to address the curse of dimensionality. The non-linear SVR solution, using ε -insensitive loss function,

$$f(x)=\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)K(x_i,x)+b \tag{11}$$

is given by,

$$\max_{\alpha,\alpha^*} W = -\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(x_i, x_j)$$
$$+\sum_{i=1}^{l}[\alpha_i(y_i - \varepsilon) - \alpha_i^*(y_i + \varepsilon)] \quad (12)$$
$$s.t. \quad \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0$$
$$0 \le \alpha_i, \alpha_i^* \le C, i = 1,2,\cdots l$$

where $K(x_i, x)$ is the kernel function performing the non-linear mapping into feature space. There are many kernel functions, such as polynomial function, radial basis function, exponential radial basis function and multi-layer perception function etc.. Table 3 illustrates the SVR solution for a exponential radial basis function with $C = 1000$, $\varepsilon = 0.0001$ and $\sigma = 18$.

TABLE III.
THE ACTUAL TRAFFIC VOLUME FROM 1996 TO 2003 AND THE FORECASTING VALUES OF SVM

Time	Actual Data	forecast value	relative error
1996	984	984.0	0.00%
1997	977	976.8	0.02%
1998	976	980.4	0.45%
1999	990	994.8	0.48%
2000	1039	1020.0	1.83%
2001	1056	1056.0	0.00%
2002	1116	1102.7	1.19%
2003	1160	1160.0	0.00%
2004		1227.8	
2005		1306.0	
2006		1394.3	
2007		1492.6	
2008		1600. 7	

The exponential radial basis function is given by,

$$K(x_i, x) = \exp(-\frac{\|x_i - x\|}{2\sigma^2}) \quad (13)$$

We can get the forecasting values of SVM listed in Table 3 and Figure 8. The sum of squares errors of three methods are described in Table4. From Table 4, we can conclude that the accuracy of the support vector machine method is higher than the other two methods.

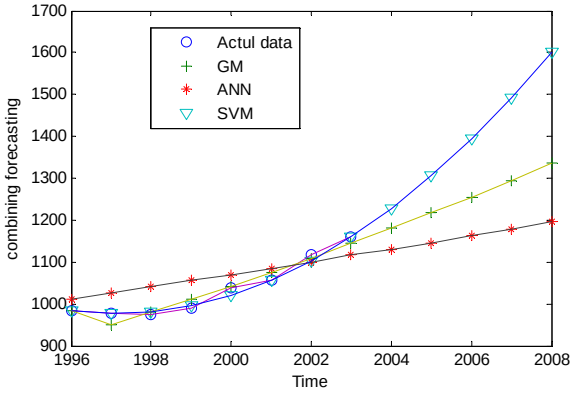


Figure 1. Three forecasting methods

TABLE IV.
THE FORECASTING VALUES OF THREE METHODS THE SUM OF SQUARES ERRORS OF THREE METHODS

Methods	GM	ANN	SVM
sum of squares errors	1859.8	15587.9	579.7
average relative error	1.18%	4.00%	0.50%
maximum relative error	2.75%	6.61%	1.83%

IV. CONCLUSIONS

The use of SVMs in the road goods traffic volume forecasting is studied in this paper. The study has concluded that SVM provide a promising alternative to time series forecasting because they use a risk function consisting of the empirical error and a regularized term which is derived from the structural risk minimization principle. Compared with single prediction methods, linear combining forecasting method and neural network combining forecasting method, the accuracy of the support vector machine method is higher.

ACKNOWLEDGMENT

This work was partially supported by Artificial Intelligence of Key Laboratory of Sichuan Province (2009RY001) and the National Natural Science Foundation of China under Grant No.60773059.

REFERENCES

[1] J. M. Bates, C.W.J Granger. Combination of forecasts, Operations Research, pp.451-468,1969， 20(4), pp.451-468.
[2] J. L. Deng. Multidimensional grey planning, Huazhong University of Science and Technology Press, pp.30-35,1990.
[3] L. M. Zhang. Models and applications of artificial neural networks. Shanghai: Fudan University Press, pp.32-47,1994.
[4] S. R. Gunn. Support Vector Machines for Classification and Regression, Technical Report, Image Speech and Intelligent Systems Research Group, University of Southampton, 1997.
[5] X. W. Tang ,C. X. Cao. Study of combination forecasting method, Control and Decision, 1993,8(1). pp.7-12.
[6] W. D. Zhou, L. Zhang, L. C. Jiao. Linear Programming Support Vector Machines. Acta Electronica Sinica, 29(11), pp.1507-1511, 2001.(in Chinese).
[7] J. H. Xu, X. G. Zhang. Nonlinear Kernel Forms of Classical Linear Algorithms. Control and Decision, 1(1), pp.1-6, 12, 2006. (in Chinese).
[8] W. S. An, Y. G. Sun. A New Method for Constructing Kernel Function of Support Vector Regression. Information and Control, 35(3), pp.378-381, 2006. (in Chinese).

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 333-335

Research of Security Identity Authentication Based on Campus Network

Guo Zhenghui [1], Han Xiujuan [2]
[1] College of Computer Science and Technology Henan Polytechnic University, Jiaozuo Henan,China
guozh@hpu.edu.cn
[2] College of Computer Science and Technology Henan Polytechnic University, Jiaozuo Henan,China
hanxj@hpu.edu.cn

***Abstract*—With the development computer technology of campus network, more and more application systems become popular. These application systems are independent each other, everyone has its own different account information in different systems and must save its different account information. Based on the above fact, the paper analyzes common authentication scheme, then propose a security single sign-on way. The method implement a security uniform identity authentication using PKI,LDAP,CAS. In this way user can access all corresponding application systems when they login only one time. This approach can allow users to easily manager their account information.**

***Index Terms*—LDAP, Authentication, SSO, PKI**

I. INTRODUCTION

With the continuous increased infrastructure in campus network and development technology of the electronic information, more and more application systems are rapidly used in campus network. These systems are independent each other, such as office automation, campus network accounting system, financial tracking system, educational management system, library loan system, etc. Each system has its own authentication database that is different from other and users in each system have their own accounting information. It is difficult for users and administrators to manage the accounting information of many different systems. So the more number of applications system, the more complicated. So these systems urgently need the support of SSO. Based on the above face, the paper presents a SSO[1-2] method after analysis of several different common authentication systems. It allows users to login once to access all mutual trust systems, at the same time SSO confirm the identity of communications and provide data security. This authentication method uses LDAP as its database that is a simplified version of X500, X509 is a part of X500. Judging from the nature of data, the certificate data store in LDAP, the information user generated come from LDAP. This paper will use CAS as authentication method and add plug for support of certificate method.

II. ANALYSIS OF COMMON AUTHENTICATION METHOD

A. General Authentication Method

General authentication is each application has its own independent authentication method and is mutually coupled from other systems. When users access different application systems, they must enter corresponding information. At present this authentication method is adopted by many application systems, and each system's authentication and authorization are different from others.

B. Authentication Method of LDAP and Radius

LDAP is Lightweight Directory Access Protocol[3], based on the X.500 standard, but significantly simpler and more easily adapted to meet custom needs. Unlike X.500, LDAP supports TCP/IP, which is necessary for Internet access. The LDAP protocol enables corporate directory entries to be arranged in a hierarchical structure that reflects geographic and organizational boundaries. LDAP directories are arranged as trees, please see Fig. 1. One of the most important features of both X.500 and LDAP is the ability to search for user-specified resources.

Radius (Remote Authentication Dial In User Service) is a networking protocol that provides centralized Authentication, Authorization, and Accounting management for computers to connect and use a network service. It runs in the application layer, uses UDP as transport, and supports a wide variety of authentication schemes.

The above two protocols can provide unified authentication methods, both run in the application layer, based on C/S mode. They supports a wide variety of authentication schemes and have a variety client

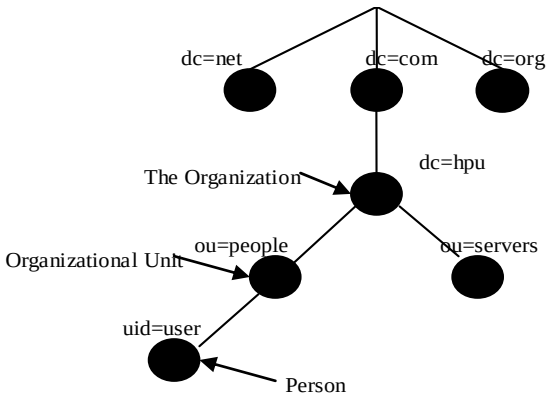


Figure 1. Example of ldap of a figure

Zhenghui Guo, male, 1978, shanxi, china, master,engineer,computer network.

development environment, such as SDK for C, SDK for JAVA, SDK for Perl, etc. Radius can store its accounting information in LDAP database. But they are not SSO; logging in different systems need enter different user information, do not have Characteristic behavior of a single sign-on.

C. Kerberos Authentication Method

Kerberos is a network authentication protocol. It is designed to provide strong authentication for client/server applications by using secret-key cryptography. It is the implementation of SSO and allows nodes communicating over a non-secure network to prove their identity to one another in a secure manner. It makes use of a trusted third party, termed a key distribution center, which consists of two logically separate parts: an Authentication Server and a Ticket Granting Server. Kerberos works on the basis of "tickets" which serve to prove the identity of users. please see authentication process of Fig. 2, it is the realization of C/S of SSO. The SSO referred in this paper is based WEB mode and is a simplified version of Kerberos to implement security unified identity of cross-domain. The service ticket stored in KDC can only be used once compared to Kerberos.

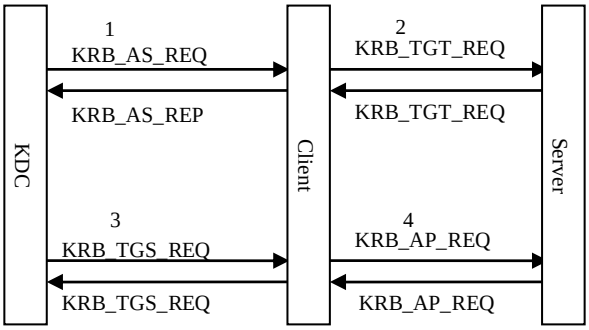


Figure 2. Kerberos authentication process of a figure.

III. Implementation of Certificate Repository

PKI is public key infrastructure, it enables users to securely and privately exchange data through the use of a public and a private cryptographic key pair that is obtained and shared through a trusted authority. The public key infrastructure provides for a digital certificate that can identify an individual or an organization. A public key certificate is a cryptographically signed digital structure that guarantees the association between at least one identifier and a public key. The X.509 document defines the format of a public key certificate and of certificate revocation list. An LDAP directory represents the perfect repository for public user information and public key certificates and offers distributed access to the data it stores[4]. Like a database schema, a directory schema defines how data is represented in the directory. In order for the directory to serve as a repository for the certificate, the directory schema should be open and extensible.

Please see Fig. 3, it show how to use LDAP as certificate database. All the basic information of users

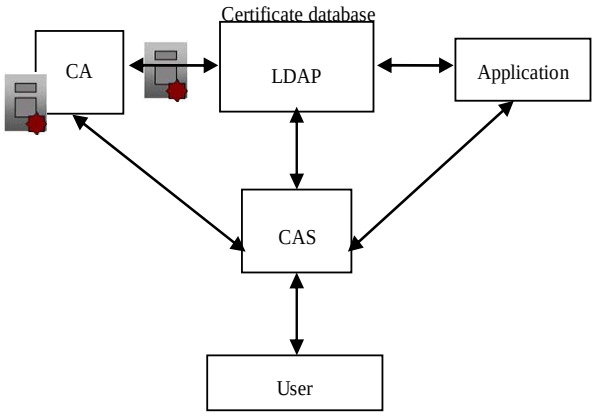


Figure 3. Certificate database based on LDAP of a figure

stored in LDAP, CA issued certificates and CRL is also stored in LDAP. All certificate original data user requested come from LDAP, then CA signature it. The certificate use the standard PKCS#12 and use pfx format[5]. User download it and import it to system to provide individual account security and personal identification. Once user has his certificate, he can use it to login in application system while he can also choose traditional user/password way. This paper recommends using certificate method, at the same time authentication system will get user information from certificate. Only the certificate is signed successful, application system can implement SSO based PKI.

IV. Implementation of SSO

SSO is an integrated part of campus network, user only need to provide a one-time credential and then can access all mutual trusted application system[6]. Now the best integration solution of application systems is based SOA that use Web Service to achieve the target. CAS is Central Authentication Service and a single sign-on protocol for the part of SOA. Its purpose is to permit a user to access multiple applications while providing their credentials only once. Please see Fig. 4, when a user accesses a site that uses CAS, that site redirects the user to CAS. Once CAS has verified a user's identity, it forwards them back to the original site. CAS attaches a unique ticket number to the URL of the protected service. The protected service sees this ticket. It sends this ticket to CAS. CAS tells the

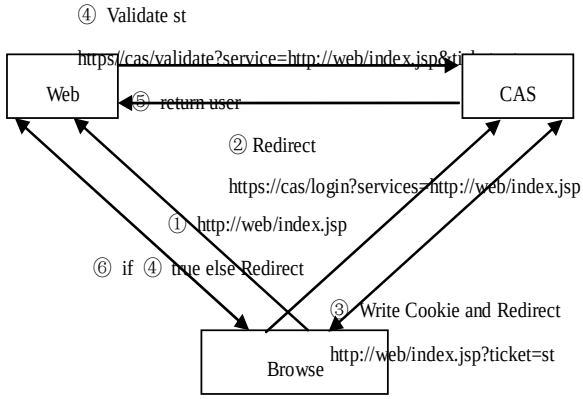


Figure 4. CAS authentication process of a figure.

protected service whether the ticket is good and if so, the Case ID that was used to obtain the ticket. The protected service reacts accordingly, allowing access if the ticket is good.

The extension of authentication is as follows:

A. Step 1

When user access a protected resource, CAS Client analysis of whether the request contains Service Ticket, if included, then skip to step C, if not, redirect to the CAS URLwith target address.

B. Step 2

If the user has certificate,system will pop up a dialog box to let user select appropriate certificate.and identify user.If user has no certificate,system will let user use traditional user/password method. After authorization authentication system will write TGC cookie and redirect target url whit service ticket.

C. Step 3

CAS client filter obtain http request,get the service ticket and post it to CAS, if valid, access applications, if not,skip to step A.

For the safety of TGC, accessing CAS use https, ticket only be use once. This paper presents dual authentication in CAS - certificate and LDAP. This requires extension of user authentication interface, CAS separate authentication interface from authentication protocol. The extended authentication provided by CAS is "AuthenticationHandler" which has support and authenticate methods where users implement their codes. In order to dual authenticate with certificate and LDAP, this paper implement "Credentials" and corresponding "CredentialsToPrincipalResolver".

V. CONCLUSION

With the development of campus network technology, independent application systems do not suite to direction of campus network. So SOA-based architecture using Web Service is the best solution for loosely couple up. At this time SSO is a part of it, login once to access the application systems of mutual trust, without second logon. Based on fact, this paper implement a SSO method using PKI, LDAP, CAS, etc and provide a security identity authentication. Users only login once can access many mutual trust systems.

REFERENCES

[1] Ma rongfei, "Research and Implementation of SSO," *Computer Engineering & Science,*Vol 31,Feb.2009,pp:145-149

[2] Ji min "Research and design of single sign-on scheme," *Computer Engineering and Design*,Dec..2009,pp:2861-2864,2914.

[3] Zeilenga, K. Named Subordinate References in Lightweight Directory Access Protocol (LDAP) Directories. RFC 3296, July 2002.

[4] Li xiaobao," A Supporting Multi-Mode Application Single Sign-On Scheme Based on PKI/PMI," *Journal of Beijing University of Posts and Telecommunications,* Vol 32,Mar.2009,pp:104-108

[5] Peng yinghui," Implementation of user authentication system based on PKCS#12 digital certification," *Computer Engineering and Design*,Aug,2009,pp:1840-1843.

[6] Jin weizu, "Solution Schema for Single Location Invalidation Based on CAS Cluster," *Computer Engineering*, Jan.2010,pp:51-54

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp.336-337

Computer Forensics System Based On Honeypot

Zi Chen Li , Xiao Jia Li , and Lei Gong
College of Computer Science and Technology, Henan Polytechnic University ,Jiaozuo,China
Email: {lrfgjj2, leigong0800}@163.com

***Abstract*—With the popularization of computer technology and the Internet, information security becomes increasingly important. The traditional passive defense has been unable to meet the needs of the people, honeypot technology as a proactive protection technology to make up for the traditional system. The paper gives a new Computer Forensics System based on honeypot which use the network deception, data control and data capture technology to achieve the network intrusion tracking and analysis.**

***Index Terms*—computer forensics, honeypot, Intrusion Detection**

I. INTRODUCTION

With the rapid development of Internet, human activities dependent on information networks are also growing. At the same time network security is tight, and the existing security measures is mainly based on the known facts of the passive protection model. Honeypot technology is an emerging network security based on active defense technology, which by monitoring the activities of an intruder, so that we can analysis of the intruder whose skills, using the tools and motivation for the invasion, thereby enhancing network security defense capacity. At the same time, honeypots can also use the custom features to phishing attacker, slow down the attack and the transfer target, effectively make up the traditional defensive deficiencies in information security technology, makes the protection system more perfect.

II. HONEYPOT WORKS

In short, the honeypot is a computer system running on the Internet which designed to lure and trick other people (such as hackers) who attempt to illegally break into others computer systems. Honeypot is mainly induced an attacker by using the network deception, makes the possible security vulnerabilities have very good camouflage place. Because honey can not provide real value to the outside service, all of its attempt to link will be considered as suspicious. Another use of honeypots is to delay the attack on the real target, make the attacker waste time in a honeypot so that the possibility of a real network services to be detected is greatly reduced and the network detection rapidly detect the attempt of the invader. Afterward, timely repair security vulnerabilities that may exist in the system and receive the enemy's offensive skills and intentions. Honeypot tools include sensitive monitor and event log. Event log to detect an intruder to access and collect information on the activities. Because any access to the honeypot system, the system is given the illusion of a successful invasion, so system administrators can not expose the system really working conditions, timely shift, record, track intruders, to collect electronic evidence, do a better computer forensics work.

III. ADVANTAGES AND DISADVANTAGES OF HONEYPOT TECHNOLOGY

Honeypot technology benefits include: the fidelity of data collection, honeypots do not provide any real effect, so the data collected very little. At the same time many of the data collected is as attacks by hackers, honeypots do not depend on the detection of any complex technology, thus reducing the false negative rate and false alarm rate. The use of honeypot technology can collect new attack tools and attack methods, unlike most current intrusion detection systems use feature matching method can only detect known attacks. Honeypot technology does not require strong resources to support, low-cost equipment can be use and it doesn't require extensive capital investment. Relative other intrusion detection technologies, honeypot technology is relatively simple, enables network administrators more easily to grasp some knowledge of hacking.

Honeypot technology also has some shortcomings, mainly: the need for more time and effort. Honeypot can only attack against the surveillance and analysis, the view is more limited, unlike the intrusion detection system can listen through the bypass techniques to monitor the entire network. Honeypot technology can not be directly protective vulnerable information systems. Honeypot deployment will bring some security risk.

IV. SYSTEM DESIGN

A. System Model

Computer forensics model is the theoretical basis of forensic system. Based on the existing model into the honeypot technology, (Figure 1) gives the model of

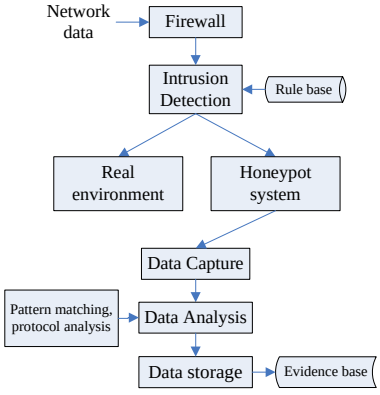


Figure 1 Model of Computer Forensics System

computer forensics based on honeypot technology. The system has four modules, intrusion detection module, data capture module, data analysis module, data storage module.

B. System Analysis

To address the need for protection for each machine are installed honeypot system costs too many problems, you can use technology to simulate virtual HoneyPot multiple systems in order to attract more attacks or intruders.

Virtual honeypot system is in a real physical machine to run some simulation software, simulation software to simulate on computer hardware, makes the simulation platform can run multiple different operating systems, such a machine becomes true multiple hosts (known as virtual machine). Virtual honeypot technology can be used to simulate the multiple systems in order to attract more attacks or intruders, software can set up a virtual honeypot system.

C. System Implementation Technology

1) intrusion detection module: Implementation of the network packet authentication. Intrusion detection system with rule base (IP or MAC) in comparing the rules for credible data packets allowed into the real system, suspicious packets redirected to the intrusion deception environment.

2) data capture module: Data acquisition functions. Packet contains a record of the intruder's actions, these records will eventually help us to analyze their use of tools, strategies and attack purposes. Forensics system to collect as much as possible all available data, and ensure that these data have not been tampered with, it needs the data tansmit to Remote Security Host. We use various means to make the honeypot system to collect data integrity and security as much as possible, through a combination of several methods, it is clear replay attack the intruder. The first record is a host firewall tool. It can record all incoming and out honeypot system connections. Not only can we set the firewall to log all the connections, but also to give us warning messages. In addition, it can record some unusual port connection attempts. The second recording tool is intrusion detection system, we use Snort, configured in Linux host. It has two functions: The first role is to capture all differences in honeypot system of network data packets. In addition, it also can found some suspicious behavior and to alert you.

3) data analysis module: Realize the characteristics of network data packets. Analysis of performance of the system determines the overall system performance. Therefore, it can take pattern matching and protocol analysis method to improve the analysis of system performance. Protocol analysis use the network protocol level and knowledge of relevant agreements quickly determine whether there are signatures, Thus greatly reducing the computational pattern matching to improve the accuracy of matching. Pattern matching is based on the signatures of network packet analysis technology. Its analysis speed, the advantages of small false alarm rate is unmatched by other analytical methods. Simple to use pattern matching, there are big drawbacks, we use the combination of protocol analysis and pattern matching methods to analyze network data packets.

4) data storage module: Realize the data transmission and preservation. Network data packets are recognized to be safe for the invasion of the transfer of data to secure evidence of machine to prevent tampering by an intruder.

D. System Implementation

According to the system structure, we can implement a system of Intrusion Deception as following steps:

1) Configuration firewall and intrusion detection systems.

2) In the server install VMware virtual machines to construct intrusion deception environment, then install a honeypot system in the virtual machine.

3) The establishment of legal rules on the server database.

4) Configuration computer for data Analysis.

5) Configuration Forensics machine for receiving the data.

V. CONCLUSION

Honeypot technology make network security shift from passive to active defense, tracking of the intruder Undeniable. Compared to other security mechanisms, honeypot easy to use, flexible configuration, occupies less resources can be effective in a complex work environment, collecting data and information relevant of a good value. With the intrusion type of diversification, the honeypot must also be a variety of interpretations, otherwise it will not be able to face the ravages of the invaders.

REFERENCES

[1] Lance Spitzner.Definitions and Value of Honeypot.[EB/OL]

[2] Michael Howard,David LeBlanc,Writing Secure Code,Microsoft Press,2002

[3] http://www.honeynet.org.

[4] Honeynet Project.Know Your Enemy GenII Honeynets.Http://www.honeynet.org/papers/gen2.2003

[5] Rajeev Motwani and Prabhakar Raghavan,Randomize Algorithms,C-Ambridge University Press,1995

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 338-341

Feature Extension for short text

Yan Tao[1]. Wang Xi-wei[2]
[1]Henan University of Urban Construction,Network Information Center,Pingdingshan
abeey2007@gmail.com
[2]Henan University of Urban Construction,Network Information Center,Pingdingshan
wangxw@hncj.edu.cn

***Abstract*—Different from the conventional word-form based automatic classification system of chinese texts，giving further consideration on words co-occurrence relationship，this paper proposes two feature extension methods based on co-occurrence relationship．The improved methods give higher accuracy to the short text classification system．**

***Index Terms*—short text classification, co-occurrence relationship, features extension.**

I. INTRODUCTION

The development of instant messaging technology and the popularization of information processing technology promoted the booming of short-text information processing technology, such as the mobile phone SMS, QQ chat, BBS, instant messaging software,has become an important channel for information dissemination.Such a rich resource of short texts make people's lives easier while bringing significant information security risks. Such as waste, harassment and frequent large quantities of text messages, seriously affecting people's lives[1], and short text classification is the realistic tasks basis to solve the short message filtering, to promote the chinese short-text classification has become an important research direction.

Most methods of the short text is mainly traditional text classification, information filtering and retrieval methods, Specific algorithm for short text has not yet formed on its own characteristics. However, compared with the long text, because of the characteristics that short text described weak signals, noise characteristics of the data ,and automatic classification system of Chinese texts based on simple word-form have been unable to meet the needs of short text classification. Therefore, the comprehensive consideration of the short text data is proposed in the course of a short text classification, mining the association relationship between the short text data to assist the classification. Currently,only a small number of research at home and abroad in this areas[2],[7] , and the results are unsatisfactory.

Previous studies show that the text feature rich or not is essential to the classification results. So, two methods are proposed for feature extension based on taking full account of the correlation between words: In this paper, take the training data as the background corpus, firstly, using FP-Growth algorithm to mine the co-occurrence relationship among the training set, and to construct the set of feature co-occurrence as expansion vocabulary, and then expanded the training and testing features using the set of feature co-occurrence respectively, the two methods were based on the the same expansion vocabulary, but expand in different ways. Finally,experiments were carried out respectively.

II. BASIC PROCESS

Expansion based on the training and testing text feature processes as shown in Figure 1 and Figure 2:

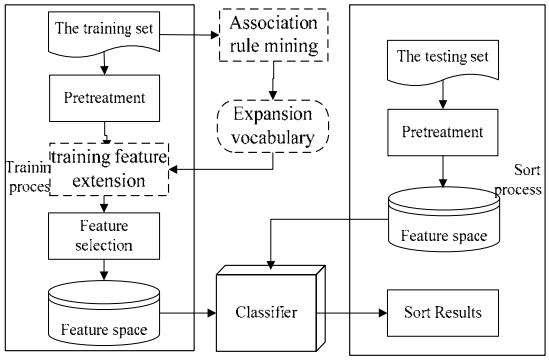


Figure 1 chinese short-text classification based on training set feature extension

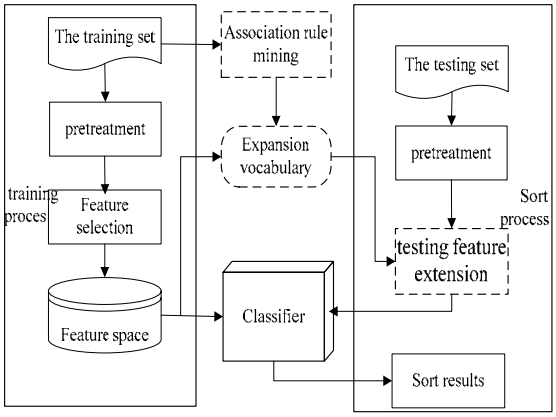


Figure 2.1 chinese short-text classification based on testing set feature extension

A. *The construction of the set of feature co-occurrence*

Agrawal proposed Association Rules is to find relationship between different data items，such as data influence of another data item. And found that rules it is conducive to data classification，can resolve the problem of exploring knowledge in short-text to some extent.

One of the most critical technology in feature extension is that the construction of extended vocabulary, data sources of extended vocabulary are usually two ways[8]:

The first is the automatic construction of machine resources (such as unlabeled test data and the background corpus, etc.);

The second is a specialist construction resources (such as the existing language knowledge base, etc.). Such as WordNet, HowNet, etc.;

support for the itemset X is $\sigma(x)$ in data mining, focus of all services including the number of items X， Suppose s is the minimum support threshold, if $\sigma(x) \geqslant$s,then it is frequent itemsets. feature with frequent co-occurrence relationship means that they appear in the same document with higher probability. Such as "telecommunications and company", " country and nation" and so on.

According to the contents of my study, the following two definitions are proposed:

Definition 1 co-occurrence word pairs : Association Rules $t_i \rightarrow t_j$ before and after entry form the co-occurrence word pairs.

Definition 2 the set of feature co-occurrence : if the latter are features in the co-occurrence word pairs,they form it.

In this paper ,we chose the first method of expansion vocabulary. The training data as the background corpus, using FP-Growth algorithm[9] to mine the co-occurrence relationship among the training set, and to construct the set of feature co-occurrence as expansion vocabulary.

The creation of the set of feature co-occurrence includes two points:one is the creation of the co-occurrence word pairs,another is the check of features. The Figure 3 is FP-Growth Extracted the set of feature co-occurrence.

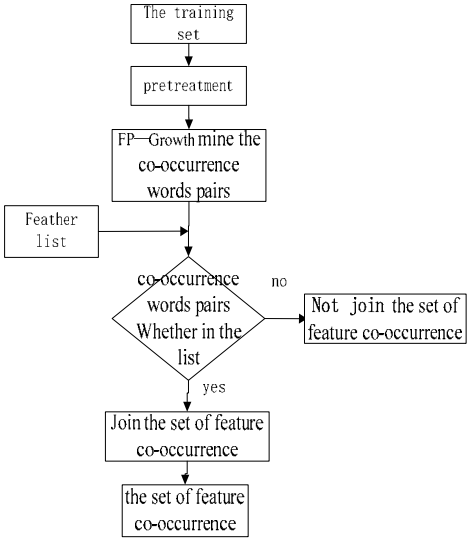


Figure 3 FP-Growth Extracted the set of feature co-occurrence

As the number of training samples of each class different, to ensure that features are set to cover a total of all categories, We calculated the set of feature co-occurrence of each class .

*1) Data pre-processing:*There are a lot of short text as the high-frequency words, but meaningless, filtering out the noise interference using chinese stop word list[10], retain only the core of chinese sentences[11], including nouns, verbs, adjectives, adverbs.

*2) Using FP-Growth for mining co-occurrence word pairs:*Input the results into FP-Growth algorithm, according to the minimum support threshold for first pass screening, then follow the minimum confidence threshold generated rules, the frequent co-occurrence set of rules generated for each class,that is set of co-occurrence word pairs.

3) Feature words check:check the latter of the co-occurrence word pairs, if it is feature ,remain it,get the final set of feature co-occurrence.

B. Feature Extension method of Training set

We believe that, once the first item which between the strong association rules appears in the text, the latter will occur to a certain probability. Specifically in the characteristic extension of the training set , the training feature is expanded according to the set of feature co-occurrence firstly, owing to co-occurrence words in training set, we only need to adjust the weight of characteristics, without new features.

Traditional automatic classification system based on simple word-form have been unable to meet the needs of short text classification, often leads significant terms to lower weight. To the method feature extension of training set, hypothesis, the first item which between the strong association rules appears in the text, the latter will occur to a certain probability, we will enhance the frequency of the latter on the basis of the number preceding paragraph and the co-occurrence probability, highlights of the contributions which interaction between co-occurrence words to the weights of the characteristics and classification.

For a word pairs $t_i \rightarrow t_j$, improve the latter characteristic weights calculated as:

$$W t_j = W t_j \bullet \left(1 + \frac{W t_i \times S \times C}{W t_j}\right) \quad (1)$$

Among them, Wt_i is the weight of t_i, Wt_j is the weight of t_j, S is the support for the co-occurrence word pairs $t_i \rightarrow t_j$, C is confidence. This formula taking the effect between frequency, support and confidence to the weight of the feature, giving further consideration on words co-occurrence relationship, solve the mere limitations based on word frequency.

Algorithm for training text feature expansion:

Take k-class for example($1 \leq k \leq 12$),

Input: feature co-occurrence set I_k;

Association rule threshold: Minimum confidence threshold C;

Minimum support threshold S;

Word frequency file of training set star.txt;

Output**:** Word frequency statistics of training set star_.txt;

Step 1 For a feature t_i in star.txt, inquiry I_k, if the co-occurrence word pairs $t_i \rightarrow t_j$ on the unique, And when C is greater than the threshold, run Step 2.if not unique,

computing S * C *(the number of the latter), extend t_i according to the term which have the maximum results, then run Step 2. If there is no co-occurrence word pairs , run step 3;

Step 2 Modify the weight of t_j according to the formula (1).

Step 3 Not extend t_i.

C. *Feature Extension method of Testing set*

Feature extension of testing set :Firstly, the feature' co-occurrence word is added as new feature according to the set of feature co-occurrence in the classification stage, and then classified.

We believe that, once the first item which between the strong association rules appears in the text, the latter will occur to a certain probability. Specificaily, in the characteristic extension of the testing set , once a feature words appear, add another feature of the relationship between two words.

We first clear the two concepts[8]:

(1) Short text concept words: Refers to the word as a verb, noun, adjective, or adverb phrase.

(2) Feature words set: Extracted from the training set focused on that part of speech as a verb, noun, adjective or adverb words, a subset obtained by feature selection.

Algorithm for testing text feature expansion:

Algorithm description:

Input: Feature co-occurrence set $I_;$

Association rule threshold: Minimum confidence threshold C;

Minimum support threshold S;

Test document;

Output: Feature space after feature expansion;

Step 1 For a concept words t_i in test document , inquiry $I_,$ if the co-occurrence word pairs $t_i \rightarrow t_j$ on the unique, And when C is greater than the threshold, run Step 2.if not unique, computing S * C *(the number of the latter), order the value of S * C *(the number of the latter),extend t_i according to the term which have the first three maximum results, then run Step 2. If there is no co-occurrence word pairs , run step 3;

Step 2 Extraction the t_j obtained in step 1,if t_j not in feature space,run step 3; If not,run step 4;

Step 3 Add t_j in the list of feature space;

Step 4 Not add t_j in the list of feature space;

Step 5 If you can not find the matching program,run step 6;

Step 6 Not extend t_i.

III. EXPERIMENT AND RESULT ANALYSIS

Datasets used in this article is collected 470252 customer comments from 12 different areas by our teams,among those comments 35104 were from the area of Finance and Economics, 28744 from Real estate, 42424 from International News, 48288 from Domestic News, 49320 from Military, 37044 from Technology, 36032 from Women, 40372 from Automobile, 39440 from Book Review, 38512 from Sports, 38660 from Games, 36312 from Entertainment. each type of texts will be randomly divided into four, A testing set, the other three are training sets.

D. *Comparison of three methods of classification performance*

1、Conventional method(Called method 1 for short): Developed by the Chinese Word Segmentation CsegTag3.0 of Tsinghua University, And remove stop words, expressed short text as a vector use tf-idf, using CHI select features， use Naive Bayes (Naive) as classifier[12] , test text Not been extended, selected features by 1000,2000,3000 ... ... 10000, try 12 categories short text classification experiments by 10 cycles.

2、Feature extension method of training set(Called method 2 for short):Extend training set based on method 1, selected features by 1000,2000,3000 ... ... 10000, try 12 categories short text classification experiments by 10 cycles. support and confidence threshold values were set to 0.3%, 2%.

3、Feature extension method of testing set(Called method 3 for short): Extend testing set based on method 1, try 12 categories short text classification experiments by 10 cycles. support and confidence threshold values were set to 0.01%, 0.5%.

The results shown in Figure 4 and Figure 5.

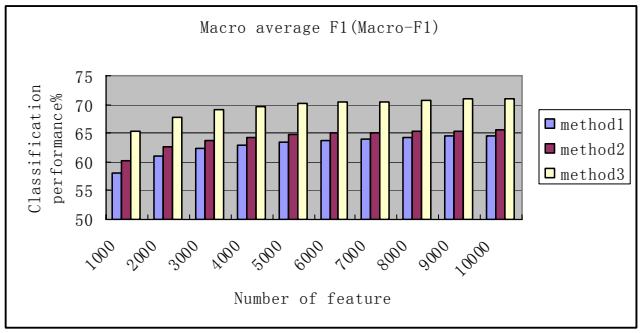


Figure 4 The macro average F1 value of three methods compared

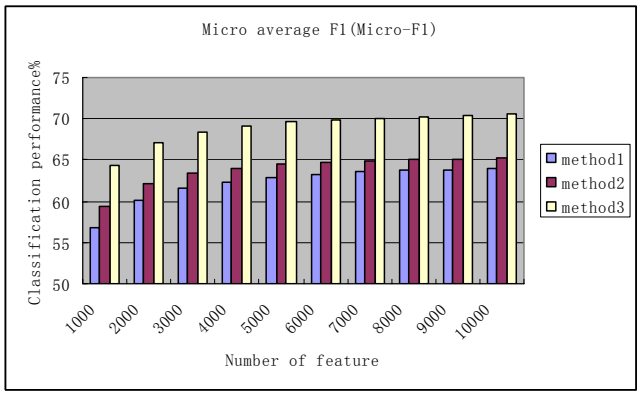


Figure 5 The micro average F1 value of three methods compared

Analysis:

(1) With the number of features increases, the classification performance of three methods increased gradually.

(2) Method 2 and Method 3 for a reasonable feature expansion, reduce the noise synonyms, class distinction information more complete , since both of its Macro-F1 and Micro-F1 the are higher than those of method 1, specially ,Method 3 show the best performance.

IV. CONCLUSION

This paper introduced the idea of feature extension, discussed association rule mining apply for text classification, proposed a method of feature extension for short text. because of the characteristics that short text described weak signals, noise characteristics of the data, in this paper, the training data as the background corpus, Firstly, using FP-Growth algorithm to mine the co-occurrence relationship among the training set, and to construct the set of feature co-occurrence as expansion vocabulary, and then expanded the training and testing features using the set of feature co-occurrence respectively, the two methods were based on the the same expansion vocabulary, but expand in different ways. Finally,experiments were carried out respectively.Studies show that the method in the multi-class short-text classification is feasible and effective, and has better performance as well as accuracy than the traditional classification.

REFERENCES

[1] 12321 Bad and the spam network report and Reception Center. the second mobile phone short message status report,2008, [EB/OL].http://www.12321.cn/viewnews.php?id=10753.

[2] Healy, M Delany, S, and Zamolotskikh, A. An Assessment of Case Base Reasoning for Short Text Message Classification[C]. In: Norman Creaney (ed.) Proceedings of the 16th Irish Conference on Artificial Intelligence & cognitive Science (AICS'05), 257-266, 2005.

[3] Zelikovitz, S and Marquez, F. Transductive Learning for Short-Text Classification Problems using Latent Semantic Indexing[J]. International Journal of Pattern Recognition and Artificial Intelligence, Vol.19(2), 143-163, 2005.

[4] Zelikovitz, s. Transductive LSI for Short Text Classification Problems[C]. In: Proceedings of the 17th International FLAIRS Conference, 556-561, 2004.

[5] Zelikovitz, S, Hirsh, H. Improving Short-Text Classification using Unlabled Background Knowledge to Assess Document Similarity[C]. In: Proceedings of ICML-2000, 1180-1190, 2000.

[6] Qiang Pu, Guo Wei Yang. Short-Text Classification Based on ICA and LSA[C]. In: Proceedings of International Symposium on Neural Networks, 2006(ISNN2), 256-270, 2006.

[7] Shen-zheng Zuo, Chun-hua Wu, Yan-quan Zhou, Hua-Can He. Chinese Short-Text Categorization Based on the Key Classification Dictionary Words[J]. The Journal of China Universities of Posts and Telecommunicationa, Vol.13(s), 47-49, 2006.

[8] Fan xing-hua. Chinese short-text classification based on Feature association[Z], Application for National Science Foundation,2008.

[9] Han Jia-wei, Pei Jian, Yin Yi-wen. Minning Frequent Patterns Without Candidate Generation[C]. In: Chen Wei-dong, Jeffrey F M, Philip A B. Proceedings of the 2000 ACM Sigmod Internal Conference on Management of Data. Dallas, Texas: ACM Press, 2000.1-12.

[10] Chinese stop word list [EB/OL]. http: //download.csdn.net/source.

[11] Wang yuan-zhen, Qian tie-yun, Feng xiao-nian. Association Rules Based Automatic Chinese Text Categorization [J]. M INI-M ICRO SYSTEMS, 2005, 26(8): 1380-1383.

[12] Wu wei. Classification of large-scale filtering of short text[D]. Beijing University of Posts and Telecommunications. 2007.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 342-345

Research of Application Model about Handset based on OSGi Service Platform

Ao Shan [1,2] , Dai Jian-hua[3]
[1] School of Computer Science and Technology Henan Polytechnic University, Henan , China
[2] Insitute of Education Economy Peking University, PKU ,Beijing , China
Email: sao@gse.pku.edu.cn
[3] The Media Management School Communication University of China, CUC,Beijing , China
Email: daijianhua66@gmail.com

***Abstract*—The OSGi (Open Service Gateway initiative) service platform to be applied to handheld devices, put forward the application requirements on handheld devices, and designed to achieve the framework of handheld devices, including a virtual machine environment on the operating system, OSGi Service Platform Framework and Bundle of application carrier. On this basis, this paper designed the application development model, proposed application-oriented service implementation mechanism and service registration model, designed the registration model about two types of services entities, at this time, realize the related application development and test instances.**

***Index Terms*—OSGi Service Platform, Handset Framework; Bundle**

I. INTRODUCTION

With the market demand increases for the intelligent of embedded devices, the development oriented handheld device software is bound to computing, communications, networking, storage, entertainment, e-commerce, and other multi-functional integration. Feature pursuit of humanity, respect for wireless Internet applications, focus on safety performance will be the standard of consumer choice; for the business, how to produce low-cost, personalized configurations, service complete product, is the ultimate pursuit [1]. Moreover, the current hand-held devices are no longer "information islands." Requirements for external networking, equipment with the communication interface is necessary, the corresponding need for TCP / IP protocol stack software support [2]. As household appliances interconnected and coordination of field devices, etc., need a new generation of handheld devices with IEEE1394, USB, CAN, Bluetooth and other communication interface. At the same time, this needs to provide the appropriate communication network protocol software and physical layer driver software.

The OSGi specifications[3] based on the application, which implements all of the access service for network equipment, equipment, shielding the complexity of different businesses [4]. This allows service providers to update and expand the service. For service providers only provide services to the user interface, so no need to modify the service user can obtain the latest service. This greatly simplifies the service development life cycle, reduce costs, and increase the effective life of equipment in the network world.

II. OSGI SERVICE PLATFORM APPLICATION FRAMEWORK FOR HANDHELD DEVICES

From the view point of OSGi Service Platform applications in handheld devices, we carried out requirements design and system architecture definitions. Handheld devices will achieve two requirements: on the one hand, hand-held device management all kinds of services on OSGi service platform, on the other hand, remote server management services for handheld devices.

A. Requirements Design

Handheld device manage all kinds of services of OSGi Service Platform. Equipment can be installed on the default system services, users can also independently control the life cycle of the authorization service.

Users can install the Services, which meet the OSGi service specification, on Framework, the user need only specify the specific installed path; Users can run the service on Framework, the services can registered to the Framework, and be used by other services on OSGi

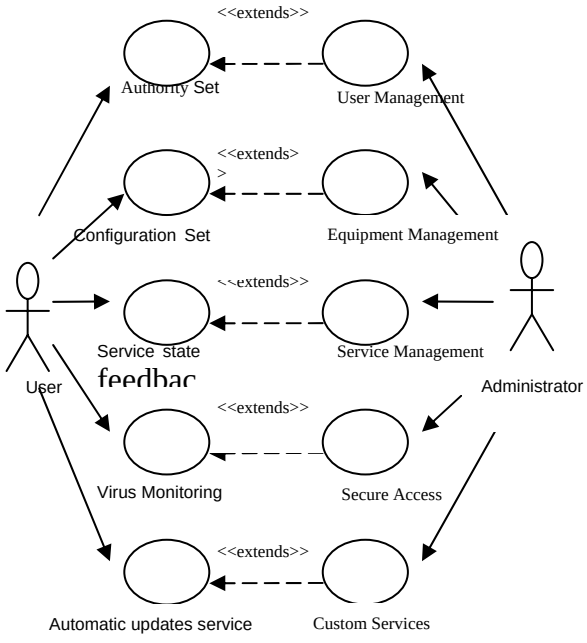


Figure 1 hand-held devices case diagram

AP-PROC-CS-10CN007

Service Platform

The user can stop the service, cancel the services which has been registered in the service Framework; user can uninstall the service from the Framework; Users can also update service, which has been installed by the specific path specified in the Framework.

After handheld device access network, it will be managed through the management proxy server, including authorization schemes, security access, dynamic monitoring and so on. Device management proxy server is a proxy of Remote server. it carries out management to equipment, service and user. Figure 1 is use case diagram.

B. System framework

Figure 2 is the environmental framework for OSGi Service Platform, OSGi Service Platform provides a common realization environment. a variety of Driver is on the Hardware; Java virtual machine is on Operationg System; OSGi Framework is on the Java virtual machine, It registered a variety of Service, run all kinds of Bundle, as well as "Bundle" form of realization of various Application. Framework is a horizontal extension. Achieving of bundle not only based on Java virtual machine implementation, but also can call the system's Native Code (non-Java code on the operating system) [5].

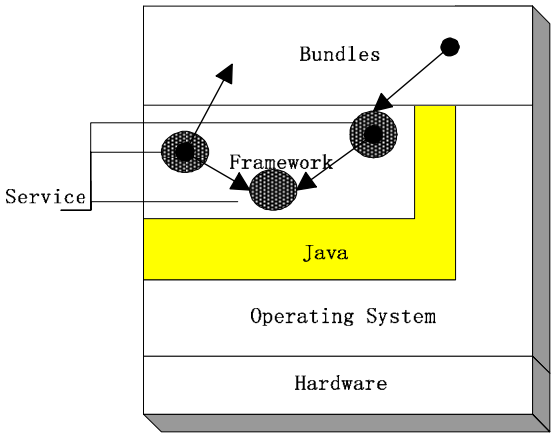


Figure 2 implement framework for OSGi Service Platform

Realization structure of handheld devices: a common operating system for handheld devices on hardware, including PlamOS, Linux, WinCE, Symbian, etc.; a special Java virtual machine, consumer-oriented micro-package devices is on operating system,;

on the Java Virtual Machine is extended over a variety of specific implementation for handheld devices: CDC (Connected Device Configuration), CLDC (Connected Limited Device Configuration), PJava (Personal Java), CVM (Component Virtual Machine); OSGi Framework is implemented on top of the Java virtual machine, and provide a variety of execution environment for Bundle. Above Framework is the application components layer and service registry layer, application components layer provide user-oriented applications (Application), service registry layer provides application-oriented services (Service), to implement dynamic service registration and Foreign expansion.

OSGi's Framework as a system Bundle, with the following special properties: It is set Bundle ID, the default value is 0, that it is the first to be launched Bundle. It cannot be managed as life-cycle, only to start, stop, update and uninstall the state; It manages all the Bundle of other activities, so it must be the path to the external leads, such as general Bundle be stated in the Manifest file. on the platform with OSGi service specification, Framework dynamically installs and updates Bundle, and manage all the Association between undle and Service. BundleA, BundleB, BundleC can respond to events, Framework achieve life-cycle management to these Bundle.

III. APPLICATION DEVELOPMENT MODEL DESIGN BASED ON OSGI SERVICE PLATFORM

From the user-oriented perspective, in the OSGi Service Platform Framework, we design registration application component layer and service layer. Application component layer is from the application component interface definitions, application components to achieve qualification, application components defined API (application program interface function that is defined) and application components related to design; service registry layer is from the service implementation model and service registration mechanism to design [6] [7], we proposed two service registration mode.

A. The application component layer

In the application component layer, various of application component interfaces is the most important. A software component to deployment is as the form of Bundle. API is the interface of application program, it defines the entrance and exit of this component.Applications based on OSGi Framework implementation, can be start, stop, pause execution and continue. Usually the application is associated with certain user interface. In the OSGi service platform, the several of the collection according to the type of application, are application container. As soon as application container is installed successfully, this application representation is registered to the platform

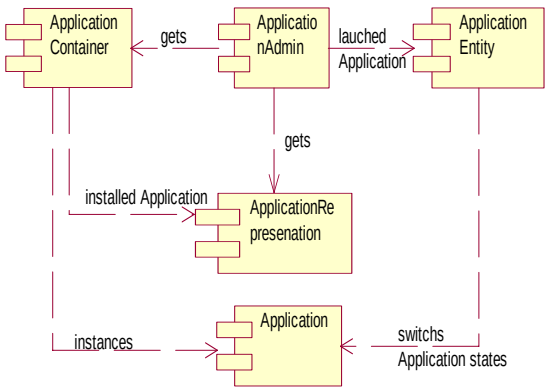


Figure 3 The relationship between the application components

service registration. Application representation includes the type of application and the application of the additional information. Figure 3 shows relationship among the application, application container, application entity, the application representation and application administer.

B. Service registry layer

Framework services Running in the OSGi Service Platform is accessed by other Bundle on the form of interface, not only to ensure Service truly independent, but also guarantee the security to code. Service A, Service B and Service C register to the Registry by Bundle A, Bundle B and Bundle C respectively. Bundle D obtain these services only through access service interface. Figure 4 is a concrete implementation example.

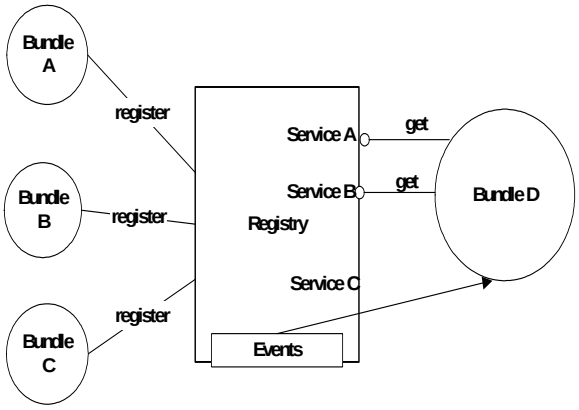


Figure 4 Service Implementation Pattern

C. Service Registry Model

This paper designed two different entities registered service mode: Mode 1: Figure 5 presents a Service registration and call patterns. BundleA and BundleB register Framework, which register the respective service entities Service Object to the Registry directly, through ServiceManagement distribution, is stored to the Framework, the only place. As an application, BundleD can direct access to services entities through ServiceMangament address.

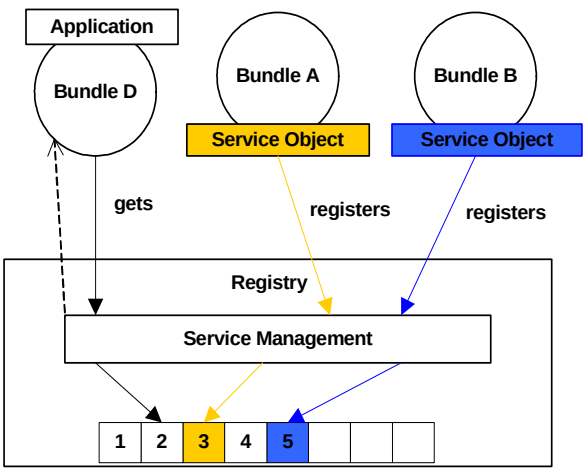


Figure 5 Service registration Mechanism: Mode 1

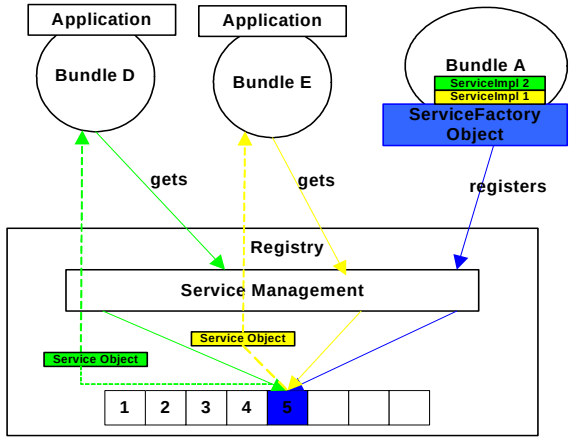


Figure 6 Service registration Mechanism: Mode 2

Mode 2: Figure 6 shows another service registration and call patterns. using ServiceFactory mechanism, different service entities of BundleA register to the Registry as unified ServiceFactory form. ServiceManagement make the distribution of addresses for registered ServiceFactory, but does not distribute addresses for the service entities. Only when Application in BundleD or BundleE obtain services, ServiceFactory.getService achieved by overloading BundleContext.getService to create the calls for service entity. Service entity is created timely by ServiceFactory only when it is acquired. This fit to register services group with more correlation, but also for the service to registration, which does not require permanent memory.

IV. DEVELOPMENT AND TESTING BASED ON APPLICATION DEVELOPMENT MODEL

Based on the above application development model, we implements an application example on DeviceTop3.0 platform. on the instance: the Framework of DeviceTop3.0 is started on PDA, PC access PDA via Telnet, and the applications are installed on the Framework, Framework manage applications. PDA display application information (see Figure 7), is starting ProductInfo application.

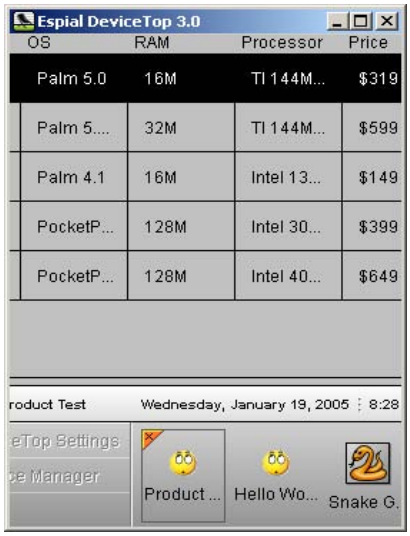


Figure 7 PDA display ProductInfo

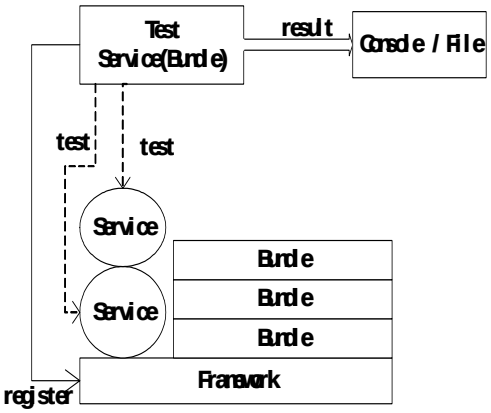


Figure 8 testing diagram for the service platform

OSGi Service Platform is a runtime environment, it can track and test running services on the system. Basic test strategy is: to registry testing service in the Framework, the service obtain testing services through the Bundle Context, so it will enter the test track, the test results can be saved directly. So as to resolute the performance testing issues about the dynamic service registration. Figure 8 is a test diagram. Test results show that the design of handheld devices to achieve better running in parallel for multiple Java programs. Framework can dynamically manage applications. Applications and services truly are separated. DeviceTop3.0 is implemented on the basis of the OSGi specification Release2.0, the extended system can successfully boot a certain port, to accept remote management of services manage the system's dynamic update, repair and send.

V. INCLUDING

Java technology is mainly adopted on Implementation of OSGi Service Platform. It is benefit of portable, high reliability, high security and multi-protocol support of Java2 technology. Embedded device is based on OSGi services management platform, it provided to users for the service platform including a service-oriented, safe and reliable, network sharing, configurable, remote management, and easy to operate. Embedded device manufacturers will be isolated from OSGi services to specialized service providers, together with the information service operators, network service operator jointly developed the intelligent network services. This article OSGi (Open Service Gateway initiative) service platforms applied to the handheld device, design and implement handheld device framework, including the a virtual machine environment on operating system, Framework and carrier Application Bundle on OSGi Service Platform, proposed for application-oriented service implementation mechanism and the service registration mode, implement two different registration mode for service entities.

REFERENCES

[1]. Zhang Shi and Huang Lin-peng, Dynamic service evolving based on OSGi, Journal of Software, 2008,19(5), pp:1201-1211

[2]. Song Ya-li and Tang Xiao-sheng, Design and implementation of smart home system based on OSGi home gateway and web service, Journal of Computer Applications, 2007, 27(6), pp:1542-1544

[3]. OSGi Alliance. About the OSGi Service Platform Technical Whitepaper Revision 3.0. http://www.osgi.org/osgi_technology. July 12, 2004

[4]. Tao Gu, Hung Keng Pung and Da Qing Zhang. Toward an OSGi-Based infrastructure for context-aware applications. IEEE pervasive computing, October-December 2004 (Vol. 3, No. 4), pp:66-74

[5]. Choonhwa Lee, David Nordstedt and Sumi Helal. Enabling smart spaces with OSGi. IEEE pervasive computing, July-September 2003 (Vol. 2, No. 3), pp:89-94

[6]. Lee Seungkeun, Kim Iniae and Rim Kiwook. Service mobility manager for OSGi framework.2006, pp:21-29

[7]. Feng zhi-yu and Huang lin-peng, Two-tier service-oriented model based on OSGi, Application Research of Computers, 2009,26（7）pp:2590-2597

[8]. Ta Li-juan, He Liang and Gu Jun-zhong, Design and implementation of a home service gateway supporting OSGi, Computer Applications and Software, 2008, 25（3）pp:186-188

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 346-348

Application of Surrounding Rock Stability Classification Based on Fuzzy ClusteringZhu

Changxing [1], Wang Fenge[2]

[1] College of Civil Engineering Henan Polytechnic University Jiaozuo, China
Email: zcx7685@yahoo.com.cn
[2] College of Computer Science Technology Henan Polytechnic University Jiaozuo, China
Email: wangfe@clc.hpu.edu.cn

***Abstract*—Fuzzy clustering is a scientific and effective clustering method, which is applied surrounding rock stability classification of underground engineering by compiling fuzzy clustering algorithm of matlab software. Surrounding rock classification data collected are defined as training and forecasting samples. The research results indicate that fuzzy clustering algorithm be better used surrounding rock classification of underground engineering.**

***Index Terms*-fuzzy clustering, clustering method, surrounding rock, classification**

I. INTRODUCTION

In recent years, with the rapid development of geotechnical engineering at home and abroad, more and more irrigation works, transportation, energy and defense projects have started in some areas. Surrounding rock classification is considered an important problem of foundation research of underground engineering. Whether evaluating result of surrounding rock stability is right or wrong, it immediately affects construction of underground engineering. Presently, many estimate methods of surrounding rock stability is in the widespread use, for example, RQD classification, Q system classification, RMR classification, etc.. Some scholars adopt new theories and methods to study the classification of surrounding rock, but fuzzy clustering is rarely applied surrounding rock classification. How to utilize known information to guide the clustering process is a research scope because of many samples known information known in the real life. Firstly, you need know sample sorts and properties in the similar questions; secondly you can classify these samples unknown according to known information. The author applies a scientific classification method based on depicted view above.

II. FUZZY CLUSTERING ANALYSIS

Fuzzy clustering proposed by DUNN and generalized by Bezdek, divides data sets $X=\{X_1,X_2,\cdots,X_n\}\subset R^{p\times q}$ into c categories, where the random sample x_i belongs to class i with probability u_{ij} .If X sample collation is divided into c categories, then n samples belong to a subjection degree of c category that is wrote following subjection function matrix:

$$u_{ij}=\begin{bmatrix} u_{11},u_{12},\cdots,u_{1n} \\ \ddots \\ u_{c1},u_{c2},\cdots,u_{cn} \end{bmatrix} \quad (1)$$

The classification results are represented by a fuzzy member-ship matrix $U=\{u_{ij}\}\in R^{cn}$ satisfying the conditions shown in formula(2)(3)(4).

$$\sum_{i=1}^{c}u_{ij}=1,\quad 1\le j\le n \quad (2)$$

$$0\le u_{ij}\le 1,\quad 1\le i\le c,1\le j\le n \quad (3)$$

$$0<\sum_{i=1}^{n}u_{ij}<n,1\le i\le c \quad (4)$$

Fuzzy C-means clustering is achieved by minimizing the objective function $J(X,\mu,\upsilon)$ that is about fuzzy membership matrix U and cluster center V .Function $J(X,\mu,\upsilon)$ is defined as formula(5):

$$\min J(X,\mu,\upsilon)=\sum_{i=1}^{c}\sum_{j=1}^{n}u_{ij}^{m}d_{ij}^{2} \quad (5)$$

Where $U=\{u_{ij}\}$ is a fuzzy membership matrix and meets all conditions defined in formula (2) (3) (4). $V=\{v_1,v_2,\cdots,v_c\}\in R^{pc}$ are focal points for clusters collation, $m\in[1,\infty]$ is the weight index. The research results of Nikhil showed that the best value range is 1.5-2.5 and the ideal value for m is usually equal to 2. The distance between the Kth sample and the center of the ith class is defined as formula (6):

$$d_{ij}=\|p_i-X_j\|,1\le i\le c,1\le j\le n \quad (6)$$

AP-PROC-CS-10CN007

Where the distance d_{ij} defined by formula (6) is Euclidean distance. FCM algorithm is repeated iteratively to optimize the objective function (5). We detail the FCM algorithm in the following part.

Step 1: Initialize the cluster center. Here, is initialized according to the matrix obtained in DSOM initial clustering.

Step 2: Calculate the membership matrix according to formula (7):

$$u_{ik}^{(b)} = \left\{ \sum_{j=1}^{c} \left[\left(\frac{d_{ik}^{(b)}}{d_{jk}^{(b)}} \right)^{\frac{2}{m-1}} \right] \right\}^{-1} \tag{7}$$

Step 3: Update the cluster center by formula (8)

$$v_i^{(b+1)} = \frac{\sum_{k=1}^{n} \left(u_{ik}^{(b+1)} \right)^m x_k}{\sum_{k=1}^{n} \left(u_{ik}^{(b+1)} \right)^m}, i = 1, 2, \cdots, c \tag{8}$$

Step 4: Repeat steps (2) (3) until the formula (6) converged.

III. EAPPLICATION OF SURROUNDING ROCK STABILITY CLASSIFICATION BASED ON FUZZY CLUSTERING

A. Choice index of surrounding rock classification

Many factors affect surround rock stability, which mainly include: the properties of rock mechanize and rock structure and structure plane, geostatic stress, groundwater, time and so on. The index of surrounding classification is mainly determined by experience, but surrounding rock stability conditions of different engineering have much difference. Classification parameters of surrounding rock are considered while the corresponding classification criteria are often formulated. According to influence factor of surrounding rock depicted above, some important parameters are selected, namely, the rock quality index RQD; uniaxial saturation compressive strength of rock R_W , integrity coefficient K_V , structural plane intensity coefficient K_f ; groundwater quantity of percolation Q_w. Specific classification indexes shown table 1.

Table I The classification indexes of surrounding rock stability

number	$RQD(\%)$	$R_W(MPa)$	K_V	K_f	$Q_w/\left[L\cdot(\min\cdot m)^{-1}\right]$	surrounding rock sort
1	>90	>120	>0.75	>0.80	<5	I
2	90~75	120~60	0.75~0.45	0.80~0.6	5~10	II
3	75~50	60~30	0.45~0.30	0.6~0.4	10~25	III
4	50~25	30~15	0.30~0.20	0.4~0.2	25~125	IV
5	<25	<15	<0.20	<0.2	>125	V

B. Engineering application

In order to prove the method that is right ,13 samples data collected from literature and are defined as training and predicting samples (as shown in tab.2).

Fuzzy clustering program is compiled by the matlab software. The former 9 samples are trained; also the later 5 are predicted. Research result testifies that method adopted is irrational (as shown in tab.3).

Monitoring and predicting result shown table2,3, comparing monitoring result with predicting result known. Predicting method is right.

Table II Monitoring parameters of surrounding rock

number	$RQD(\%)$	$R_W(MPa)$	K_V	K_f	$Q_w/\left[L\cdot(\min\cdot m)^{-1}\right]$	surrounding rock stability
1	28	26.0	0.32	0.30	18	IV
2	50	40.5	0.38	0.55	10.5	III
3	46	38	0.28	0.32	6	IV
4	65	54	0.23	0.52	13	IV
5	26	28	0.25	0.3	15	V
6	91	48	0.57	0.55	5	III
7	23	25	0.28	0.17	14	V
8	87	42	0.58	0.66	12	III
9	41.5	25.0	0.22	0.52	12	IV
10	93	156.5	0.78	0.82	3.2	I
11	52.0	25.0	0.22	0.52	12	III(IV)
12	24.2	12.5	0.13	0.18	125	V
13	76	63.9	0.65	0.62	10	II

Table III. Predicting sort of surrounding rock stability

number	$RQD(\%)$	$R_W(MPa)$	K_V	K_f	$Q_w/\left[L\cdot(\min\cdot m)^{-1}\right]$	surrounding rock stability
9	41.5	25.0	0.22	0.52	12	IV
10	93	156.5	0.78	0.82	3.2	I
11	52.0	25.0	0.22	0.52	12	III
12	24.2	12.5	0.13	0.18	125	V
13	76	63.9	0.65	0.62	10	II

CONCLUSIONS

The method of fuzzy clustering analysis may effectively classify surrounding rock stability, which overcomes ambiguous classification shortcomings of traditional and common gather, and realizes science and accuracy of clustering. According to known information clustering, it not only can economize classification time, but also check up the validity and correctness of classification. So the fuzzy clustering with known information in the real world has an important significance.

REFERENCES

[1] LIEW A W, YAN H, LAW N F. Image segmentation based on adaptive cluster prototype estimation [J]. IEEE Transactions on Fuzzy Systems, 2005, 13(4): 444−449.

[2] Zhu K J，Su S H，Li J L Optimal number of clusters and the best partition in fuzzy C-mean[J]. Systems Engineering—Theory＆Practice，2005，25(3)：52-61.
[3] LIU Y T. A genetic clustering algorithm for data with non-spherical shape cluster [J]. Pattern Recognition Letter, 2000, 33(1): 1251−1259.
[4] PHAM D L. Fuzzy clustering with spatial constraints [C]//Proceedings of the IEEE International Conference on Image Processing. New York: IEEE Computer Society, 2002, 2: 65−68.
[5] Zhang S H，Sun J X，Zhu K J. Sampling fuzzy C means clustering algorithm based on genetic ptimization[J]. Systems Engineering—Theory&Practice，2004，24(5)：121—125.
[6] CHEN Song-can, ZHANG Dao-qing. Robust image segmentation using FCM with spatial constraints based on new kernel-induced distance measure [J]. IEEE Transactions on Systems, Man, and Cybernetics, 2004, 34(4): 1907−1916.
[7] YU Jin-hua, WANG Yuan-yuan, SHI Xin-ling. Image segmentation with two-dimension fuzzy cluster method based on spatial information [J]. Opto-Electronic Engineering, 2007, 34(4): 114−119.(in Chinese)
[8] WANG X, WANG Y, WANG L. Improving fuzzy C-means clustering based on feature-weight learning [J]. Pattern Recognition Letter,2004, 25(10): 1123−1132.
[9] AHMED M N, YAMANY S M, MOHAMED N. A modified fuzzy C-means algorithm for bias field estimation and segmentation of MRI data [J]. IEEE Trans on Medical Imaging, 2002, 21(3):193−199.
[10] LI Xiao-he, ZHANG Tai-yi, QU Zhan. Image segmentation using fuzzy clustering with spatial constraints based on Markov random field via Bayesian theory [J]. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, 2008, 91(3):723−729.
[11] GAO Xin-bo. Fuzzy cluster analysis and its applications [M]. Xi'an: Xidian University Press, 2004: 50−54. (in Chinese)
[12] CLERC M, KENNEDY J. The particle swarm: Explosion, stability, and convergence in a multi-dimensional complex space [J]. IEEE Transaction on Evolutionary Computation, 2004, 6(11): 58−73.
[13] Cai Guangzhou.Research on BP neural network model of surrounding stability classification[D]. NanJing:HeHai University, 2001.
[14] Zhou Min,Lu zhiwu,Su Chao.[J]. Application of neural network of surrounding rock classification Underground Cavern. Guandong Water Resources and Hydropower , 2004, 1: 23~27.
[15] Jinzhu Hu,Chun Fang,Bin He,etc..A novel text clustering method based on dsom-fs-fcm[J], International Symposium on Distributed Computing and Applications for Business Engineering and Science,2008,pp354-359.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 349-352

Evaluation and development of software Engineering Supervision

Dong Feng[1], Zhang Qiu-xia[1],Li Hua[2]
[1]HuangHe Science and Technology College, ZhengZhou, China
Email:{ df ,zqx} @hhstu.edu.cn
[2] Sias International University, ZhengZhou, China
Email: qiuxiang6@sina.com.cn

***Abstract*—In this paper, we indicate the role of software engineering supervision plays in each phase of the life circle of software engineering, even we describe the evaluation model of the software engineering supervision. It analyzes and demonstrates the necessity of increasing supervision between the proprietor and the software engineering contractors. Though the effective evaluation of the software engineering, we can enhance the pirmorgraphics and normativizy of software engineering construction.**

***Index Terms*—software engineering supervision，life cycle, Evaluation Model**

I. INTRODUCTION

Software engineering supervision is refer to the engineering supervision organization established by law which process a appropriate qualification , authorized by the proprietor of organization, according to national law and regulations, technical standards and the contract of information systems engineering supervision[1]. To regular the implement of information systems project. China's software engineering supervision is the beginning stage. The relevant standards and specifications are still in short [2]. In the relationship of engineering supervision, the proprietor (construction organization) authorizes the supervisory side the management and controlling power of the project. Supervision organization, on behalf of the proprietor, handles the project management activities. Some Project Supervision Company based on relevant information about technical specifications and software engineering project contract, used his or her own experience, to supervise, the practice of most supervision company is scattered, making it very hard to form a supervision evaluation system with date to measure. This paper is based on the studying and analysis about the characteristics of software engineering, to demonstrate the necessity to increase the need for supervision between the proprietor and the software engineering contractors, making a preliminary discussion for the possible development of software engineering supervision in the future.

Ⅱ. THE FUNCTION OF SOFTWARE ENGINEERING SUPERVISION IN EACH PHASE OF THE LIFE CYCLE OF SOFTWARE ENGINEERING

A. Whole life cycle of the software Engineering

Whole life cycle of software engineering is in accordance with the definition of ISO. The full life cycle of software engineering phase can divide into the implementation phase, use phase and maintenance phase, of which is further subdivided into the implementation phase of preparation, design and construction. Together with the actual situation in our country, the software engineering life cycle has four phases. The first stage is "born" in phase, "the decision-making stage of system", once the system has made decision, the system enters the second stage, design stage, in which stage we set up system model. The third stage is "production" phase; the system was put into the development and construction. The fourth stage is the "operation and maintenance" phase, in which phase the system go into operation.

B.The function of software engineering supervision in the whole life cycle of software engineering

Software engineering supervision is a kind of society security structure in software engineering field. It is an independent third party organization to provide service on planning, organization, coordination and communication, controlling and management, monitoring and evaluation for engineering, the purpose is to support and ensure the success of software engineering. Software engineering supervision should monitor the every stage that after the design stage of software.

1) The supervision in the designing phase of software engineering.

Super vision engineer operate with owners closely, assessing, the qualification of the designing department and the design staff, checking the designing plan, reviewing the designing progress, evaluating the designing achievements and relate documents and follow checking the cooperation of internal designing.

2) The supervision at the stage of software engineering implementation

The supervision engineer should practice synchronization systematic tracking to supervise the construction stage and construction process. The problems of systematic tracking mainly are: inspecting and approving the plan of constructing and construction process which putted by the contractor unit, examining the situation of the implement of software engineering, collecting and collation the documents on project technology, writing the conclusion of the object, to make

a good foundation for the construct department's running, management and asserting of the new system.

3)The supervision at the stage of running and asserting of the software engineering.

At the running and asserting stage, supervision engineer should inspect whether the configuration of the operation system, application system and other software conform to the designing project, examine the similarity between the system's function and the contract, check the situation that the plan of training staff, help the user make the rules about running and management of the system.

C. Evaluation model of software engineering supervision.

The purpose of information engineering monitor is getting profits [3-4]. Business owner and the project constructor become dual relationship after reaching a contract during the process of software engineering supervision and evaluation. Whatever, the business owner or the developer is hard to solve these two problems by himself; however, the effect of supervision is able to reduce the pressure of the dissymmetry between the two parties. So, it's necessary to the appear of the third party's supervision. The supervision, business owner and the developer become a triplet organized relationship.

Although software engineering supervision adds the cost of the project, it also increases the success rate of the project and reduces the lost of the net social capital. The canonical managing way, means and process of supervision not ensure the project's quality, but decrease the assert costs of the project, thus they reduce the business costs. Keep inspecting by the mean of analyzing the data, quoting the final result as the same time as software engineering supervision. The purpose of evaluating the software engineering supervision is to achieve the software project monitoring point and reach the quality requirement in certain time and costs. Its model is shown as "Fig 1".

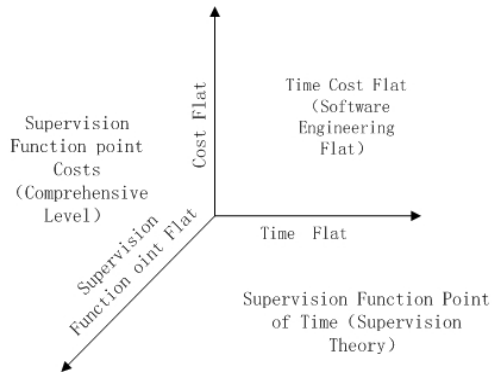


Figure 1. Software Project Supervision Assessment dimensional model.

Software engineering supervision is outsourced project, the statistic meaningful formula of it's assessment model is as in (1):

$$SSE = [(\frac{C_1}{CD_1} * \frac{SFP_1}{SFPD_1}) + (\frac{C_2}{CD_2} * \frac{SFP_2}{SFPD_2}) + ... + (\frac{C_i}{CD_i} * \frac{SFP_i}{SFPD_i})] / T_i \quad (1)$$

Where by SSE represents software supervision quality, C1 stands for how much it cost at the first milestone, CD1 means the cost value of the beget at the first milestone; CO2 stands for how much it cost at the second milestone, CO2 means the cost value of the beget at the second milestone, According to the condition at the milestone, T1 represents the milestone which is calculating currently.

These three purpose come across all the processes of software engineering supervision, there is no weight in the first formula that we provide, the tendency in the software engineering project is more and more projects need the weight factors, therefore, the important level of the targets that they care about is different milestones, the effects of software engineering project's weight shown as "Fig.2".

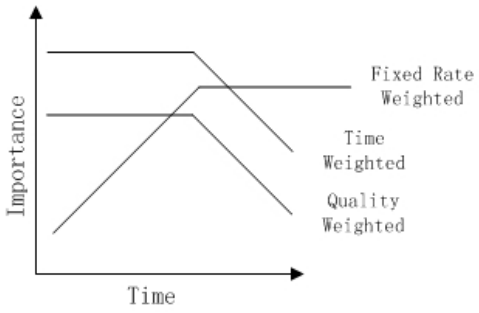


Figure 2. Note how the caption is centered in the column.

Following, we give out quantification formula of the assessing model of software engineering supervision as in (2)

$$SSE = [(\frac{C_1}{CD_1} * \frac{SFP_1}{SFPD_1}) * \lambda_1 * \mu_1 * \beta_1 + (\frac{C_2}{CD_2} * \frac{SFP_2}{SFPD_2}) * \lambda_2 * \mu_2 * \beta_2 + ... + (\frac{C_i}{CD_i} * \frac{SFP_i}{SFPD_i}) * \lambda_i * \mu_i * \beta_i] / T_i \quad (2)$$

Represent cost weight, time weight, quality weight respectively, the numeric area of the subscript is from 1 to i.

The whole model of software engineering supervision assessment contains five factors: supervising object, supervising purpose supervising content, security safeguard, supervising implementing [5-7]. We can get measurable calculating results by comprehensive calculation and analysis, so it can let software engineering supervision have a fundamental basis. At present, the risk of constructing software engineering project is larger and the constructing market needs further regulate. In order to reduce the risk of the software engineering construction and regulate the project construction market, protect the benefits of the business owners and the contractor units, it's important and urgent

to conduct organized and normalized supervising assessment for the software engineering construction.

III. THE SOFTWARE ENGINEERING SUPERVISION NECESSITY AND DEVELOPMENT TREND

Software engineering technology achieved a fast development in recent decade, every all of life put large amount funds into the information construction nowadays, the control system of industry's automation, the ERP system in enterprises, the digitizing school yard system and the software like these engineering projects spring up and have leap-forward development. However, because the origination, development not complete, the supervision software engineering can't hold a candle to the mature supervision to the constructional works. The mature level of project supervision is worse, so the development of software engineering supervision still has a much large space in the future, and the software engineering supervision will have more and more important effects on software engineering construction supervision, there are few problems shown in the picture to software engineering supervision waiting solving, and the development of software engineering supervision will also go with the solution of these problems, a few issues to be addressed as "Fig.3."

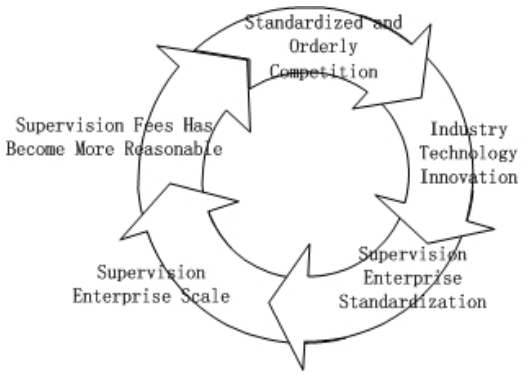


Figure 3. Classification of Software Project Supervision

A. An orderly competition mechanism.

The supervision market of information project hasn't formed an orderly competition mechanism. At present, the domestic supervision market of information project is not well regulated. The governmental supervision strength upon the supervision market is not enough, which result in the chaos in the supervision market. Competition among enterprises relies on the competition of relationship and price to a large extent. A great number of supervision units lower their prices to obtain jobs and even take illegal measures, such as reciprocations, commissions to attract customers. There should be a good supervision method for related departments to settle these problems, or this situation will be very hard to be improved.

B. Poor technological innovative capacity.

After more than a decade's development, the supervisions of information project still hasn't achieved innovative breakthrough both theoretically and practically. Alt hasn't been combined with the advanced experience home and abroad. Comparing with the development areas, there is a gap in such aspects as thinking, methodologies, business scope and operating mechanism, etc. The supervision industry of information project should strengthen the development and promotion of the advanced applied technology; make compulsory promoting measures to make the industry a substantial improvement in technological innovative capacity.

C. The standardization of information supervision enterprise.

Some information project supervision enterprises takes on inner management confusion, only cares about money while doing less work with a bad sense of service. On a whole, there is much difference among the managing levels of the supervision enterprise standardization. Generally, supervision enterprises are relatively small in scale, weak in financial and technological strength, which makes it hard to complete with the big companies from home and abroad. Also, some supervisors with a loose discipline who take for their own interests under the name of work have made great damage to the reputation of this industry. Some nominal supervision units having qualification certificate and business license, but without regular employees, don't have qualifications in supervision. To gain the supervision job, several people are gathered temporarily. Once the mission is over, the staff is on the dissolution, let alone to carry out standardized work. So, the supervisors don't have high qualities. It's necessary to improve the present situation that the tantalization of supervision enterprises is not high.

D. Enlargement in scale of information project supervision enterprises

At present, the scale of supervision enterprises of information project is generally small and the technological level is not high. The knowledge they possess is relatively old. There are also enterprises that employ supervisors after they accepted the job to save their expenses, thus make the unbalanced qualities among supervisors. So, the scale of information project supervision enterprises needs to be enlarged.

E. Relatively how cost of information project supervision

Nowadays, the cost of information project remains at a low level, which makes it hard for super visitation enterprises to develop. It's harder to attract higher-level people into the supervision job. The supervision work maintained at a low level circulation. The majority of supervision enterprises just make ends meet thus makes eating the most important problem. While the main energy having been put on to survival, it is hard to develop. There are the facts we have to confront with.

To sum up, with the information, construction on its full swing, software engineering supervision is gradually accepted and applied in our country and will play a more

and more important role in the information construction. The supervision enterprises should pay attention to their own qualities and management and require themselves according to specifications. The software engineering supervision will pave its way to standardized assessment, institutionalization. Specialization and socialization step by step and gradually evolve into a mature and perfect market to play a greater role in the tide of information construction.

REFERENCES

[1] Wei-hong Li, Modern computer, "comparison between Information System Engineering supervision and Project management", March 2007, pp.62-63.

[2] Xiang Yong,, Study paper in Chong qing Architecture University, "Project supervision Management Performance Theory under Asymmetric Information", February 2006, pp.115-118.

[3] Hong- chang An. Beijing University of Technology. Benefit achievement and risk-driven Information Engineering Management, February, 2006, pp.71-72.

[4] Ben-pang Ren, Shandong University, "analysis and research of software engineering supervision model with the center of organization and coordination" February 2006, pp.61-63.

[5] He Feng, Science and Technology consultation, "Discussion the perfection of Software Engineering supervision", March 2007, pp.74-76.

[6] Liu Yang, Tian jin Management University, "Software Project Supervision objective control and effectiveness Evaluation and study" February, 2004, pp.40-41.

[7] Li Dong, Computer World Network of China, "IT project supervision comparison of three models", March 2004, pp.43-45.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 353-356

Research on Hardware I/O Passthrough in Computer Virtualization

Bencang Liu[1], Lishen Yang[1], Xiaoming Qin[2]
[1] College of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, China
Email: {bcliu, yangls}@hpu.edu.cn
[2] Dept. of Computer Science and Technology, Jiaozuo Teachers College, Jiaozuo, China
Email: xmjzhn@jzsz.cn

***Abstract*—Processors have evolved to improve performance for virtualized environments. But A commodity I/O device has no support for virtualization. A VMM can assign such a device to a single guest with direct, fast, but insecure access by the guest's native device driver. Discover one such I/O performance enhancement called device passthrough. With the technology, the guest operating system can use the hardware device as if it were a non-virtualized system. In this paper, we describe the innovation and present preliminary results how to improve performance of PCI devices using hardware support from Intel (VT-d) or AMD (IOMMU).**

Index Terms—**passthrough, emulation, hypervisor, virtualization, linux**

I. INTRODUCTION

Platform virtualization is about sharing a platform among two or more operating systems for more efficient use of resources. But platform implies more than just a processor: it also includes the other important elements that make up a platform, including storage, networking, and other hardware resources. Some hardware resources can easily be virtualized, such as the processor or storage, but other hardware resources cannot, such as a video adapter or a serial port. Peripheral Component Interconnect (PCI) passthrough provides the means to use those resources efficiently, when sharing is not possible or useful. This article explores the concept of passthrough, discusses its implementation in hypervisors, and details the hypervisors that support this recent innovation.

Passthrough I/O [1, 2, 3], let device emulation (virtual devices) [4] works in two hypervisor architectures. The first architecture incorporates device emulation within the hypervisor, while the second pushes device emulation to a hypervisor-external application.

Device emulation within the hypervisor is a common method implemented within the VMware workstation product. In this model, the hypervisor includes emulations of common devices that the various guest operating systems can share, including virtual disks, virtual network adapters, and other necessary platform elements.

The second architecture is called user space device emulation. As the name implies, rather than the device emulation being embedded within the hypervisor, it is instead implemented in user space. QEMU [8] provides for device emulation and is used by a large number of independent hypervisors. This model is advantageous, because the device emulation is independent of the hypervisor and can therefore be shared between hypervisors. It also permits arbitrary device emulation without having to burden the hypervisor with this functionality. The same idea exists with the hypervisor. The security of the hypervisor is crucial, as it isolates multiple independent guest operating systems. With less code in the hypervisor (pushing the device emulation into the less privileged user space), the less chance of leaking privileges to untrusted users.

Another variation on hypervisor-based device emulation is paravirtualized drivers. In this model, the hypervisor includes the physical drivers, and each guest operating system includes a hypervisor-aware driver that works in concert with the hypervisor drivers.

Regardless of whether the device emulation occurs in the hypervisor or on top in a guest virtual machine (VM), the emulation methods are similar. Device emulation can mimic a specific device or a specific type of disk. The physical hardware can differ greatly—for example, while an IDE drive is emulated to the guest operating systems, the physical hardware platform can use a serial ATA (SATA) drive. This is useful, because some type support is common among many operating systems and can be used as a common denominator instead of all guest operating systems supporting more advanced drive types.

II. DEVICE PASSTHROUGH

In the two device emulation models discussed above, there's a price to pay for sharing devices. Whether device emulation is performed in the hypervisor or in user space within an independent VM, overhead exists. This overhead is worthwhile as long as the devices need to be shared by multiple guest operating systems. If sharing is not necessary, then there are more efficient methods for sharing devices.

So, at the highest level, device passthrough is about providing an isolation of devices to a given guest operating system so that the device can be used exclusively by that guest. Two of the most important reasons are performance and providing exclusive use of a device that is not inherently shareable.

A. Architecture of I/O Passthrough

We propose a novel I/O virtualization technique, virtual passthrough I/O (VPI/O). VPI/O allows the guest's native driver to have direct access to a commodity device (one that does not have self-virtualization support) most of the time. The VMM can assure, however, that the guest does not maliciously or inadvertently program the device to affect the VMM (hypervisor) or the other guests. Furthermore, the VMM can hand-off the physical device from one guest to another. Furthermore, the VMM can hand-off the physical device from one guest to another. The architecture of the virtualization technique was shown in Figure 1.

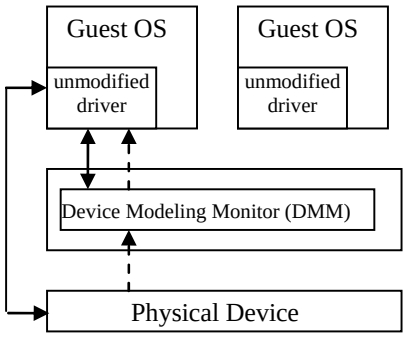


Figure 1. Passthrough architecture

The essential idea is that the VMM maintains a formal model of the I/O device that is driven by guest/device interactions. The model can be far simpler than a driver or virtual device implementation, and must only be sufficiently detailed so that when faced with an interaction, the model can determine:

• Whether the device is serially reusable after the interaction.

• Whether a DMA is about to start, and which host physical addresses will be involved.

With such a model, the VMM is able to determine whether a device interaction should be allowed to continue down to the physical device, and at what points a device can be context-switched to a different guest. Thus the VMM can multiplex a single commodity physical device across multiple guests, each of which uses a native driver. If every guest/device interaction involves an exit into the VMM, the performance will be terrible. The practicality of virtual passthrough I/O hinges on the extent to which exits can be avoided through modeling and systems techniques, and/or the extent to which the overhead of an exit can be reduced.

The device model is conceptually a state machine with additional scratchpad information (e.g., DMA addresses). The edges are annotated with the device requests (e.g., I/O port reads/writes, interrupts) that trigger them, as well as with checking functions. A checking function is called before a state transition occurs, and must approve the state transition. If state transition is denied, the device request fails, and no state transition occurs. Optionally, a notification of failure can be delivered to the guest. The checking functions reflect VMM policy. As side effects, they also can change the hooked I/O list.

B. Performance Concerns

For performance, the Virtual Functions (VF) device has direct access to its own registers and I/O Memory Management Unit (IOMMU) technology allows translation of guest physical addresses (GPA) into host physical addresses for direct I/O the VF will achieve near native (or bare metal) performance running in a guest OS. Each VM using a VF device will get the benefits of higher throughput with lower CPU utilization compared to the standard software emulated NIC. Another significant benefit of a Virtual Function device using Direct I/O is that register reads and writes do not have to be trapped and emulated. The CPU paging features are used to directly map the VF device MMIO space into the guest. Trapping and emulating register reads and writes are very expensive in terms of CPU utilization and extra task switches.

Near-native performance can be achieved using device passthrough [7]. This is perfect for networking applications (or those that have high disk I/O) that have not adopted virtualization because of contention and performance degradation through the hypervisor (to a driver in the hypervisor or through the hypervisor to a user space emulation). But assigning devices to specific guests is also useful when those devices cannot be shared. For example, multiple video adapters in a system could be passed through to unique guest domains.

Finally, there may be specialized PCI devices that only one guest domain uses or devices that the hypervisor does not support and therefore should be passed through to the guest. Individual USB ports could be isolated to a given domain, or a serial port (which is itself not shareable) could be isolated to a particular guest.

III. COMPARISON TO DEVICE EMULATION

Early forms of device emulation implemented shadow forms of device interfaces in the hypervisor to provide the guest operating system with a virtual interface to the hardware. This virtual interface would consist of the expected interface, including a virtual address space representing the device (such as shadow PCI) and virtual interrupt. But with a device driver talking to a virtual interface and a hypervisor translating this communication to actual hardware, there's a considerable amount of overhead—particularly in high-bandwidth devices like network adapters.

Xen popularized the PV approach [5], which reduced the degradation of performance by making the guest operating system driver aware that it was being virtualized. In this case, the guest operating system would not see a PCI space for a device (such as a network adapter) but instead a network adapter application programming interface (API) that provided a higher-level abstraction. The downside to this approach was that the guest operating system had to be modified for PV. The upside was that you can achieve near-native performance in some cases.

Early attempts at device passthrough used a thin emulation model, in which the hypervisor provided software-based memory management by translating guest operating system address space to trusted host address space. And while early attempts provided the means to isolate a device to a particular guest operating system, the approach lacked the performance and scalability required for large virtualization environments. Luckily, processor vendors have equipped next-generation processors with instructions to support hypervisors as well as logic for device passthrough, including interrupt virtualization and direct memory access (DMA) support. So, instead of catching and emulating access to physical devices below the hypervisor, new processors provide DMA address translation and permissions checking for efficient device passthrough.

IV. SUPPORT FOR DEVICE PASSTHROUGH

Both Intel and AMD provide support for device passthrough in their newer processor architectures [6]. Intel calls its option Virtualization Technology for Directed I/O (VT-d), while AMD refers to IOMMU. In each case, the new CPUs provide the means to map PCI physical addresses to guest virtual addresses. When this mapping occurs, the hardware takes care of access and protection, and the guest operating system can use the device as if it were a non-virtualized system. In addition to mapping guest to physical memory, isolation is provided such that other guests are precluded from accessing it. The Intel and AMD CPUs provide much more virtualization functionality.

Another innovation that helps interrupts scale to large numbers of VMs is called Message Signaled Interrupts (MSI). Rather than relying on physical interrupt pins to be associated with a guest, MSI transforms interrupts into messages that are more easily virtualized. MSI is ideal for I/O virtualization, as it allows isolation of interrupt sources.

Using the latest virtualization-enhanced processor architectures, a number of hypervisors and virtualization solutions support device passthrough. You'll find support for device passthrough (using VT-d or IOMMU) in Xen and KVM as well as other hypervisors. In most cases, the guest operating system (domain 0) must be compiled to support passthrough, which is available as a kernel build-time option. Hiding the devices from the host VM may also be required (as is done with Xen using pciback). Some restrictions apply in PCI (for example, PCI devices behind a PCIe-to-PCI bridge must be assigned to the same domain), but PCIe does not have this restriction.

Additionally, you'll find configuration support for device passthrough in libvirt (along with virsh), which provides an abstraction to the configuration schemes used by the underlying hypervisors.

V. PROBLEMS WITH DEVICE PASSTHROUGH

One of the problems introduced with device passthrough is when live migration is required. Live migration is the suspension and subsequent migration of a VM to a new physical host, at which point the VM is restarted. This is a great feature to support load balancing of VMs over a network of physical hosts, but it presents a problem when passthrough devices are used. PCI hotplug (of which there are several specifications) is one aspect that needs to be addressed. PCI hotplug permits PCI devices to come and go from a given kernel, which is ideal—particularly when considering migration of a VM to a hypervisor on a new host machine. When devices are emulated, such as virtual network adapters, the emulation provides a layer to abstract away the physical hardware. In this way, a virtual network adapter migrates easily within the VM.

VI. CONCLUSIONS

We have proposed a new technique for I/O virtualization of commodity I/O devices, virtual passthrough I/O (VPIO). VPIO is an intermediate option between passthrough I/O and traditional fully emulated virtual devices. It provides some of the performance of the former, while maintaining the security/protection of the latter. VPIO implements a device state model in the VMM that vets guest access to the physical network card. The overhead of running the model is much less than the overhead of fully virtualizing the device.

The next steps in I/O virtualization are actually happening today. For example, PCIe includes support for virtualization. One virtualization concept that's ideal for server virtualization is called Single-Root I/O Virtualization (SR-IOV). This virtualization technology (created through the PCI-Special Interest Group, or PCI-SIG) provides device virtualization in single-root complex instances (in this case, a single server with multiple VMs sharing a device). Another variation, called Multi-Root IOV, supports larger topologies. In a sense, this permits arbitrarily large networks of devices, including servers, end devices, and switches (complete with device discovery and packet routing).

With SR-IOV, a PCIe device can export not just a number of PCI physical functions but also a set of virtual functions that share resources on the I/O device. In this model, no passthrough is necessary, because virtualization occurs at the end device, allowing the hypervisor to simply map virtual functions to VMs to achieve native device performance with the security of isolation.

The key challenge in further improving the performance of VPI/O is to decrease the number of exits and their costs even more. It is clear that while we can reduce the number of device requests and events that we need to intercept through careful device modeling, the high cost of interceptions and VM exit/entry in the VMM is the most problematic issue with the VPIO model [9].

Virtualization has been under development for about 50 years, but only now is there widespread attention on I/O virtualization. Commercial processor support for virtualization has been around for only five years. So, in essence, we're on the cusp of what's to come for platform and I/O virtualization. And as a key element of future architectures like cloud computing, virtualization will

certainly be an interesting technology to watch as it evolves. As usual, Linux is on the forefront for support of these new architectures, and recent kernels are beginning to include support for these new virtualization technologies.

References

[1] LIU, J., HUANG, W., ABALI, B., AND PANDA, D. High performance vmm-bypass i/o in virtual machines. In Proceedings of the USENIX Annual Technical Conference (May 2006).

[2] RAJ, H., AND SCHWAN, K. High performance and scalable i/o virtualization via self-virtualized devices. In Proceedings of the 16th IEEE International Symposium on High Performance Distributed Computing (HPDC) (July 2007).

[3] SHAFER, J., CARR, D., MENON, A., RIXNER, S., COX, A. L., ZWAENEPOEL,W., AND WILLMANN, P. Concurrent direct network access for virtual machine monitors. In HPCA '07: Proceedings of the 2007 IEEE 13th International Symposium on High Performance Computer Architecture (Washington, DC, USA, 2007), IEEE Computer Society, pp. 306–317.

[4] SUGERMAN, J., VENKITACHALAN, G., AND LIM, B.-H. Virtualizing I/O devices on VMware workstation's hosted virtual machine monitor. In Proceedings of the USENIX Annual Technical Conference (June 2001).

[5] BARHAM, P., DRAGOVIC, B., FRASER, K., HAND, S., HARRIS, T., HO, A., NEUGEBAUER, R., PRATT, I., AND WARFIELD, A. Xen and the art of virtualization. In Proceedings of the 19th ACM Symposium on Operating Systems Principles (SOSP) (October 2003).

[6] AMD CORPORATION. AMD64 virtualization codenamed "pacific" technology: Secure virtual machine architecture reference manual, May 2005.

[7] ADAMS, K., AND AGESEN, O. A comparison of software and hardware techniques for x86 virtualization. In ASPLOS-XII: Proceedings of the 12th international conference on Architectural support for programming languages and operating systems (New York, NY, USA, 2006), ACM, pp. 2–13.

[8] BELLARD, F. Qemu: A fast and portable dynamic translator. In Proceedings of the USENIX Annual Technical Conference, Freenix Track (April 2005).

[9] XIA, L., LANGE, J., DINDA, P., AND BAE, C. Investigating virtual passthrough I/O on commodity devices. Operating Systems Review 43, 3 (July 2009).

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 357-360

Performance Research of Modulation for Optical Wireless Communication

Gao Yan[1,] Wu Min[2]
[1]College of Computer Science &Technology Henan Polytechnic University Jiaozuo, Henan
[1]School of Information Science & Technology Southwest Jiaotong University Chengdu, Sichuan
Email: gaoyan@hpu.edu.cn
[2]College of Computer Science &Technology Henan Polytechnic University Jiaozuo, Henan
Email: cathy-21@126.com

***Abstract*—There are many modulation methods suitable for optical wireless communication. OOK, PPM are adopted widely in optical wireless communication for its high average-power-efficiency. DPPM, DPIM and DH-PIM are three new modulation methods for optical wireless communication, which may be the substitutes of PPM because of their better performance in power efficiency and bandwidth efficiency. In this paper, in combination of the characteristic of the atmospheric optical wireless channel, the bandwidth efficiency, transmission capacity, power efficiency and slot error rate of the typical modulation schemes as OOK, PPM, DPPM, DPIM and DH-PIM for atmospheric optical wireless communications are analyzed. Theoretical analysis and simulation results by matlab show that DPPM, DPIM and DH-PIM are more applicable for the future optical wireless communication.**

***Index Terms*—optical wireless communication; modulation; slot error rate**

I. INTRODUCTION

With the increasing of information communication, expansion of network bandwidth resources and improvement of communication flows have become important issue. Currently the main means of communication transmission contains microwave, fiber and so on. Compared with wire communication, a lot of non-ferrous metals can be saved and complex terrains can be crossed by microwave communication. As a new communication technology, optical wireless communications have the advantages of optical fiber communication and mobile communications, with wide bandwidth and without the need for application of frequency. Therefore, in recent years, the research on wireless optical communication was gotten more attention. But the wireless optical communication in the atmosphere were influenced by atmospheric absorption, scattering and turbulence lead to signal attenuation, while the average transmission power is limited owing to the requirements for safety of human eye [2]. Thus higher requirements of modulation are proposed.

This work is supported by the Natural Science Foundation of Henan Province Department of Education (the serial number is 2008B470002) and Henan key project (the serial number is 082102210079).

AP-PROC-CS-10CN007

II. THE CLASSIFICATION OF WIRELESS OPTICAL MODULATION

A variety of wireless optical communication modulations are proposed in the present study. The ways involved in this article are the focus of attention in this field, including the on-off keying modulation, pulse position modulation, differential pulse position modulation, digital pulse interval modulation, double-pulse interval modulation and improved differential pulse position modulation.

A. On-Off keying Modulation

In the digital wireless optical systems, as the simplest way, the on-off keying modulation is based on intensity modulation with direct detection [3]. The generation of optical pulse is achieved by opening and breaking of lasers, when sending information "1", the light pulses are sent; when sending information "0", the laser is shut down completely.

B. Single-pulse position modulation (L-PPM)

Single-pulse position modulation converts a binary M-bit data group to a single pulse signal at a particular time slot in time segment which are composed by $L=2^M$ time slots, each time slot is called chip.

C. Differential Pulse Position Modulation

Differential pulse position modulation is one of the methods of modulation which is improved on the basic of PPM. For one sign of L-PPM, its time slot is aptotic L bits, one of them is 1, and the other are 0. Then the code number of L-DPPM is indefinite, it is composed by s string of low level and single high level followed [4]. The signal after high level in a code block of the PPM modulation signal is removed by the DPPM modulation signal.

D. Digital Pulse Interval Modulation(DPIM)

As similar to the DPPM, the symbol length of the DPIM is unfixed and can be divided into unprotected slots and protected slots, one protected slot is mostly adopted by protected DPIM modulation to reduce the impact of intersymbol interference effectively. The modulation symbols S_k (k is the decimal number expressed by the symbol) contain k+2 time slots, after

each starting time slot L, the pulse adds a protected empty slot and adds k empty slots for expressing information. When demodulation in the receiver, after determining the pulse time slot received, it only needs to count the empty time slot and subtract one for them. In the receiver, therefore DPIM only need clock synchronization without symbol synchronization, this greatly simplifies implementation of the system.

E. Double-Pulse Interval Modulation (DH-PIM)

DH-PIM modulation is more complex, the time slot included by each symbol is also mutative, but the symbol adopts two kinds of starting pulse. The symbol S_k is formed by a head slot and m empty time slots followed .The head time slot is included by α+1 time slots (α is integer). Considering two forms of head H1 and H2, theH1 initial pulse width is α/2 time slot, followed by (α / 2) +1 protected time slots; H2 pulse width is a time slots, followed by one time slot. When $k<2^M-1$, the head time slot of symbol S_k is H1, otherwise it is H2.

III.Performance Analysis and Comparison of Bandwidth Demand

Because the bandwidth of the receiver is limited by the large volume capacitance of light receptors, bandwidth of the wireless optical communication system is much smaller and much better. Under the condition of same bit rate, bandwidth requirements of the five kinds of modulation are analyzed and compared, assuming that the modulation order are M. For the OOK system, the bit rate of information are R_b, then the share of bandwidth are $B_{ook}=R_b$; in the PPM system, for the same bit rate R_b, the bandwidth of PPM are $B_{PPM}=2^MR_b/M$; in the DPIM system, for the same bit rate R_b, the bandwidth of DPIM are: $B_{DPIM}= (2^M+3) R_b/2M$. the bandwidth of DPPM as follows: $B_{DPPM}=(2^M+1)R_b/2M$; the average time slot length of DH-PIM modulation are $L_\alpha=\alpha+2^{M-2}+1/2$; the cycle of time slot are: $T_\alpha=M/(L_\alpha R_b)=2M/[(2\alpha+2^{M-1}+1)R_b]$. then the bandwidth as follows: $B_{DH\text{-}PIM}=(2^{M-1}+2\alpha+1)R_b/2M$. The results of normalized band width according to OOK by matlab simulation [1] are showed in the figure 1.

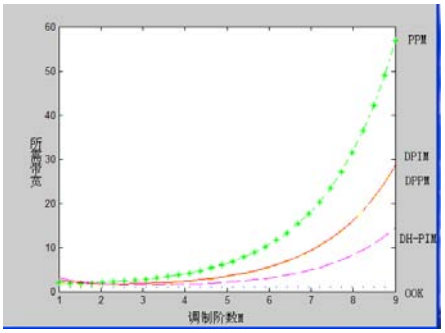


Figure 1. The comparison of bandwidth on OOK PPM DPPM DPIM and DH-PIM

It can be seen that the bandwidth demand of PPM is the highest and the bandwidth demand of DPPM is the second highest, the bandwidth demand of OOK is the lowest, along with the increase in the order of modulation M, the bandwidth demand of PPM, DPPM, DPIM and DH-PIM modulation is higher and higher.

IV.Performance Comparison of Average Transmitted Power

Because of the eye-safe and portable requirements of mobile communication device, the transmitted power of atmospheric wireless optical communication is greatly restricted. as far as possible to improve utilization rate of power is demanded [7]. For the OOK (NRZ), assuming that P_1 is the power of launching optical pulse "1", in the event that the probability of appearing "0" and "1" is same, its average transmitted power is: $P_{OOK}=P_1/2$. Because 2^M time slots are contained by one PPM symbol, only one time slot send optical pulse. So the average transmitted power of L-PPM is: $P_{PPM}=P_1/2^M=P_{OOK}/2^{M-1}$, $(2^M+1)/2$ time slots are contained by one DPPM symbol, so $P_{DPPM}=2P_1/(2^M+1)=4P_{OOK}/(2^M+1)$. By the same token, $P_{DPPM}=2P_1/(2^M+3)=4P_{OOK}/(2^M+3)$ i*n* the way of DH-PIM, the average width of head pulse is 1.5 pulses (supposing that the two head pulses is equal probability). $\alpha/2$ slots are employed by each pulse, so the average time slot in length of head pulse is $L_\alpha=3\alpha/4$ and the average transmitted power is $P_{DH-PIM}=\frac{3\alpha P_1}{4L_\alpha}=\frac{3\alpha P_1}{2(2^{M-1}+2\alpha+1)}$

The results of normalized average power comparison for OOK by matlab simulation are showed in figure 2 under the same peak power [8].

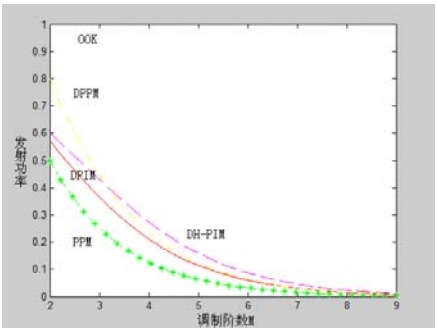


Figure 2. The comparison of average transmitted power on OOK PPM DPPM DPIM and DH-PIM

It can be seen that the power utilization efficiency of PPM, DPIM, DPPM and DH-PIM is higher than OOK; with M increases, the trend is more obvious; power utilization efficiency of PPM is the highest and the DPIM is the second highest. When M is less than 4, the power utilization efficiency of DH-PIM (α=1) is superior to DPPM; When M is more than 4, the power utilization efficiency of DPPM is superior to DH-PIM (α=1), As M increases, the average transmit power of DPPM and DPIM tend to equal, DH-PIM power utilization ratio is only better than OOK, and its power utilization ratio decreases as α increases. The average transmission power gradually decline as M increases except OOK.

V. Performance Comparison of Transmission Capacity

The size of the transmission capacity represents the ability to transfer information per unit time, and it also was important performance index of atmospheric wireless optical communication. Under the condition of same time slot width, the transmission capacity of OOK, PPM, DPPM, DPIM, DH-PIM modulation were analyzed and compared, their order of modulation were M, the size of bit rate were used to measure size of transmission capacity [5]. Supposing the slot width are τ, then transmission capacity of OOK are1/τ, the average symbol length of PPM are $2^M\tau$, the average symbol length of DPPM are $(2^M+1)/2\tau$, the average symbol length of DPIM are $(2^M-3)/2\tau$, the average symbol length of DH-PIM are($2^M+2\alpha+1)/2\tau$, Each symbol corresponds to M bits of binary information, so the transmission capacity of PPM, DPPM, DPIM, DH-PIM are respectively as follows:

$$C_{PPM}=\frac{M}{2^M\tau}=\frac{M}{2^M}C_{OOK}$$

$$C_{DPIM}=\frac{2M}{(2^M-3)\tau}=\frac{2M}{2^M-3}C_{OOK}$$

$$C_{DPPM}=\frac{2M}{(2^M+1)\tau}=\frac{2M}{2^M+1}C_{OOK}$$

$$C_{DH-PIM}=\frac{2M}{(2^M+2\alpha+1)\tau}=\frac{2M}{2^M+2\alpha+1}C_{OOK}$$

The results of normalized transmission capacity for M by matlab simulation are shown in Figure 3.

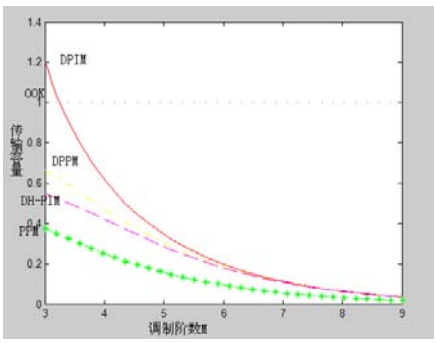


Figure 3. The comparison of transmission capacity on OOK PPM DPPM DPIM and DH-PIM

From figure 3, when M=2, transmission capacity of OOK are the highest, transmission capacity of DH-PIM (a=1) are the same with DPPM, transmission capacity of DH-PIM (a=2) are the same with DPIM, and the transmission capacity of DH-PIM (a=1) is higher than DH-PIM (a=2) and transmission capacity of PPM is the lowest [4]. When M is greater than 2, OOK transmission capacity is the highest, DH-PIM is superior to DPPM, and DH-PIM transmission capacity increases as a decrease, DPPM is superior to DPIM, transmission capacity of the OOK is the minimum. As M increases (except for OOK), other modulation transmission capacity of other modulation is lower and lower and tends to equal.

VI. Performance Analysis and Comparison of Error Rate

The intensity modulation / direct detection (IM / DD) were adopted in the system of optical wireless communication. It was assumed that the additive white Gaussian noise (AGWN) only exist to discuss conveniently, the mean value of noise n(t) is 0 and variance is σ_n^2 [6]. At the same time the bandwidth of receiver is very wide. Then the x (t) which is gotten in the input of sample decision device are $\sqrt{S_t}+n(t)$ when pulse "1" is sent, or x (t) is n (t) without pulse. The signal peak power in the input of decision device are S_t. Decision threshold is supposed b, $P_{1/0}$ is the probability that "1" are misjudged "0"and $P_{0/1}$ is the probability that "0"are misjudged "1", they respectively are as follows:

$$P_{01}=(1/2)\{1+erf[(b-\sqrt{S_t})/\sqrt{2\sigma_n^2}]\}$$

$$P_{10}=(1/2)\{1-erf[b/\sqrt{2\sigma_n^2}]\}$$

$$erf(x)=\frac{2}{\sqrt{\pi}}\int_0^{\pi}\exp(-u^2)du=1-erfc(x)$$

Then the error rate are $P_{se}=P_1P_{0/1}+P_0P_{1/0}$. P_1 and P_0 respectively are the probability of "1"transmitted and "0" transmitted. $P_0+P_1=1$. Probability of occurrence for information "1"and "0" are supposed to equal. For OOK, $P_1=P_0=1/2$. Obviously the best decision thresholds are:

$$P_{se,OOK}=\frac{1}{2}P_{0/1}=\frac{1}{2}P_{1/0}=\frac{1}{2}erfc(\sqrt{S_t/2\sigma_n^2}/2)$$

By the same token available, the error rate of PPM, DPPM, DPIM and DH-PIM are respectively:

$$P_{se,PPM}=\frac{1+erf[(b-\sqrt{S_t})/\sqrt{2\sigma_n^2}]+(2^M-1)[1-erf(b/\sqrt{\sigma_n^2})]}{2^{M+1}}$$

$$P_{se,DPPM}=\frac{1+erf[(b-\sqrt{S_t})/\sqrt{2\sigma_n^2}]+[(2^M-1)/2][1-erf(b/\sqrt{\sigma_n^2})]}{2^M+1}$$

$$P_{se,DPIM}=\frac{1+erf[(b-\sqrt{S_t})/\sqrt{2\sigma_n^2}]+[(2^M-1)/2][1-erf(b/\sqrt{\sigma_n^2})]}{2^M+3}$$

$$P_{se,DH-PIM}=\frac{(3\alpha/2)(1+erf[(b-\sqrt{S_t})/\sqrt{2\sigma_n^2}]+[(4L_m-3\alpha)/2][1-erf(b/\sqrt{\sigma_n^2})]}{4L_m}$$

The derived function which is gotten by finding derivative to b on both sides of the above four formulas is supposed to be "0". Then the optimum threshold b can be obtained. For the PPM: $l=2^M-1$; for the DPPM: $l=(2^M-1)/2$; for the DPIM: $l=(2^M+1)/2$; for the DH-PIM: $l=(2^M+\alpha+2)/3\alpha$.

Under the optimum threshold, the curve of error rate relative to SNR (Signal to Noise Ratio) is shown in figure 4, and the SNR is defined as $S_t/2\sigma_n^2$. From that the error rate decreases as SNR increase for one modulation. When the SNR is certain, the error rate decrease as M increase(except OOK); When the M is certain, the PPM error rate is the minimum and the OOK error rate is the maximum, the DPIM error rate is inferior to DH-PIM, the DH-PIM error rate increase as α increase. The DPPM error rate is briefly inferior to DH-PIM (α=1) when M is equal to 3; and the DPPM error rate is between DH-PIM (α=1) and DPIM when M is equal to 4; the DPPM error

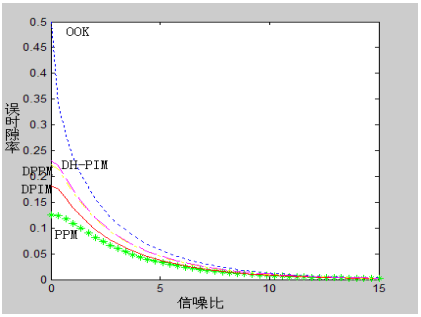


Figure 4. the comparison of error rate on OOK PPM DPPM DPIM and DH-PIM

rate tend towards to DPIM when M is greater than 4. With the increasing of M, the error rate of PPM, DPPM, DPIM and DH-PIM (α=1) tend to the same.

VII. CONCLUSION

In this paper, combining the wireless optical channel characteristics of the atmosphere, the performance of bandwidth demand, transmission capacity, power requirements and error rate is analyzed on the five kinds of modulation of atmospheric optical wireless communication. Theoretical analysis and simulation show that OOK is the easiest way without symbol synchronization, the bandwidth demand is the minimum and the transmission capacity is the maximum, but the power utilization is too low and the error rate is large; the power utilization of PPM with symbol synchronization is greatly improved but the bandwidth utilization is the lowest; comparing to the OOK, the DPIM have the higher power efficiency, comparing to the PPM, the DPIM have higher bandwidth utilization, and the DPIM do not need the symbol synchronization at the receiving end; DPPM is relatively close to the DPIM in all respects; DH-PIM modulation method is an improvement of DPIM which use pulses of different lengths in the head slot, the average transmission power and error rates of DH-PIM are slightly higher than the DPIM, but more good bandwidth efficiency is gotten more than DPIM and PPM, its advantages become more pronounced with the symbol length (M) increases. The several modulation methods have their own advantages and disadvantages, and therefore DPPM, DPIM and DH-PIM (α = 1) have more advantages, more suitable for future wireless optical communication systems.

REFERENCES

[1] Sun Yi, Wu Lei. Simulink Communication Simulation Development Manual [M]. Beijing: National Defence Industry Press, 2006.10.

[2] Hu Zongmin, Tang Junxiong. Atmospheric optical wireless communications systems in the digital pulse interval modulation [J]. Communications, 2005, 26 (3):75-79.

[3] Wang Hongxing, Zhang Tieying, Zhang Tieying, et al. Wireless Optical DH-PIM with DPIM modulation performanceStudy [J]. Laser Technology, 2007, 31 (1):95-96.

[4] HU ZM, TANG JX. Digital pulse interval modulation for atmospheric Optical wireless communication [J] .Journal on communications, 2005, 26(3):76 ~77.

[5] Li Yuquan, Zhu Yong. Optical Principle and Technology [M]. Beijing: Science Press, 2006.

[6] Yang Xiaoli. Optoelectronic Technology Foundation [M]. Beijing: Beijing University of Posts and Telecommunications Press, 2005:115-119.

[7] Ke Xizheng, Xi Xiaoli. The Survey of Wireless Laser Communication [M]. Beijing: Beijing University of Posts and Telecommunications Press, 2004:148-157.

[8] Li jianxin, Liu naian. Analysis and Simulation of Modern Communication System [M]. Xian: Xi Dian University Press,2004:48-58.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 361-364

Design and Implementation of Vulnerability Scanning Distributed System

Zhang Ping [1], Tao Bin [2]
[1] College of Henan Engineering,Zhengzhou, China
Email: zpings@sina.com
[2] College of Henan Engineering,Zhengzhou, China
Email: tb3190@126.com

***Abstract*—In this paper, we discussed the basic principle and the common shortcoming in the existing network vulnerability scanning systems. Based on these, distributed system based components of network vulnerability scanning system and the communication and collaboration of Agents are been described in detail. The system uses plug-in technology to strengthen the expansibility. Its key factors are friendly interface, faster scanning, and particularly scanning results in detail.**

***Index Terms*—Vulnerability Scanning, Plug-in technology, Agent Technology, Distributed System**

I. INTRODUCTION

At the present stage, the network information systems are increasingly complex, increasingly rich Internet applications, and network security problems exposed more and more drawn worldwide attention. The study of computer security, in particular how to improve the safety performance of the computer must first be to know ourselves. Network scanning is an important way to study the network security situation of other computers and it could initiatively find vulnerability in the system and promptly repair [1]. Therefore, network scanning play an important role in the research of network security.

II. WORKS AND LIMITATIONS IN NETWORK VULNERABILITY SCANNING SYSTEM

A. scanner works

Network vulnerability scanner test target remote host TCP/IP services to different ports, recording the answers given. In this way, many target host can collect all kinds of information (such as: whether the anonymous login can be used, whether the IP directory writable, it can use Telnet, http it is root run). Access to target host TCP/IP ports and their corresponding network access services related to information, it should match to vulnerability scanning and network vulnerability database system, and vulnerability exists if the matching conditions are considered. In addition, it is one way of achieving scanning module through simulated hacker attack techniques to attack on the target host system of security vulnerability scanning, such as weak passwords and other test. Vulnerability exists if the simulated attack succeeded.

B. The existing limitations of network vulnerability scanning system

Existing network vulnerability scanning system (Nessus, ISS, etc.) and more of a centralized or C/S structure, the central control node or server side is responsible for vulnerability scanning and store the scan results. The serious flaws of the structures is that when more scans the target host, the control node or server-side information processing will become a "bottleneck", resulting in performance degradation. Therefore, the larger network (such as the Intranet) for network vulnerability scanner scans often requires multiple servers to work together. At the same time, regarding the needs for managing conveniently, it requires all the scan servers in network can be remotely controlled. To solve these problems, we could introduce Agent technology to design and implement distributed network vulnerability scanning system. Using the Agent to develop the function module in vulnerability scanning, and through the relevant Agent to complete vulnerability scanning, it could avoid performance "bottleneck" and increase the scanning efficiency for the completion of large-scale network vulnerability scanning task.

. DISTRIBUTED SYSTEM MODEL FOR NETWORK VULNERABILITY SCANNING

With the multi-Agent structure model, each scan run on host vulnerability scanning task are independent, achieving the distribution of network vulnerability scanning technology.

In this paper, the network vulnerability scanning model for self-Agent for the organizational unit, according to their different functions, they could be classified into five categories: Scanning Agent (SA), Communication Agent (CMA), control scheduling Agent (CDA), state inspection Agent (SCA) and the User Interface Agent (UIA). Network vulnerability scanning process is as follows: The system model based on Agent, Agent mutual cooperation to accomplish a variety of network vulnerability scanning task. Users enter the network vulnerability scanning tasks through interface Agent, and it could control Agent and broadcast the scanning operation mission that will be submitted, and conduct consultations Agent task decomposition with the other scanning the server control scheduling. After the consultation process and send the task decomposition by Agent to complete the

communication. Scanning Agent responsible for tasks related to vulnerability scanning, and scan the results to control the scheduling Agent; control scheduling Agent collect all the scan results and send to the user interface Agent, the Agent user interface, report on the results generated and displayed to the user. The vulnerability scanning process of the entire network could be shown in Figure 1.

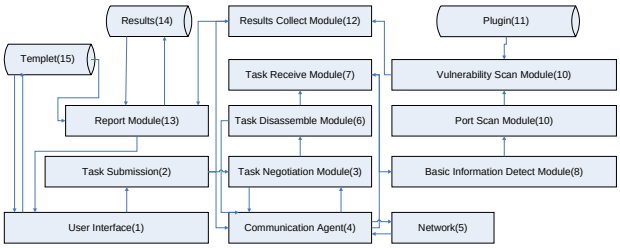


Figure 1. Network vulnerability scanning process.

. IMPLEMENTATION OF THE VARIOUS COMPONENTS

A. Communication Agent

Communication Agent specializing in communications services, and there is only one communication Agent in each scan host. When the Agent on the host with other hosts on Agent Communication, it is first distributed data communication Agent, then Agent communication objectives according to the data to be sent forward to the target host communication Agent, Agent communication on the target host again purpose of the data sent Agent. Communication Agent records the Machine Agent communication and collaboration in the host contact, it can provide routing services for data packets and its main task is to receive and transmit data without controlling ability. The contents of their communications for a quaternion group [2], that <communication content> ∷ = <sender> <recipient> <time> <data flow>.

When the Agent to communicate with other Agent, it sends the packet to this machine communication Agent, Communication Agent under the destination address for forwarding. If it is a local Agent, the Agent receiving forwarded to the destination port, otherwise the communication transmitted to the target host Agent, then forwarded to the purpose of the Agent Communication Agent. Agent communication between the content of mainly short information, such as Agent task control scheduling and task allocation of broadcast information, these short messages are encapsulated in UDP packets, sent to the Communication Agent. Large amount of data transfer through TCP protocol, such as Agent Update Agent to send another update when the NASL script vulnerability update the Agent may request that the machine's communication with the destination host on Agent Communication Agent to establish TCP connection, and then to transfer files.

B. Control Scheduling Agent

Control Scheduling Agent is use to manage the vulnerability scanning scheduling tasks. Since this system uses no central control of distributed structures, no specific task to complete submission of the host and the final results of the collection, each scan host must have a control task scheduling Agent to be responsible for scheduling. Control Agent is responsible for scheduling the control of the host with the collaboration Scheduling Agent to interact and coordinate task decomposition and scheduling. Scan a large network into multiple small tasks need to scan more than one task can be executed in parallel host scanning, receiving the control task scheduling Agent responsible for the coordination of other hosts, according to their computing power decomposition task, the task will be decomposed Send to a host of other collaborative control of scheduling Agent. Scheduling Agent receives control mission, the task of redistribution to multiple scan Agent, Agent on the target by the scan to scan. After the distribution in the scanning machine task is completed, control scheduling Agent will scan results back to the receiver control task scheduling Agent.

The first task Control Scheduling Agent is to inspect its own security system Agent [3], and it is only on each host. It checks the communication collaboration Host Agent and the Agent of the machine state, and is responsible for reporting to the administrator. Since each host is only one communication Agent, ensuring the normal operation of Agent Communication is the focus of the entire security system. If Communication Agent is damaged, then the host will not be able to collaborate with other hosts. Similarly, the normal operation of the machine Agent is also important, regularly check the state examination Agent communication, and collaboration with the host state of the local Agent, if there are unusual circumstances arise, it is responsible to the administrator requesting that the administrator check the errors.

C. User Interface Agent

Agent user interface is user-oriented; it provides users with a friendly graphical interface. Provides three functions:

- Interactivity: provide users with a graphical interface to interact with . The user interface on the input that wanted to scan the target and a number of control parameters, the system would submit the task of user input to control scheduling Agent.
- Vulnerability Update functions: update for vulnerability information. Vulnerability information to update vulnerability scanning system is an essential feature [4], we use the NASL (Nessus Attack Scripting Language) to describe the vulnerability of information and store in NASL script file. Updating of the machine's vulnerability keeps consistent with other servers in vulnerability information, and it would send vulnerability updates information to the control scheduling Agent which sends message to inform other servers update the vulnerability database through the communication Agent.
- Report generation function: undertake the results of the scan into results reported , and generate different types of results reported according to different templates. Report on the results can be divided into three types, namely,

managerial, management-level and technical staff level. Different levels of reports describing the results of different depths and different forms of expression.

D. Scan Agent

Scanning Agent is the basic function of this model unit. The system has a plug-in library to store all the script plug-ins and can be extended by adding plug-ins. According to user's requirements and parameters settings to read from the database that corresponds to the script plug-ins, and sort into the right plug-dependence, and then interpreted plug-in exploit code, and scan and detect vulnerability on the target system the through simulated attack.

1) Detection modules of basic information

Detection modules of basic information complete some basic information to detect on the target host, such as whether to boot the operating system types [5]. The module determines whether the target host is used in turn to the target host to send ICMP echo request packet, it indicates the target host is active if you receive a corresponding ICMP echo request packet. Some personal firewall installed on the target host can be shielded such messages, resulting in the illusion that did not start. However, this system is mainly aimed at providing network services to host, so that the host usually does not block ICMP echo packets. Modules also roughly determines the target host's operating system according to ICMP echo reply message in the TTL set, such as the TTL value is close to 256, the target host for the UN IX system; TTL value is close to 128, the target host is the Windows system [6].

2) Port scan module

The control parameters of the corresponding platform came the 1-1024 or 1-65535 scanning TCP port or a special user-defined port when the port scan sub-module is called. Scanning mode uses an open port scanning techniques, namely, using TCP connect scanning technology to design the scanning module, which is the most basic of TCP scanning. Usually by calling the socket function connect () to connect to the target computer and complete a full three-way handshake process[7]. If the port is in listening state, the connect () would return successful, otherwise, this port is not available that does not provide services. One of the biggest advantages of this technology is that it does not need any permission and any users of of system have the right to use this call. Another advantage is more stable and reliable than other scanning methods (such as port scanning or half hidden prescribing port scanning). However, one of the defects of this approach is: scanning does not cover. Often used as a scanner software applications, TCP's connect would repeat and use in a concentrated way and such scanning behavior is easily to find at one end of scanning, and the target computer's log file would show whether a series of connections and connection service error message and close it quickly. Besides, with the rapid development of firewall technology most firewalls can shield these scans. Some other way of scanning the firewall those have been considered subtlety might also be identified and screened out [8]. However, we are developing scanning systems from the system administrator's view, so the above problems are not existent unless it is used to scan other hosts illegally.

3) Vulnerability scanning module

The system combines several common classification methods combined with the causes of vulnerability as well as direct threat and harm to the system brought by vulnerability, so the vulnerabilities could be classified as follows: Finger abuse, Windows attacks, backdoor, CGI abuses, remote file access, RPC, firewall, FTP, SMTP, access to remote root, a denial of service and other categories. The same type of vulnerability scanning is a plug-in [9]. The main process of implementing the plug-ins could be demonstrated in Figure 2.

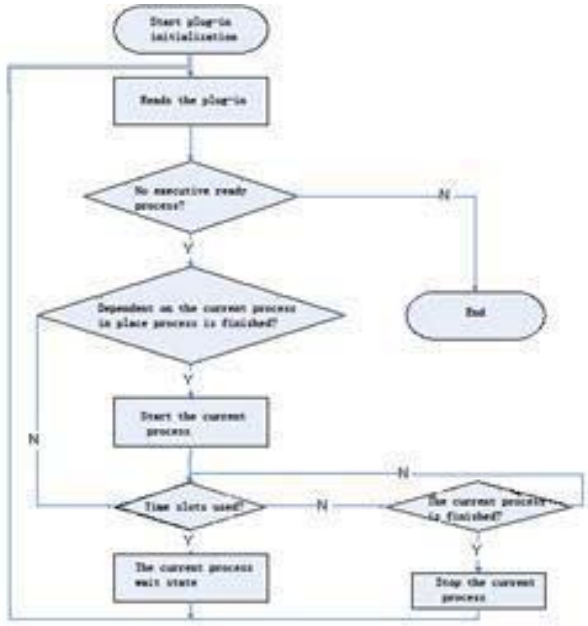


Figure 2. The main process of implementing the plug-ins.

4) Scanning plug-in library

This is a major part of the system, through calling the library vulnerability scanning scripts the system formatted a vulnerability scanning plug-in by using simulated attacks, As long as in accordance with the plug-in interface standard provided by software, the third-party included software developers or programmers individually can easily add functions for the software. Plug-ins are generally presented in the way through Dynamic Link Library (Windows platform is a DLL, in the Linux/Unix

TABLE I.
PART OF THE SCRIPT TAG

Script tag	Function
Scrip_t_id()	Scan tag script
Scrip_t_scantype()	Script Scan type
Scrip_t_description()	Vulnerability description of the corresponding script
Scrip_t_port()	Determine a particular type of port state

environment is a share object library) [10]. Plug-ins could run only be attached to the host program and could not be run independently. Users are free to increase or uninstall the existing plug-ins, but the application does not need to re-compile and link. Part of the script tag in Table 1.

This plug-ins provides a unified interface, offering conveniences for the scanning module, and the expandability becomes strong and provides the extent of vulnerability and repair program.

5) Scan results recording module

The system will scan statistics to document for the administrator to check.

Ⅴ. CONCLUSIONS

This paper proposes a network vulnerability scanning system, which uses distribution structure included multiple scanners host and completes the collaboration between the host through the specific communication, at the same time ,it also to protect the security of each module and communication in system with certain security mechanisms. The system has good scalability, which can easily by adding new plug-in to add a new host and new collaboration vulnerability detection method. The system overcomes the traditional limitations of vulnerability scanning system and is suitable to complete the vulnerabilities scanning task of large-scale network.

REFERENCES

[1] Fanglu Guo,Yang Yu,Tzi- cker Chiueh.Automated and Safe Vulnerability Assessment[J].IEEE Computer Society,2005.

[2] Michitaka Yoshimoto,Bhed Bahadur Bista,Toyoo Takata. Development of Security canner with High Portability and Usability [J].IEEE Computer Society,2005.

[3] Andrew Storms.Using Vulnerability Assessment Tools To Develop an OCTAVEòRisk Profile [C]. SANS Information Security Reading Room,2004.

[4] Andrew Storms.Using Vulnerability Assessment Tools To Develop an OCTAVEòRisk Profile [C]. SANS Information Security Reading Room,2004.

[5] Anonymous. Maximum Security (4th edition) [M]. SAMS, 2002, 2212230.

[6] Liu Bo, Liu Hui, HU Huaping, Huang Zunguo. Design, implementation and application of Computer Vulnerability Database system [J]. Computer Engineering and Science, 2005. 26 (7): 31 - 33.

[7] Cao Yuanda, Li Xianfeng, Xue Jingfeng. Plug-in technology in vulnerability scanner research [J]. Microcomputer development, 2005,15 (9): 72 - 74.

[8] Zhang Yuqing, Dai Zufeng, Xie Chongbin. Security scanning technology [M]. Beijing: Tsinghua University Press, 2004:10 - 11.

[9] Chen Tieming, Cai Jiaying , Jiang Rongrong, Feng Xiancheng. Plug-in based security scanning system full of holes. Computer Engineering and Design. 2004 (2) :194-196.

[10] Ma Hengtai, Jiang Jianchun, Chen Weifeng, et al. Agent-based Distributed Intrusion Detection System [J]. Journal of Software, 2000, 11 (10): 131 221 319.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 365-368

The Research of Agent Union Algorithm Based on MAZE

Xue Xiao[1,2], Li Huiqin[1]
[1](College of Computer Science and Engineering, Henan Polytechnic University, Jiaozuo Henan 454000, China)
[2](National CIMS Engineering Center, Tsinghua University, Beijing 100084, China)
E-mail: {jzxuexiao, lihuiqin6}@126.com

***Abstract*—The Collaboration of Multi-Agent is one of the most important issues in the study of Multi-Agent Systems, and agent union has become the priority and hot spot as a form of collaboration when studying multi-Agent system theory and technology. Though, there exists a variety of coordination algorithms, it is difficult to compare and verify these algorithms. So, the simulation platform called MAZE is developed, that users can use it to compare and verify various algorithms about Agent. Although there have been a variety of Agent infrastructure platforms, such as JADE, which has been widely used. However, it is difficult to extend and customize the customer-centric application based on these platforms. This paper will summarize some common patterns (e.g. coalitions)to support the development of the MAS system by the design and implementation of MAZE which is a multi-Agent system simulation platform. Users can test and compare various Agent Alliance algorithms.**

***Index Terms*—Agent union, platform of Agent, algorithm simulation**

I. INTRODUCTION

The Agent theory and technology has been studied as a research field of artificial intelligence since the late 1970s. Agent technology is developing so rapidly that it has become a hotspot among the domestic and foreign scholars due to its important role in the domain of computer science[1].Agent coordination occupy an important position in the multi-Agent system development, so the research on the synergetic algorithm about Agent has become the focus of domestic and foreign scholars. At the same time, Agent union has also attracted extensive attention as a form of Agent coordination. Because of the selfishness of Agent, we can not assume that each Agent have a common goal when developing Agent platforms, while, in order to complete its goal favorably, it also needs the help from other Agents and the co-operation. In the process of Agent coalition formation, task decomposition, the distribution of benefits, and the timing of coalition formation and dissolution have become the difficult and important point when researching Agent union. At present, various algorithms about Agent coalition have come up with the research from domestic and foreign scholars. However, it is difficult to compare and validate among these algorithms, so, the development of Maze simulation platform is for this problem. The simulation system based on the Maze that we have developed is designed by HDA model [2], therefore, the simulation and realization of Maze is a challenge for AO[3] to develop actual software methods.

With the development of modern computers, it is no longer a stand-alone system but a large-scale distributed system. The computer and information processing system are becoming increasingly complex because of the close relationship between computers and users. So, the traditional centralized model can not meet the need of the adaptation of large-scale distributed information processing systems, but the computing based on Agent and the high-level interaction led by Agent can do. Nowadays, JADE (Java Agent DEvelopment Framework) [4] is the most popular form of distributed Agent Platform. It is a software framework to develop Agents with a FIPA-Compliant Agent Framework[5], especially for the interoperability of intelligentized MAS [4].

The reminder of the paper is organized as follows: The second section describes the status of Agent coalition algorithm in domestic and foreign countries, as well as a brief description on Agent simulation platform. Next, the details of the Maze system are presented, and the design method of the Agent will be introduced thoroughly. The simulation of the secondary invite/bid mechanism that advanced by Qin Haiou is given in section four. Finally, the conclusion achieved by the experience is presented, and it describes the future direction of work.

II. RELEVANT WORK

Rosenschein and Zoltkin have made various research on Agent union since 1993 when Agent union has been proposed. Agent coalition has been one of the most important problems in the field of researching Multi-Agent systems, and Agent union of itself as a form of Agent coalition has been widespread concerned by scholars home and abroad. How to form the global optimal coalition formation is an important issue in the process of forming an Agent union. On the one hand, it needs to ensure that proceeds gained after the union has been formed increased. On the other hand, it also needs to meet the interests of individual Agent, which can guarantee Agents join the union voluntarily. At present, various algorithms are researched by domestic and foreign scholars, such as the method of Shapley[6]and the two

Foundation items: Provincial Science and Technology Department of basic research projects(092300410216); Doctor Fund of Henan Polytechnic University(648227); Project of young teachers of Henan Polytechnic University in 2009(649100)

sides auction algorithm put forward by Ketchple[7]. These algorithms became an important basis for the reference algorithm, though they have some limitations. Other domestic and foreign scholars also have made various Agent coalition algorithms, such as the strategy of DCF referred in literature[8], and a strategy of behavior based on coalition formation proposed by Luo Yi and Shi Chun-yi[9]. Wei Wei and other scholars proposed a strategy to form Agent coalition based on relation web model[10]. However, some shortcomings still exist in these studies: (1) It is not in accordance with the actual that assume the tasks among each other is independent. (2)The formation and dissolution of Agent alliance is accompanied by the states of the task, which result in the low utilization rate of Agent, and the temporal relationship among sub-tasks was not considered. (3)The distribution of the benefits after the task did not meet the maximize needs of selfish Agent.

Though a number of characteristics of these algorithms above have been presented, there is still lack some suitable platforms to make further comparisons of these algorithms. With the advancement of Agent technology, a number of Agent simulation systems have been developed at home and abroad, such as JADE developed by the Parma University, and ZEUS developed by British Telecom[2, 4, 11]. However, it is difficult to extend and customize the customer-centric application based on these platforms, which have been an obstacle for us to apply agent technology in practice. Therefore, this paper will summarize some common patterns (including autonomous mechanism, communication mechanism, collaboration mechanism, coordination mechanism)to support the development of the MAS system by the design and implementation of Maze, which is a multi-Agent system simulation platform.

In this paper, the secondary strokes/tender mechanism[12] advanced by Qin Haiou is simulated on the platform of MAZE to test the function of the mechanism.

III. THE SYSTEM OF MAZE

A. *System Overview*

Maze based on the scenes of officers catching robbers , which fully demonstrate the variety features of Agent and Multi-Agent, such as, the autonomy, initiative, communication capabilities, coordination and collaboration capabilities of Agent. Each Agent in Maze is corresponding to a soldier or a bandit, and they comply with the following operating rules:

- Each Agent has its own physical critical value, the speed threshold, the confidence of the critical value, location, personal sense of accomplishment, etc., which determines the autonomy and the collaboration of Agent, and these values are mutually related, for example, the higher physical critical value, the faster speed of the Agent;
- Each Agent belongs to two different organizations, they can find and attack their enemies within limitations, and the life of the Agent is over when their physical value decreased to zero;
- Each Agent is autonomous in a certain degree, they can apply to their teammates for help and also can give the assistance that their teammates requested, or refused;
- The task of Agent is to search the enemy and annihilate the enemies in the maze, according to their own situation, Agent can attack the enemies and ask their teammates for help or escape when they find their enemies;
- There is a safe area (the gray areas in the figure), where Agents can have a rest to restore their strength and life values;
- Agents can send messages to each other and annihilate the enemy collaboratively.

main interface of Maze is shown in Figure 1, it has four parts: ①the top part is the toolbar for users to draw the maze, controlling the run of Agent; ②the left is the parameter setting interface, which users can set the form of the organizational structure, the property value of Agent ,as well as the message format when Agent interacting; ③the right of the large view is the running demonstration interface, users can observe the process and the results of the Agent running after they have drawn a maze and put the Agent in the maze; the lower right of the main interface is the message display interface when Agent interact with each other, users can observe the frequency of interaction and the interactive content. Users can control the running of Agent by the buttons of "Start", "Pause" and "Stop" after the preliminary work has been done.

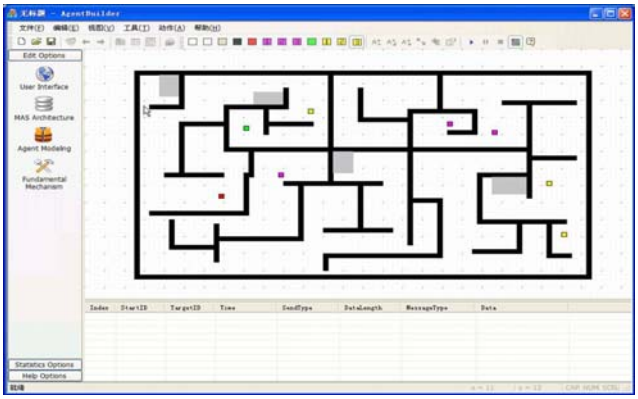


Fig. 1 Maze program's main interface diagram

B. *The design of Agent*

In order to keep users can determine the thinking and behavior of the Agent conveniently, so, when designing Agent, we need to ensure the autonomy, reactivity and preactivity of Agent; on the other hand, we also need to take the social nature of Agent into account to ensure that Agent can interact, collaboration and consultation with other Agents. According to the Agent architecture in the HAD[2], the class diagram about Agent is shown in Figure 2: Agent class inherits from the Entity class, which contains two modules: TeamModule and StateMachine. TeamModule is responsible for the function that Agent can interact with the outside Agent, and, StateMachine is responsible for the autonomy of Agent. Agent class

extends the basis of features by increasing behavior functions to achieve more advanced features, such as the nature of self-serving. Agent in each group only have one leader and three soldiers, users can add new Agents of their own needs, and give them different roles.

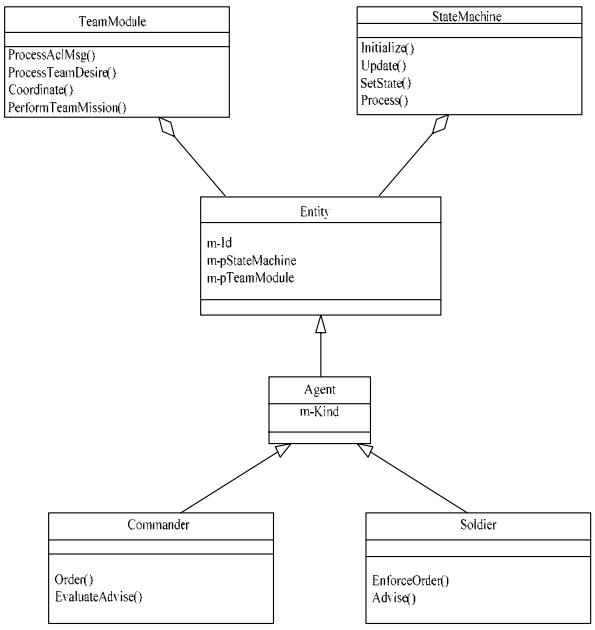


Fig. 2 The Agent model diagram in MAZE

Agent union emerged at the time when they need help from other Agents to finish its own tasks that it can not complete by itself. The union that has been formed should keep the interests of the league increased, on the other hand it also need to meet the interests of the member Agents. In this paper, the simulation platform MAZE can provide users with a good scene to test union algorithms. In the scene of the officers catching robbers, all the Agents are divided into two competing organizations: Police (soldiers) consisted of Commander and Soldier and the Enemy (robbers) consisted of Robber and Subrob. Each Agent has been given a role in the scene, and they fulfill their own responsibilities and obligations according to the role. There will appear Agent union among teammates when Agent is opposing to their enemies. About the problem of the distribution of benefits, users can consider that give some medals to the Agent according to its performance of the Agent who took part in the union, and can promote the Agent who has obtained a certain number of medals. So, the scene that provided in Maze can help users test union algorithms conveniently.

IV. THE SIMULATION AND ANALYSIS OF AGENT UNION ALGORITHMS

C. Simulation model

The theory of the algorithm was proposed in literature[12], but users can not make a decision about the function of the algorithm, and can not compare their algorithms with other algorithms. So, we give a simple simulation of the union mechanism that proposed by the literature[12].

Firstly, the task of Agents are allocated by the method that proposed by Chen Yuwu and Cao Jian[13], and then build the union according to the secondary strokes/tendering mechanism., the process is as follows:

(1)Task decomposition:

The Agents in the two groups have a general goal that destroying the enemy. The overall goal is composed of four sub-goals, that is, destroy each enemy. These sub-goals can carry out at the same time. However, the overall goal is finished only if all the sub-goals have been completed. So, the method that we adopted was the AND branch unit. Each Agent is in dependent before the beginning of the task, and the union is built when the task needs Agents to cooperate with each other, and Agents return to independent when the task is completed.

(2) The formation of the union mechanism:

Union is happened when completing the sub-goals, that is, an Agent need others help to destroy the enemy when it found its ability is limited:

1) The Agent sends tender notice to the contractor as a inviter.

2) Contractors make the biding document after they check the cost.

3) The inviter evaluates the biding document, and select the appropriate contractor, and send a confirmation message to the contractor.

4) The contractor begins the project after it has received the confirmation message, and gets the remuneration that from the inviter when the task has been finished competently.

5) The contractor give a cooperation evaluation about this task.

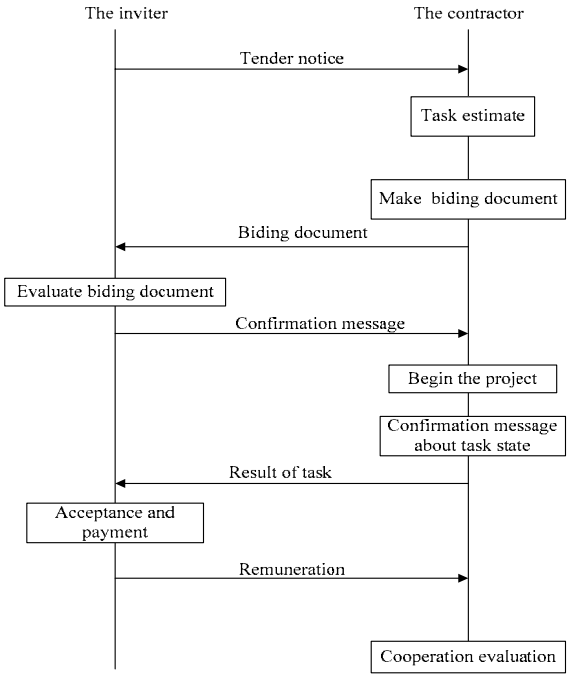
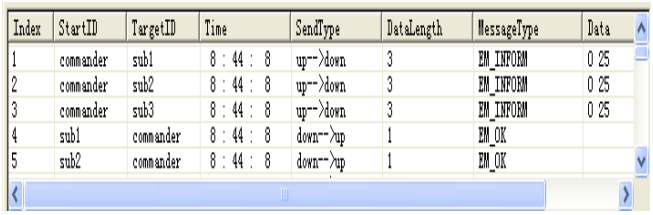


Fig. 3 The formation of union

D. Analysis of results

The state of their communication of attack each other is shown in Figure 4:



Index	StartID	TargetID	Time	SendType	DataLength	MessageType	Data
1	commander	sub1	8：44： 8	up-->down	3	EM_INFORM	0 25
2	commander	sub2	8：44： 8	up-->down	3	EM_INFORM	0 25
3	commander	sub3	8：44： 8	up-->down	3	EM_INFORM	0 25
4	sub1	commander	8：44： 8	down-->up	1	EM_OK	
5	sub2	commander	8：44： 8	down-->up	1	EM_OK	

Fig. 4 Message view

According to the message view, the state of their communication present a distribute shape, and does not require a central decision-making Agent. So, this algorithm meet the simplicity and distributivity.

V. CONCLUSION

In the field of multi-Agent system, Coordination and collaboration of Agent has been the concern of the domestic and foreign scholars. So, Agent union became a research focus as a form of collaboration. Although various algorithms about Agent union have exist, the platform for users to compare these algorithms are lacking in. Therefore, the simulation platform that this paper addressed is developed in allusion to this problem. However, some shortcomings still exist in the platform, that users can not observe the characteristics about the union algorithms, such as, effectiveness, stability, timeliness and non-reducing. So, the next step is to solve the shortcomings above, and make MAZE a perfect simulation platform.

REFEREBCES

[1] Glaser, N.(1996).The CoMoMSA Methodology and Environment for Multi-Agent System Development. Zhang, C., and Lukose ,D.(Eds.)Multi-Agent Systems and Applications III.

[2] Xue Xiao, Agent-oriented software design and development method, Electronics Industry Press, 2009.1.

[3] Paul Kearney, Jamie Stark, Giovanni Caire, Francisco J. Garijo, Jorge J. Gomez Sanz, Juan Pavon, Francisco Leal, Paulo Chainho, and Philippe Massonet. Message: Methodogy for engineering systems of software Agents. Technical Report EDIN 0223-0907,Eurescom,2001.

[4] Bellifemine, F., Poggi, A., and Rimassa, G.(2001). Developing Multi-Agent Systems with a FIPA-Compliant Agent Framework. Software Practice and Experience,31,pages 103-128.

[5] Poslad, S., Buckle, P., and Hadingham, R.(2000).The FIPA-OS Agent Platform: Open Source for Open Standards. Available at http://fipa-os.sourceforget.net.

[6] Shapley L S.A value for n-person games. In: Roth A E ed. The Shapley Value. Cambridge: Cambridge University Press,31-40.

[7] Ketchple S. Forming coalitions in the face of uncertain rewards. In: Proc AAAI-94, Seattle ,US , 414-419.

[8] Klusch Matthias, Gerber Andreas Dynamic coalition formation among rational agents[J]. German Research Center for Artificial Intelligence,2002,17(3):42-47

[9] Luo Yi, Shi Chunyi. The Behavior Strategy To Form Coalition In The Agent Cooperative Problem-Solving .In:Chinese J. Computers, 1997, 20（11）:961-965.

[10] Wei Wei, Liu Hong. Strategy to Form Agent Coalition Based on Relation Web Model [J]. Computer Application , 2006, 23（10） : 41-43.

[11] Nwana,H.S.,Ndumu,D.T.,Lee,L.C.,and Collis,J.C.(1999).ZEUS:A Toolkit for Building Distributed Multi-Agent Systems.*Applied Artificial Intelligence Journal*,1(13),pages129-185.

[12] Qin Hai-ou. A New Formation Mechanism of Dynamic Agent Coalition. In JI SUAN JI YU XIAN DAI HUA,2009,168(8)：161-168.

[13] Chen Yu-wu, Cao Jian. An Algorithm for Coalition Formation with Complex Tasks. In:Computer Engineering&Science,2010,32(5):72-76.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 369-372

Research of Distributed Algorithm based on Parallel Computer Cluster System

Xu He-li[1], Liu Yan[1]
[1] School of Computer Science and Technology/Henan Polytechnic University, Jiaozuo 454003, China
xuhl@hpu.edu.cn
hpuliuyan@126.com

***Abstract*—Parallel computer cluster technology is an important development direction of high-performance parallel computer system. Parallel computer system is the best choice in institution, which has the high frequent operation requirement. This paper gives the ideal scheme based on the analysis for the complex degree of common distributed algorithm, and points out the advantages of distributed algorithm applied on the high-performance parallel computer system.**

***Index Terms*—parallel computing, cluster system, distributed arithmetic**

I. High Performance Parallel Cluster System and Distributed Algorithm

The variation of parallel computer system structure development is very quick, which mainly embodies in two aspects. One is the improvement of performance in calculating node, and the other is the enhancement of communications technology between nodes. Over a long time, large-scaled integrated circuit technology has been developed with high speed in accordance of Moore's law. The development of element density of the chip and clock frequency leads the improvement of microprocessor performance which is as parallel computer basic dealing unit. At the aspect of communication technology, the speed of switching for traditional crossover switch improves quickly. And the new high-speed network technology and application to parallel computer, thus greatly improve the rate of communication between the nodes. Parallel computer is a set of communication and mutual cooperation processing unit which with the rapid solution of large the problem. Computation system could be contained multiple processors of a computer, it could also be a cluster that interconnected a number of independent computer. Requirements of mankind in calculation and performance are endless. From the angle of the integrated system: system resources to meet growing demand for performance and functions. From the angle of the application: appropriate into the application to implement the larger or more careful calculation. Cluster is a set of computer that put them as a whole to provide users with a group of network resources. Cluster systems are generally be divided into four parts : Management unit or node, the unit or node, network and software of cluster management. Cluster system has the virtue of high scalability, high performance, high cost efficiency, high availability, etc, that has attracted increasing attention.

Linux can operate on the very popular computer without purchasing the expensive hardware equipment. Linux cluster which possesses strong reliability and load capacity can be constituted based on adding corresponding cluster software to the several computer moving Linux system. Linux cluster technology gives full play to computer and internet's speciality. Nowadays, Linux parallel cluster system has been the most popular high performance computing platforms, and takes the great proportion among high-performance parallel computers. Linux cluster system can be as cheap parallel program test environment, also can be designed to real high-performance parallel. Its system scale can be single, few networked computers until includes thousands of large-scale parallel system. The nodes of the calculations are used for high performance cluster system in structure and use the software tools usually different networks, database used to provide the service.

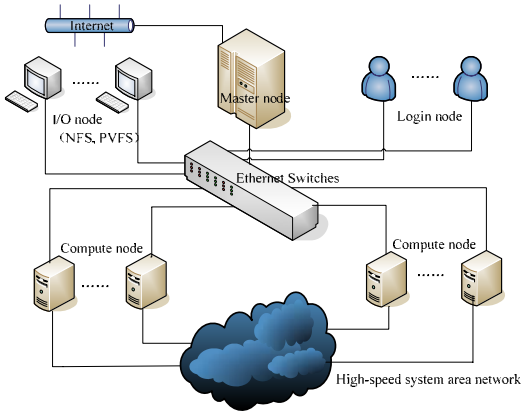


Figure 1. Classic Linux cluster system

Distributed algorithm is sharing information mutually in two or more software. This software can be run not only on the same machine, but also on several computers which are connected through internet. Distributed algorithm mainly studies how to make a problem that needs very huge calculating ability to solve into many small parts. Then put these parts into a number of computers to process. Finally, put these calculations together to get the ultimate result. Distributed algorithm to raise compute power is made full use of computing resources out of stand-alone. It needs not only hardware support, but also software design and support of the process structure itself. Tightly coupled to a serial calculation, because of the close correlation of data and

control, this process is difficult form break and increasing the complexity of the software design in parallel environment. Consequently, a process in support of parallel is to improve the system compute power of the cluster. If people make use of some characteristic of distributed arithmetic in cluster system to develop arithmetic that have distributed parallel computing power. It will improve the cluster system's computational power and solve larger and more complex system optimization and design.

II. TWO DISTRIBUTED ALGORITHM AND THEIR RELATED ANYLASIS OF COMPLEX DEGREE

The distributional algorithm is to complete some tasks in coordination with several nodes linking through communication. In the distributional algorithm, supposing the record element is distributes in certain nodes in the local storage, during various nodes may the arbitrary form interconnecting network connect group of advancements the correspondence to exchange the information by way of a fixed group correspondence C. Correspondence both sides agree in a transmission, in another receive regulations. Making N = (P, C) is a communication network (P is group of advancements, C is group of channels), the data-in I distributes in each advancement, the so-called distributional algorithm is opposite in N and I pair of question Q solution.

A. Distributional determines the k- selection algorithm

Algorithm 2.1 MIMD-AC on model determination k-selection algorithm

Input: $B = \{b_1, \ldots, b_n\}$, $S = \{s_1, \ldots, s_p\}$, $L \subseteq S \times S$

Output: The Kth element

Begin

- Through to the scanning that has a spanning tree, the root node may calculate the total element number: $|B| = \sum_{i=1}^{p} |Bi|$. If $|B| = 1$, notify the root node in the element node elements to the root node, the algorithm end, Otherwise the implementation steps;
- Each process is distributed in five elements in the bureau of deposit (in the local storage). But because the process may have remnant, so total remnants may be big, it is only a fraction and assumptions at most. To solve this problem, can make every process gets from child node at first, and then make five to a group, after its parent node to change. So may have O (p) exchange;
- Local to each of the five elements in value;
- Takes the parameter by M, the recursion transfer asks in M value m;
- Each advancement i of its bureau saves the element to divide three according to m sub-to gather BLi, BEi, BGi, they include separately <, =, > m these elements. Through to the spanning tree from the leaf to a root scanning, may calculate in the root node $|BL| = \sum_{i=1}^{p} |BLi|$, $|BE| = \sum_{i=1}^{p} |BEi|$, $|BG| = \sum_{i=1}^{p} |BGi|$. Once the |BL|, |BE|, |BG| calculate, root node can accord B' and C' decision algorithm based on selected k m, continued to end recursion. The root node to all other nodes broadcast this decision, so I know every node in BLi, BGi and set which should be as the next recursively parameters, the step to exchange information for O(p);
- According to the new parameter B' and k', algorithm may automatic recursive calls. In distributed environment, the recursion transfers when its entrance and the export complete by the root node. It distributed counts the existing active element number. If it is many, hen the root node notice other nodes, their recursion transfers their partial procedure. When only left over an element, the root node makes other nodes this element transmission for it, thus obtained k various elements. This time each advancement might transfer from the recursion promotes does not need with the root to further discuss then ended.

End

B. Median algorithm for distributed

Algorithm 2.2 Distributional asks the value algorithm on the MIMD-AC model

Input: A and B series.

Output: Distribution of m.

Begin

- PA through correspondence receive from PB B in value element 2 (B); PA 1 (A) carries on a in value element with 2 (B) the comparison; If 1 (A) < 2 (B), then $\lceil n/2 \rceil$ smallest elements of A join to A1; If 1 (A) >2 (B), then 1+ $\lfloor n/2 \rfloor$ biggest elements of A join to A2; If 1 (A) =2 (B), then $\lceil n/2 \rceil$ smallest elements of A join to A1.
- PB through communication from the value of the PA element 1 (A), The median PB will B elements 2 (B) and 1 (A) were compared; If 1 (A) < 2 (B), then $\lceil n/2 \rceil$ biggest elements of B join to B2 ; If 1 （A）>2 （B),then 1+ $\lfloor n/2 \rfloor$ smallest elements of B join to B1; If1（A）=2（B）,then $\lceil n/2 \rceil$ biggest elements of B join to B2, $\lfloor n/2 \rfloor$ smallest elements join to B1.

End

III. THE IMPLEMENTATION OF DISTRIBUTED ALGORITHM IN CLUSTER SYSTEM

A. *The differences and relations of distributed computing and parallel computing*

The distributed computing and the parallel computing are different. The goal of the parallel computing is to solve the single problem by using the multi-processor. However, the goal of the distributed computing is mainly to provide the convenience, including usability, reliable and physical distribution. The interactions between processors are frequent in parallel computing. It was usually have a fine granularity, low cost and considered reliable. While in distributed computing, the interactions between the processors are not frequent. It was have coarse granularity interactive features and often thought to be unreliable. Parallel computing pays attention to the short execution time, but the distributed computing focuses on the normal operation time.

Certainly, the parallel computing and distributed computing are close related. Certain characteristics relates to the degree (interactive frequency among processors), although we have not carried on the explanation to this kind of intersection. Another characteristics are relate to emphasis （speed and reliability), and we all know that these two characteristics are very important in parallel and distributed computing system. Thus, the two different types calculating system behalf a point which different but adjoin in a multidimensional space.

B. *Analysis the complexity of the tow types distributional algorithm*

The complexity measurement of distributed algorithm is standard by the overall sending messages. The complexity of algorithm contains time, space and traffic. Complexity is the major communications costs in distributed algorithms.

Distributional determines the k- selection algorithm the order of complexity analysis: Supposes |B|=n, then the order determined that k- choice recursion transfer number of times f (n) is: $f(n) \le 2+f(n/5) + f(3n/4)$,so $f(n) \le O(n^{0.9114})$. Each time the recursion transfer needs O (p) news exchanges. Therefore algorithm 2.1 need the news exchange number is $O(p \bullet n^{0.9114})$. The algorithm spatial order of complexity is linear. Storage space which needs as for the algorithm, each time the recursion transfers when regarding studies is p each advancement involves at most to 4p various elements. Even if therefore the number of degree regards as a constant, in certain advancements, possibly accumulates $O(n^{0.9114})$ various elements. Therefore the algorithm 2.1 space requirements are $O(n^{0.9114})$.

Distributional strives for the value algorithm order of complexity analysis: Makes A_i and B_i is P_A and P_B ith step asks time the value subset. Uses the present algorithm, even if time value, P_A and the P_B ith iteration occupies c|Ai-1| the time (c is with some constant which i has nothing to do with). Because $|A\ i\text{-}1| \ge 2|Ai|$, P_A needs the total time is O (|A|), therefore the algorithm total time is also a linear function. PA needs the time was probably equal to that asks in A and B by the non-distributional algorithm the value time. It is not difficult to see the algorithm the spatial order of complexity is also linear. At each iteration, P_A and P_B swap news, the scale of reduced problems for at least 1 1/2, and exchange information for a total of 2log n. Rodeh has proved distributed in the value for the communication cost is lower, so log n in constant factor 2.2algorithm is best. The total communication cost algorithm for the exchange of information $\log_{k+1} n$.

K-choice of distributed determination algorithm is applied to cluster system, the synergy of recursion inlet and outlet in the root. The root node distribution of active element existing technology, if not, then the remnant of the root node will inform all other processes to these elements, and then the root node of the first k. If still have many active elements, the root node and notify all other processes are called recursively local program. Median algorithm is distributed for distributed determine k - selection algorithm of $k = \lfloor n/2 \rfloor$ special case, in the extent than k - selection algorithm. The algorithm is suitable for high speed communications with the parallel computer application, or through the Ethernet connection between PCs and WSs execution. Study these two algorithms is to reduce the communication cost and Clusters of environment based on performance evaluation.

C. *Application of distributed algorithm in the cluster system*

Cluster computing system involved different systems architecture. For some users, cluster system is a multiprocessor collection which closes integration and work together to solve a single problem. To other users, cluster system might mean a computer network which consists by the separate processor. The processor join together in order to realize the resources sharing. Although the computing capabilities of high performance parallel system have increased, the rapid development of science and technology require more computing ability. There are three ways to improve the computing performance. The first is improved the device operation speed. The second is improved system structure. The third is focus on the computing ability to important area by using computational algorithm, and ignore some minor problems in order to improve the efficiency.

Distributed algorithm is a parallelism algorithm based on MIMD asynchronous communication model. This model can crystallize the distributed computing model. It means that the model contained in a processor network which have a unique identifier. But the processor can not understand the whole network overall. The only way of communication in network is processor and its neighbors to exchange information, and only limited but they do not know at that time. Suppose that one message will contain three different types of value： a starting value, an identifier and a number. Algorithms can start in the collection of nonempty processors and each processor has been worked out some function eventually. The characteristics of distributed algorithms are spatial scattered. When this feature combined with time parallelism, it is distributed parallel processing. In fact, distributed and parallel is inseparable. In parallel machine, the processor itself is decentralized also. In distributed

systems, the software's execution also cannot leave the synchronization. Early stage, high performance parallel computer cluster system is realized by parallel machine. Figure 2 showed a method of the distribution and parallel. It makes the system resources can be fully optimization, and exert the parallelism between nodes.

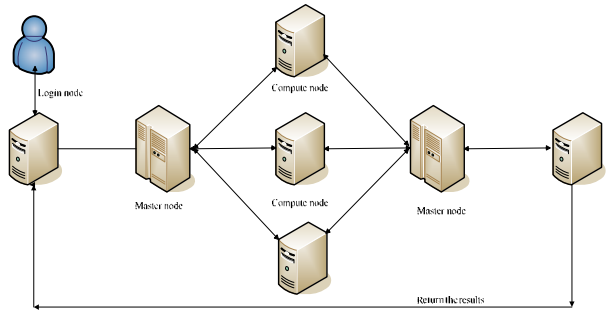


Figure 2. Classic Linux cluster system

In recent years, the high cost mainframe and data center become history gradually when the performance of computer increased. It is replaced by desktop computer and minicomputers. The function of parallel machine has been replaced by the distributed cluster system which consists of computer. Application of distributed algorithm in cluster system is a tendency in the future. However, distributed algorithm melt the network in different topology into Spanning Tree first. And then, design algorithm on Spanning Tree. Distributed algorithms and centralized algorithm have a big different in design of the methods and techniques, that is because the distributed system and centralized system have an essence distinguish in model and structure. The basic characteristics of centralized algorithm do not exist in distributed algorithm. Distribution and concurrency are two basic features in distributed algorithm. Execution of distributed system exist some instability factors. Because of these differences, the design and analysis of distributed algorithm are more complex and more difficult than centralized algorithms. These many questions remain to be solved.

IV. CONCLUSION

This article gives the analysis and discussion of the distributional algorithm based on application on the high performance parallel computer cluster system. Moreover, it introduces in detail about distributional determined the k- selection algorithm and distributional asks the value algorithm, and analyzes its algorithm complexity. The basis may enhance computer's performance and the computing power using the computation algorithm, proposed that applies distributional algorithm this viewpoint on the high performance parallel machine colony system. The colony is the isomorphism, and the distributional is isomerism. To apply the distributional algorithm on the cluster system to make it become the synthetically systematical construct which reaches to the parallel in spatial and time area in order to promote the parallel computer cluster system's performance.

APPENDIX A ALGORITHM COMPLEXITY

TABLE1. ALGORITHM COMPLEXITY

Algorithm	Time complexity	News complexity
Distributional determines the k-selection algorithm	Space complexity $O(n^{0.9114})$	$O(pn^{0.9114})$
Median algorithm for distributed	Space complexity O(n)	2log n

ACKNOWLEDGMENT

I would like to thank my colleagues on the HPU-HPC team for their contributions, insights, and support.

This paper is supported by the high-performance grid computing platform of Henan Polytechnic University.

REFERENCES

[1] Bansal S, Kumar P, and Singh K. “An improved two-step algorithm for task and data parallel scheduling in distributed memory machines,” *Parallel Computing*, 2006, pp.759-774.

[2] Gatani Luca, Re Giuseppe Lo, and Gaglio Salvatore. “An efficient distributed algorithm for generating multicast distribution trees,” *Proceedings of the International Conference on Parallel Processing Workshops*, 2005,pp.477-484.

[3] Li Yufeng, Qiu, Han, Lan Julong, and Yang Jianwen. “Analysis of the centralized algorithm and the distributed algorithm for parallel packet switch,” *Parallel and Distributed Computing, Applications and Technologies, PDCAT Proceeding*, 2006, pp.156-161.

[4] Czygrinow A, Hanckowiak M, and Szymanska E. “Distributed algorithm for approximating the maximum matching.,”Discrete Applied Mathematics, pp.62-71, September 2004.

[5] Al Hajj Hassan M, Bamha M. “An efficient parallel algorithm for evaluating join queries on heterogeneous distributed systems,” *16th International Conference on High Performance Computing, HiPC 2009 – Proceedings*, pp.350-358, December 2009.

[6] Talby David, Feitelson Dror G. “Improving and stabilizing parallel computer performance using adaptive backfilling,” *Proceedings - 19th IEEE International Parallel and Distributed Processing Symposium*, pp.84, April 2004.

[7] Vasupongayya Sangsuree1, Chiang Su-Hui1, and Massey Bart. “Search-based job scheduling for parallel computer workloads,” *Proceedings - IEEE International Conference on Cluster Computing, ICCC*, 2005.

[8] Talby David, Feitelson Dror G. “Improving and stabilizing parallel computer performance using adaptive backfilling,” *Proceedings - 19th IEEE International Parallel and Distributed Processing Symposium, IPDPS* 2005.

[9] Soveiko Nick, Nakhla Michel S,and Achar Ramachandra. “Comparison study of performance of parallel steady state solver on different computer architectures,” *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 2010, pp.65-77.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 373-375

Flexible Skinning Research in Reverse Engineering Based on Cross-Sectional Fitting

Wu Xiaogang[1] , Chen Dan[2] , and Zheng Chunying[3]
[1]College of Zhijiang , Zhejiang University of Technology, Hangzhou, China
E-mail: wxg@zjc.zjut.edu.cn
[2]School of computer & computing science, Zhejiang University City College, Hangzhou, China
corresponding author E-mail: chend@zucc.edu.cn
[3]Information Technology Department, Zhejiang Financial College, Hangzhou, China
E-mail:chunyingzh@126.com

Abstract—**To keep surface shape smooth and avoid transformation during dealing with surface reconstruction, an improved method about surface flexible skinning based on cross-sectional data is proposed. The method produces a continuous skinning surface, it improves the weakness that surface shape loses the true and calculation is unsteady, by running in practical system demonstrate the effectiveness of improvement.**

***Index Terms*—flexible skinning, algorithm, reverse engineering, surface reconstruction**

I. INTRODUCTION

Surface skinning is a proceed to construct a smooth surface using a set of cross-sectional curves, these cross-sectional curves may have different degrees, maybe is rational or non-rational, and maybe defined over arbitrary knot vectors, so surface skinning can produce many unexpected shape. In order to keep surface shape smooth, curves compatibility is necessary during dealing with different curves, skinning proceed can lead to largely increasing of number of control points, and it can produce distortion of surfaces shape, unsteady of calculation, and parameterization-related problems, etc.. In this paper an improved algorithm is proposed, the results demonstrate the effectiveness of the approach.

II. SURFACE SKINNING

In order to understand the skinning process, a few NURBS formulas will be concisely introduced, some detailed discussion about NURBS can be referred to corresponding references [1-6].

A NURBS curve of degree p is a piecewise polynomial curve defined as follows:

$$C^{w}(u) = \sum_{i=0}^{n} N_{i,p}(u) P_{i}^{w}$$

In the formula Piw，i=0,⋯,n form control polygon which is defined by a set of control points Piw with weight, Ni,p(u) i=0,⋯,n are the B-spline basis functions defined by a knot vector:

U=｛u0,⋯,um｝, ui≤ui+1 i=0,⋯,m-1

U is used as the following form:

U=｛a, a, ⋯ , a, u_{p+1}, ⋯ , u_{m-p-1},b,b, ⋯ , b｝

To extend NURBS curve , a NURBS surface of degree (p, q) is defined as:

$$S^{w}(u,v) = \sum_{i=0}^{n}\sum_{j=0}^{m} N_{i,p}(u) N_{j,q}(v) P_{i,j}^{w}$$

In the above formula Pi , j w i=0, ⋯, n ； j=0,⋯, m form control polygon which is defined by a set of control points Pi , j w with weight, Ni， p(u), i=0,⋯,n are the B-spline basis functions defined by knot vector U and V, U=｛u0,⋯,ur｝, ui≤ui+1 i=0,⋯,r-1 , V=｛v0,⋯,vs｝, vj ≤ vj+1 j=0, ⋯ , s-1 , wi,j is weight. To simplify computation we set weight wi,j to be 1.

NURBS surface skinning can be described as the following:

Given a set of sectional-curves:

$$C_k^w(v) \qquad k = 0,\cdots k$$

According to the value of parameter interpolate curves to get NURBS surface. if skinning process in U direction has finished, then the surface which we can get has following form:

$$S^w(u_k, v) = C_k^w(v), \qquad k = 0,\cdots,k$$

In the above formula uk , k=0, ⋯, k can be got through skinning process.

The algorithm of surface skinning can be described as following:

- given a set of cross-sectional curves $C_k^w(v)$ $k = 0,\cdots k$
- deal with curves $C_k^w(v)$ $k = 0,\cdots k$ compatibility
  get a knot vector V, the max index of control points is $\hat{m}$
- compute the knot vector U
  from j=0 to $\hat{m}$ repeat
    from i=0 to k repeat
      set R_i^w value with the jth control point
    end i
    interpolate curves $C_j^w(u)$ according to the knot vector U
  end j
- compute the surface control points $P_{i,j}{}^{w}$ i=0, ⋯, n ； j=0,⋯, $\hat{m}$ from $C_j^w(u)$ j=0,⋯, $\hat{m}$

- construct surface according to control points and the knot vectors

In the above algorithm the curve compatibility can be described as following:

- find out the max degree p of cross-sectional curves;
- use the Cohen algorithm[1] to increase degree , make curves degree raise to p. then change original knot vector into new knot vector raised degree;
- merge knot vector of all curves;
- refresh every curve using merged knot vector, use the algorithm of knots insert, make every curve have the same knot vector and the same number of control points.

The above skinning process may appear some abnormal phenomena, such as shape distortion; parameterization-confused of surface, appearing minus weight; surface continuity being very low; the amount of control points being largely raised, and so on. The reasons maybe uncontinual parameterization and the use of rational form. Some researchers have used non-rational curves which are continual prameterization to approximate cross-sectional curves, but these methods are not enough practical and effective. The problems maybe independent approximation will produce uncontinual parameters and largely raise the number of control points because of merging knot vector. The key to solve skinning problems is to supply a method of curve fitting which have continual parameterization and a changeable knot vector to avoid largely increasing of knots number, the following is the corresponding improved algorithm.

III. SURFACE SKINNING ALGORITHM BASED ON CROSS-SECTIONAL DATA

Firstly change cross-sectional data into point data. Because the form of Bezier curve is comparatively simple and not effected by knot attribution, so decompose NURBS into Bezier, the point data will be decided by curve border two degree derivative and tolerance error.

Based on point data, approximate NURBS curves, in order to avoid the largely increase of control points, a candidate knot vector will be passed, the algorithm can be expressed as following:

- compute the value of parameter based on point data;
- define original the number of control points,then use above approximate algorithm to approximate curves, and modify the candidate knot vector accordingly;
- compute the error of curves, modify the number of control points based on the error, make the error not exceed the permitted max error ε ;
- modify candidate knot vector by adding new knots;
- output the curve C and the modified candidate knot vector;

The key of this algorithm is how to select knots from a given input knot vector, the main idea is following:

- compute an original knot vector for each point set;
- define a flexible interval for original knot vector;
- if there are knots included in input knot vector between flexible interval, then use the nearest knot from input knot vector; if there isn't knot between flexible interval, then add the original knot to input knot vector.

Total above relation, the surface skinning can be described as following:

- original cross-sectional curves $C_k^w(v)$ $k = 0, \cdots, S$ (p, q) are degrees in u, v directions and ε is the max error accepted
- get scattered data points on curves $C_k^w(v)$
- initialize candidate knot vector $\hat{V}$ NULL, and repeat dealing with s+1 curves as following:
- fit curves using candidate knot vector $\hat{V}$ and the limited error ε ;
- add new knots to candidate knot vector $\hat{V}$.
- do s+1 curves compatibility, get knot vector V, the max index of control points is $\hat{m}$;
- initialize the knot vector NULL
  from j=0 to $\hat{m}$ repeat
  from i=0 to k repeat
  give R_j with the jth control point of $C_i(v)$
  end i
  fit curve $C_j(u)$ according to R_j and the candidate knot vector
  add new knots to $\hat{u}$
  end j
- deal with $C_j(u)$ j=0$\cdots$ $\hat{m}$ compatibility ,the knot vector merged is V, the max index of control points is $\hat{n}$;
- give the value of surface control points $P_{i,j}$ i=0$\cdots$ $\hat{n}$, j=0$\cdots$ $\hat{m}$ with the curves control points $C_j(u)$ j=0$\cdots$ $\hat{m}$;
- construct surface according to control points and the knot vector

IV. SURFACE DISPLAY

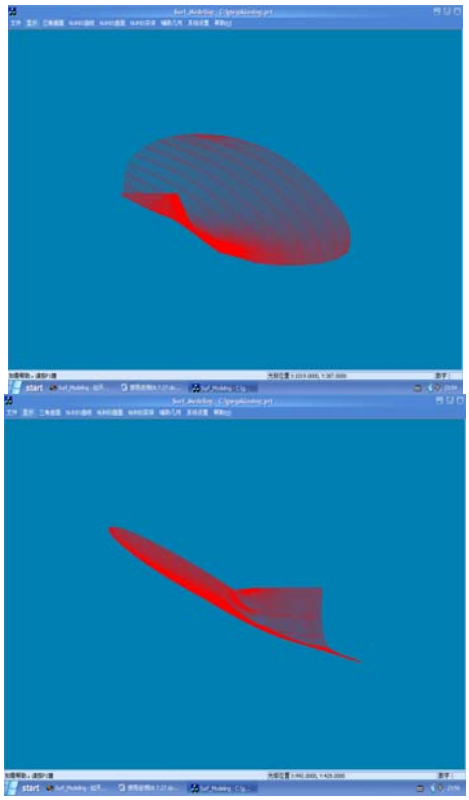


Figure 1 the surface of skinning reconstruction

Surface display use OpenGL functions to acquire the display effect of material, lighting, illumination, and veins. Using deBoor algorithm[1] to compute points data of surface, the process can be described as following: define parameters（u,v）of surface, compute points data of curves based on $\hat{m}$ +1 control points in v direction, get $\hat{m}$ +1 points as midst points to construct midist polygon, then in u direction compute the points data based on deBoor algorithm using parameter u, at last get the points data of surface P（u, v）. Fig 1 display the skinning surface using the above algorithm.

V. CONCLUSION

Firstly the above surface skinning process get scattered point data from cross-sectional curves, then fit curves based on these point data, skinning reconstruct these curves, lastly target surface is composed. the parameterizations of skinning surface will not be effected by different distribution of each cross-sectional curve, through flexible selection of knot vector, reduce the number of control points, The independent fitting of each curves reduce these phenomena that surface shape loses the true and calculation is unsteady. In practical system of surface modeling by using VC++6 .0 and OpenGL graphics functions implement above algorithm, assisting pickup point data of surface, interactive 3D treatment and lighting, the result shows that it is an effective surface fitting method.

REFERENCES

[1] Xinxiong Zhu.. Modeling technology for free curve and surface[M]. Beijing: Science press, 2000. (in Chinese)

[2] LA Piegl, W Tiller. Parameterizations for surface fitting in reverse engineering[J]. Computer-Aided Design, 2001, 33(3):593-603.

[3] Xujing Yang, Guangyong Sun, Qing Li, A New NURBS Tool Path Generation Algorithm for Precise Sculptured Surface Machining[J] Advanced Materials Research , 2010, (97-101) : 2477-2480

[4] Xianbing Liu, Fahad Ahmad, Kazuo Yamazaki and Masahiko Mori.Adaptive interpolation scheme for NURBS curves with the integration of machining dynamics[J].International Journal of Machine Tools and Manufacture, 2005, 45(4-5):433-444.

[5] LA Piegl, W Tiller. Biarc approximation of NURBS curves[J]. Computer-aided Design, 2002, 34(2): 807- 814.

[6] T.J.R. Hughes, J.A. Cottrell, Y. Bazilevs. Isogeometric analysis: CAD, finite elements, NURBS, exact geometry and mesh refinement[J]. Computer Methods in Applied Mechanics and Engineering, 2005, 194(39-41): 4135-4195

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 376-378

The Application of OptiSystem in Optical Fiber Communication Experiments

Xiang Yang[1], Yang Hechao[2]
[1]College of Computer Science&Technology Henan Polytechnic University Jiaozuo,China
Email:xiangyang@hpu.edu.cn
[2]College of Computer Science&Technology Henan Polytechnic University Jiaozuo,China
Email:yanghechao@hpu.edu.cn

***Abstract*- The basic components in Optisystem are introduced in this paper. In order to overcome the traditional shortcomings of the experiment in optical fiber communications, We use the Optisystem software to design the fiber-optic communications system and the simulation results are presented, which can not only enhance the understanding of each component of the fiber-optic communications system and its the function and provide guidance in real experimental design for the students, but also lay a solid foundation for the fiber-optic communication systems research in the future.**

***Index Terms* -Optical Fiber Communication;Simulation; OptiSystem**

I. INTRODUCTION

Optical fiber communication technology stood out from the optical communication and has become one of the main pillars of modern communications. It plays an important role in modern telecommunications networks. Optical fiber communication as a new technology, in recent years, its rapid development and the broad range of application, are rare in the history of communications. It becomes the denotation of the new technological revolution in the world. As a main transmission of various information tools, it is of great importance in the future information society. Now, optical communication systems are becoming increasingly complex [1]. These systems often include multiple signal channels, different topology structure, nonlinear devices and non-Gaussian noise sources [2], which make their design and analysis quite complex and require high-intensity work. Optisystem will allow the design and analysis of these systems become quickly and efficiently.

The traditional optical fiber communication experiments are usually conducted in the experimental box. The various components of optical devices in these boxes are encapsulated in comparison. So in the experiments, students often only do their work in accordance with the instructions on the experimental procedure step by step. It is difficult to understand the various parts of optical fiber communication system functions for the students and therefore, they lack the ability to create designs and fail to reach the effect required in the classroom instruction. When the OptiSystem software is introduced to the teaching of the experiments, it not only help the students to have a deep understanding of all parts of the optical fiber communication systems, but also have a clear visual impression on the optical fiber communication characteristics of the various components, which can give full play to its innovative design capabilities.

II. OPTISYSTEM

OptiSystem is an innovative optical communication system simulation package which was explored by optiwave company in order to meet the academic requirement of the system designers, optical communications engineers, researchers. It integrates design, test and optimize all types of broadband optical network physical layer functions such as virtual optical connection. From the long-distance communication systems to LANS and MANS, it can be well used. It has a huge database of active and passive components, including power, wavelength, loss and other related parameters. Parameters allow the user to scan and optimization of device-specific technical parameters on the system performance. OptiSystem has powerful simulation environment and real components and systems of classification definitions. A fiber optic communication system model is based on the actual system-level simulator. Its performance can be attached to the device user interface library and can be completely expanded to become a widely used tool. OptiSystem meet the booming market to a strong photon and becomes a useful tool for optical system design requirements [3].

III. PILOT PROJECT AND ARRANGEMENTS

Here, we will simulate the relative basic optical fiber communication experiment using the basic OptiSystem models and then presents the simulation results.

A. Simulation of wavelength division multiplexing experiment

WDM (Wavelength Division Multiplexing, WDM) is an important progress in the development history of optical fiber communication technology. The basic principle of the WDM is that the light signals with different wavelengths is put together at first, and then coupled to fiber optic cable lines in the same fibers for transmission. At last the receiver separates the different wavelengths by signal processing, restores the original signal and sends them to different terminal [4]. Figure 1 is a schematic map of WDM systems [5].

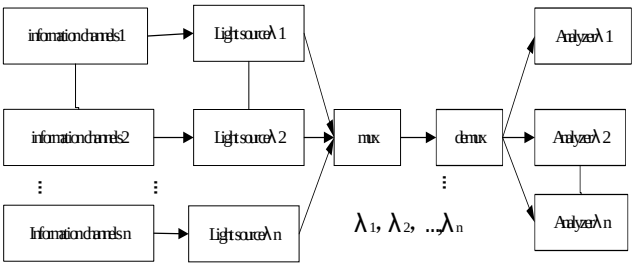


Figure 1 The schematic map of WDM systems

According to Figure 1, the related modules of the system are transferred in OptiSystem, then consists a system diagram, shown in Figure 2. The system consists of laser, wavelength division multiplexing, optical, demultiplexer and optical spectrum analyzers and other devices integral. We set the light source composed of four lasers with emission frequency 193.1THz, 193.4 THz, 193.7 THz, and 194.0 THz respectively. The light signals from the four lasers are put together through the WDM combine, and then coupled into the optical fiber. At last, the signal wavelength demultiplexer separates the combined signal in the terminal which is represented by optical spectrum analyzer. According to Figure 2, connect the system and run the simulation program, then the simulation results can be obtained.

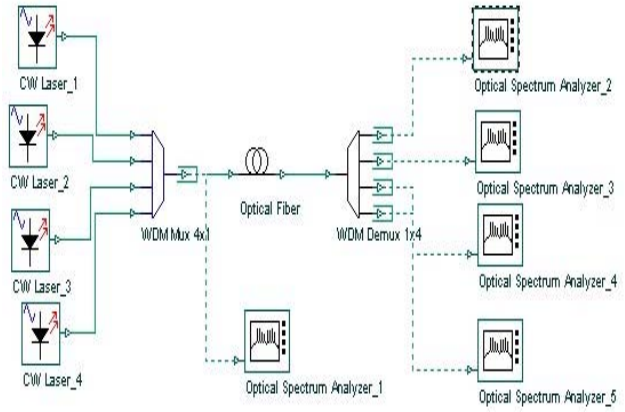


Figure 2 OptiSystem WDM system diagram

Figure 3 shows the frequency spectrum for the WDM signal after the combined. After demultiplexing, the frequency spectrum of each channel is shown in Figure 4.

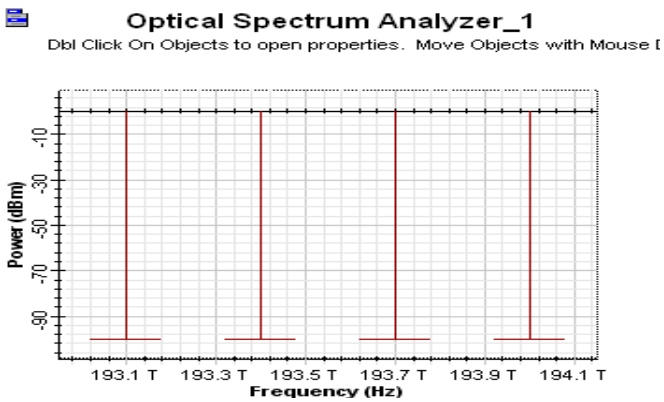


Figure 3 Multiplexed signal spectrum after the WDM mux

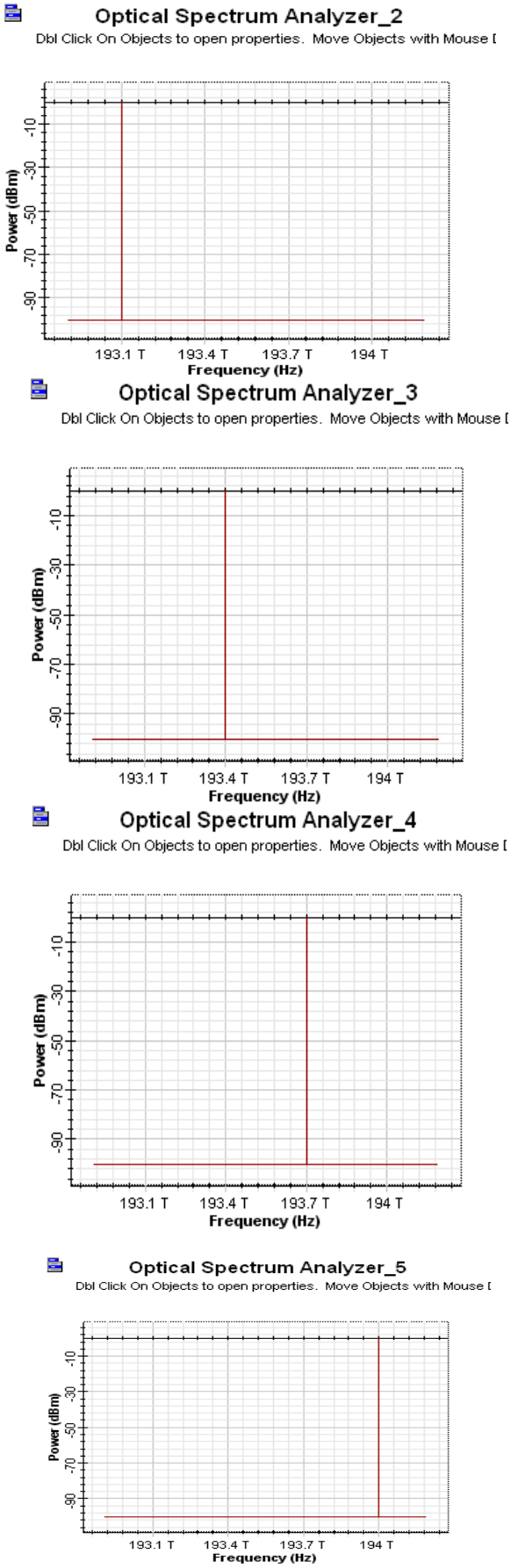


Figure 4 The channel spectrum analysis chart after demultiplexing.

The simulation results show that the system implements the basic functions of WDM systems. In the actual experiment, it can be realized by properly adjusting the parameters of each device to obtain better experimental results.

B. Simulation of optical fiber amplification experiment

The application of optical amplifiers in communication is a major breakthrough in the history of optical fiber communication technology. It replaces the traditional electronic relay station and make the dream of all-optical communication becomes a reality, in which erbium-doped laser amplifiers has the fastest development [6]. Figure 5 is an optical amplifier system based on EDFA, which is designed with Optisystem. The signal and pump light are combined together through the ideal MUX. Then they enter into the erbium-doped fiber amplifier. By comparing the spectrums of light changes before and after amplified, we can observe the amplification effect. Connect the system according to Figure 5 and run the simulation. The center wavelength of the signal light and pump light used here are 1550nm and 980 nm respectively [7]. The signal spectrums of the optical before and after amplified are shown in figure 6(a) and (b) respectively. From this figure, we can see that the intensity of the signal is significant enlarged. Moreover, optical signal spectrum before and after amplified has similar shape. This means that this system has achieved the purpose of optical amplification.

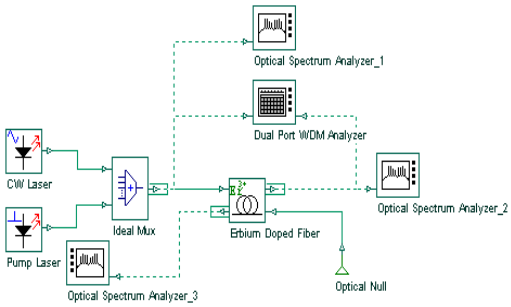
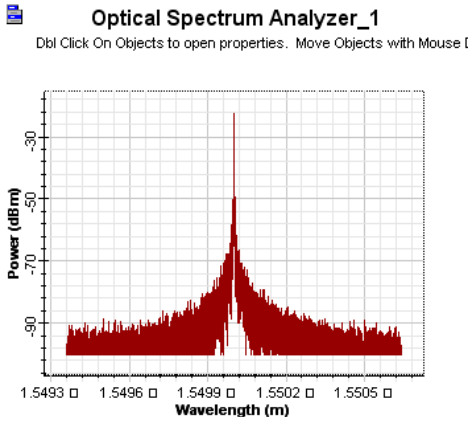


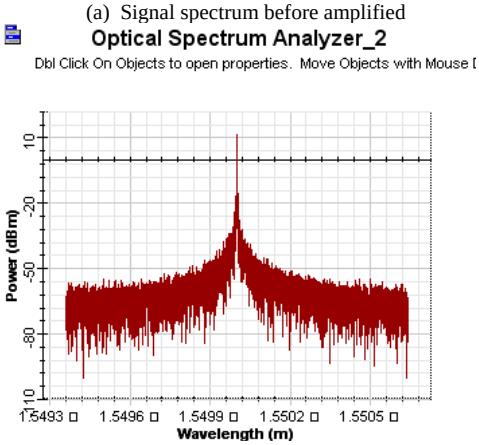
Figure 5 Optisystem optical amplification system chart

IV. CONCLUSION

Optisystem provides a flexible platform for virtual experiments which help students to grasp the more abstract principle of optical fiber communication systems. Using this software, it is beneficial to train the abilities of students, such as independent analysis, design and ability to solve practical problems. Moreover, it helps to enable students the ability of connecting theory with practice, finding some problems in the experiment, grasping the soul of theoretical knowledge. The results show that the actual teaching has greatly enhanced the students' interest and curiosity and lays a solid foundation for their future research work.



(a) Signal spectrum before amplified



(b) Signal spectrum after amplified

Figure 6 The signal spectrums of the optical before and after amplified.

REFERENCES

[1] Sun Qiang, Zhou Xu. "Optical fiber communication system and its applications," Beijing: The Press of Tsinghua University, 2004.

[2] Shi Shunxiang, Chen Guo Fu, Zhao Wei et al.. "Nonlinear optics," Xi'an: The Press of Xidian University, 2003.

[3] Automation net "OptiSystem software for the design of optical communication system ," http://www.zidonghua.net.cn.

[4] Liu Zengji, Zhou Yang Yi, Hu Liaolin. "The optic fiber communication," Xi'an: The Press of Xidian University, 2001.

[5] Zhang Baofu, Tan Xiao, Jiang Huijuan. "The principle and experiment lectures of optic fiber communication system," Beiing: The Press of Electronics Industry, 2004.

[6] Zhang Mingde, Sun Xiaohan. "The principle and system of optical communication," Nanjing: The Press of Southeast University, 1998.

[7] Wang Jingshan, Shen Xinjie, Sun Wei. "The optical fiber communication devices," Beijing: The Press of Defense Industry, 2003.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15,August 2010, pp. 379-381

Research on SOA-Based Heterogeneous Systems Access Performance

Shufen Liu[1,2], Yanyang Zeng[1], Chuanhong Huang [3], Peng Xu[3]
[1] Institute of Computer Science and Technology,Henan Polytechnic University, Jiaozuo, China
liusf@mail.jlu.edu.cn
[2] Institute of Computer Science and Technology,Jilin University, Jilin, China
zyyhost@126.com
[3] System Engineering Research Institute,Beijing, China
huangch0403@163.com,xupeng_2007@live.cn

***Abstract*—SOA entities heterogeneous systems integration, application system integration by connecting cross-platform, feature independent, reusable service. Mutual visits between heterogeneous systems are based on different platforms, different languages, different environments. In this paper, we research information sharing among heterogeneous systems, business component reuse, combine Web services communication technology, and for access to the system performance is analyzed, propose a reasonable deployment of multi-service center program, the program can improve system scalability, reliability, and access performance, initially showed a good value.**

***Index Terms*—SOA, access performance, Web services, heterogeneous systems**

I. INTRODUCTION

In today's rapid development of information technology, network technology is continuously improved, Web-based applications systems are introduced, most of these applications have a lot of data for processing. Centralized information processing frequently used in the early course of business, which does not show concern on sharing application logic and data between different applications. With the growth of the business and the development of information technology, more and more heterogeneous systems have appeared, leading to the problem that the information interaction and interoperability in systems are very difficult, and then leading to a number of urgent problems as "islands of information"[1].

SOA provides services for other applications by some mechanisms such as publishing、discovery and binding services etc . Combining SOA with web services technology has become a new direction in the field of computer information at present, which is a good solution to these heterogeneous systems communication and other issues. SOA has the advantages of loosely coupled, coarse-grained interoperability etc, the basic idea is services at the core, the heterogeneous system integration into a useable, standards based services, and so that it can be reassembled and used. [2]By using SOA architecture design ideas, you can minimize the coupling between systems and improve reusability.

Web services applications is growing rapidly in heterogeneous system, the number of Web services doubled. Heterogeneous systems based on SOA using techniques of web services composition when they mutual access , the existing Web services composed according to business process logic, which makes the composition service to provide more powerful and more complete functionality, enabling Web services reuse. If the communication between heterogeneous systems is very frequent, and the flow of data is relatively large, then the visit will be considerable pressure. This will be likely result in data loss, network congestion, etc. Moreover, once the server machine failure occurs, it will directly lead to paralysis of the whole structure of communication.[3]This article describes two common access programs, we analysis and comparison the programs access performance, put forward a more reasonable multi-service center communication program, the program fully to achieve loosely coupled SOA characteristics , make the whole system highly scalable, reliability and good access performance.

II. SOA AND WEB SERVICES

For now, Web Services is the most suitable technology set to achieve SOA, SOA is able to rapidly develop a large extent due to the maturity of Web Services standards and the popular of application, which provides the foundation for the realization of SOA architecture .SOA is a conceptual model, Web services is a framework structure defined by the protocol stack which constitute by a set of protocols. Which defines the communication between different systems programming framework for loosely coupled[4].

Web services as a good SOA implementation technology, enables service-oriented architecture has better features compared with the past architecture (such as C/S), highlighted in:

1) The distribution of the overall structure: application functional elements are deployed to multiple systems, in local or remote network. Web services can make full use of HTTP which the industry have been widely used on the Internet as the underlying transport protocol, going through the corporate firewall, to achieve the interaction across the enterprise boundaries;

2) Open standards: compared to traditional applications design pattern, loosely coupled system based on Web services is easily re-configured to achieve

the replacement of functional elements, interactive simple, has good scalability and flexibility;

3) Open standards: SOA using open standards is a critical success factor, the traditional distributed computing technologies such as CORBA, DCOM (Distributed Component Object Model), RMI (Remote Method Invocation) are to be based on the specific implementation, particular software provider, achieve very complicated, hinder the advance of the application;

4) Application process manageable: in the SOA, each separate service is designed to be business oriented functional elements, and it also as a business process or workflow component. A well-designed service through a clear description in its input and output, make other services know how to call on the services by understanding the description, particularly in the Web services architecture specifications using the WSDL description, which enables automated processing of calls.

Web services technology provides a realization of SOA platform to enable SOA has become the mainstream of IT. Early SOA model defined by Web Services formed standards, around three basic components of the architecture model: service requester, service provider, service registry.

II. Three Kinds Of Heterogeneous System Access Deployment Plan Analysis And Comparison

Heterogeneous systems mutual access is based on SOA architecture as middleware, Web Service as the means of communication. This access method used Web Service Access platform independence, hidden business components platform-related, and also achieved a great reuse of business components.

The following major research on how to achieve mutual visits between SOA-based heterogeneous systems by the proposed deployment program in this paper[5].

A. Peer-to-peer deployment program

Peer-to-peer deployment program is a routine deployment plan, this deployment program requires each heterogeneous systems endpoint deploys a Web Service. Any of a heterogeneous system access with other systems, mutual visits directly by web service in their own end as middleware, Figure 1 is a kind of peer-to-peer deployment program.

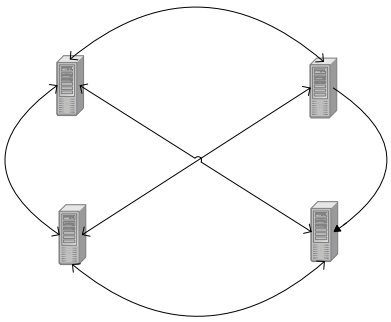


Figure 1. Peer-to-peer deployment program.

Obvious shortcomings of this peer-to-peer deployment program:

1) With the increase of heterogeneous systems, the connection between Web Service servers growing, each system, will access the total number of connections follows the following formula:

SUM= P_n^2 (SUM denotes the total number of connections, n denotes the number of heterogeneous systems).

According to this algorithm, each additional heterogeneous systems, it will increase 2n (n is the total number of heterogeneous systems before the addition of a heterogeneous system) connections.

2) These connections are not conducive to management, and bring troubles for heterogeneous system maintenance and management, and will bring enormous pressure on Web Service on the end of the heterogeneous systems when a system access to other systems frequently.

Therefore, peer-to-peer deployment program is not the best deployment of Web Service.

B. Central server deployment program

Central server deployment program used a central server as access transit station, as shown in Figure 2. Each heterogeneous system access to other heterogeneous

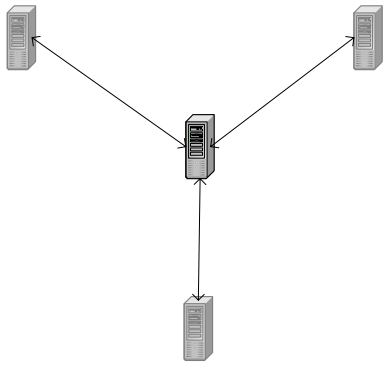


Figure 2. Central server deployment program.

systems must be based on a central server for the transit.

Compared with peer-to-peer deployment program, central server deployment program has obvious advantages:

1) Each distributed application system only has one connection with and the central server, the total number of connections as the following formula:

SUM=n(n denotes the number of distributed applications).

Using this approach, each add a distributed application system, just add one connection.

2) This connections all connect between central server and heterogeneous systems.

However, the seemingly perfect central server deployment program also has its defects which can not be ignored: this program is not considered of the pressure on central server, when the communication among heterogeneous distributed application systems more frequently, the data transmission ratio is large, will lead to access rate slowing, and may cause network congestion,

the poor performance defect determines that it is not the best deployment of Web Service program.

C. Multi_centers server access deployment

The prototype of the multi_centers server access deployment program comes from the center, deployment scheme, and improved based on the original program (Figure 3 is the multi_centers server access deployment program).

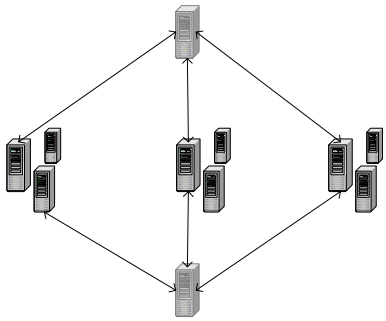


Figure 3. The multi_centers server access deployment.

1) In order to relieve the accessing pressure of server center, Web Service components on the Web Service component layer can be classified into different clusters according to some way. For example: application services, data Service etc. According to the actual needs, each class or several classes corresponds to a server.

2) Each center is a cluster server.

In multi_centers server access deployment program, the center server of the center, deployment program can expand to multi_centers server. Each server is equal on its position and composed of a group of servers. These servers' positions on single center server are equal, and managed by the unified management tool.

The multi_centers server access deployment program compared with the former two access deployment program, the advantages as follows:

1) Scalability: The number of the center server can be expanded according to one's own need or the actual need.

2) Management Convenience: Comparing to the peer-to-peer deployment program, through classifying the components on the web services layers, The advantage of access connectivity of distributed applications system is convenient management between distributed application system and the center server.

3) Efficiency: compared to the access, deployment program, the deployment according to the classification of the web services components greatly reduce the accessing pressure of a single center server. And the same time, the accessing pressure of center sever further improved by load balancing on each server [6].

4) Reliability: Each server is composed of multiple servers, if one server is failed, which will not affect the normal operation of server center, and will also not affect the normal operation of lightweight SOA framework [7].

III. CONCLUSION

Multi-center server deployment program simultaneously uses two strategies, Web Service classification and load balancing, to reduce the pressure on a single server, and improves the reliability of SOA architecture. [8]The program can effectively improve the system scalability, reliability, and access performance, is a good framework for enterprise application integration, and shows a good value initially in enterprise heterogeneous systems.

REFERENCES

[1] Alonso G,Casati F.Web services and service-oriented architectures [C]. Proceedings of the 21st International Conference on Data Engineering, 2005.1147-1148.

[2] Michael Stevens.The benefits of a service-oriented architecture[J].EAI Journal, 2002,4:20-22.

[3] StalM. Using Architectural Patterns and Blueprints for Service Oriented Architecture. Software, IEEE, 2006, 23 (2):54-61.

[4] Benatallab B, Dumas M, Fauvet M C. Overview of Some Patterns for Architecting and Managing Composite Web Services[J]. ACM SIGecom Exchange, 2002, 3(3): 9-16.

[5] Paul Patrick. Impact of SOA on enterprise information architectures.Proceedings of the 2005 ACM SIGMOD international.

[6] Wang Ende. Developing novel enterprise management information systems based on SOA[J] . Journal of Jilin University:Information Science Edition ,2006 ,24 (3) :322-329.

[7] Liu Xinfa , Ye Xiaojun. Construct SOA with XMLBeans and Web service technologies[J] . Computer Engineering and Design ,2007 ,28 (6):1320-1323.

[8] Herzum P. Web services and service-oriented architectures [J].Cutter Distributed Enterprise Architecture Advisory Service Executive Report,2002,4(10):35-63.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15, August 2010, pp. 382-385

Research of Web Data Mining Based on XML

LiFen Gu[1], JunXia Meng[2]
[1]Department of Computer and Information Engineering JiaoZuo Teachers College, JiaoZuo , China
Email: jzgulifen@163.com
[2]College of Information Engineering JiaoZuo University, JiaoZuo , China
Email:mjx_2005@126.com

Abstract—With the broad use of internet,Web data mining has gradually become the focus of current research on data mining. At the same time, XML technology is becoming the actual standard for data-organizing and data-exchanging by and by in the new-generation intemet.The integration of this two kinds of technology,Web data mining based on XML has become an important task of Web data mining.This paper designs a XML-based web data mining model, explains the process of HTML documents transformed to XML documents and analyzes the key technology in the process, and utilizes traditional data mining methods to complete Web data mining through XML.

Index Terms—XML, Web data mining, semi-structured

I. INTRODUCTION

Along with the rapid development of the Internet, more and more database and information systems join into the network, network exists large amounts of data. Facing so many complicated Web space, How to find the required information that have become an important problem in the vast network. Although users can rely on search engines quickly, efficiently and accurately to find the related information, to find the user need information is still very difficult. Recent years, web data mining based on XML provides an effective method to solve this problem.

II. XML AND WEB DATA MINING

A. XML

The Extensible Markup Language (XML)[1] is released by the World Wide Web Consortium (W3C) in Feb,1998. XML's purpose is to define a data-exchange standard via the Internet, to meet the requirements of increasing network application, and to ensure the good reliability and interoperability on interacting via the Internet.

Much of the information appeared on the current Web in HyperText Markup Language HTML document, Users through the browser to obtain information of these HTML document. HTML document may be written by manual or using HTML tool. Because the HTML document does not aim to automatically extract, but for expressing the information content. Therefore many of the HTML document on the Web is not standardized format, and extracting data is more difficult from the unstandard document than the structured document.

XML overcomed the shortcomings of HTML, standardized the documents on the Internet, gave mark a certain meaning, and reserved the advantages of the HTML- concise, suitable for transmission and browsing. XML Set the advantages of SGML and HTML in a whole, and become the core of the next generation of the Internet. XML have the advantages of scalability, structural, platform independence, self-describing, flexibility, standardability and simplicity.

B. Web Data Mining

Data mining [2] is a data- extracting process, which extracts the unknown implied and user interested information from the large amounts of data. As the development of Internet, A large amount of information is obtained from the Web, so Web data mining becomes a new research content. Relative to the data of the Web, the data structure in traditional database is very strong, but the data on the Web is most characteristic semi-structured, therefore the Web data mining is much more complicated than the data warehouse data mining.

1) Heterogeneous database environment

From the perspective of database-research, the information on the Web site can also be considered a more larger and complex database. Each site on the Web is a data source which is heterogeneous. Thus the information and organization are not the same between each site, finally to be a huge heterogeneous database environment. If you want to use these data for data mining, we should firstly study the integration of heterogeneous data between sites. Only integrate the data of the sites and provide a unified view of users, we may get the needed from the huge data resources. Secondly, we need to solve the problem of data query on Web. Because, If the required data cannot be obtained effectively, we will not do the data analysising, integrating, and processing.

2) Semi-structured data structure

The data on the Web is different from with the traditional database data, the traditional database has a certain data model which can describe the specific data according to a model. While the data on the Web is very complicated, it has not the specific model, and the data of each site, with reports and dynamic variability, is designed independently. Therefore, the data on the Web has a certain structure. But for the readme levels, it becomes a kind of completely structured data-Semi-structured data. Semi-structured structure [3]is the biggest characteristic of the Web data.

3) To solve the problems of semi-structured data

Web data mining technology firstly solve the query and integration of the semi-structured data source model

and semi-structured data model. To solve the query and integration of the heterogeneous data on the Web, we must have a model to clear the data on the Web. For the semi-structural characteristics, we need to find the key point that is a semi-structured data model. Not only to define a semi-structured data model, also we need a kind of extracting technology. That is to say, it can automatically extract semi-structured model from the existing data. The semi-structural model and the semi-structured data model extracting technology must be the prerequisite of the Web data mining.

C. Web data mining methods

Comparing with the traditional data and the data warehouse, the information on the Web is unstructured, semi-structured, dynamic and easily confused. So it is difficult to take the data mining directly from the data on the Web pages without the necessary data processing. The typical Web mining process[4] as follows:

4) Find resources: The task is to obtain data from the target Web document, it is sometimes not only limited information resources online, including Web document email, files, newsgroups, or Web log data and even through the Web form of trade in the database.

5) Information selecting and preprocessing: The task is to eliminate the useless information and conduct the necessary information from the obtained Web resources. For example, automatically removing advertising links from Web document, excess format markers, identifing paragraphs or fields, and making the data into a neat logical form or a relation table.

6) Patten discovery: Automatically taking the patten discovery, and it may happen in one same site or among multiple sites.

7) Mode analysis: Validating and explaining the mode of the previous Step. It can accomplished by the machine automatically, or by interacting with the analysts.

III. WEB DATA MINING SYSTEM MODEL BASED ON XML

According to the general flow of Web data mining and the related XML technology, this paper designed a Web data mining system model which based on XML. The basic idea of this model is to obtain the target web page and turn it to XML document for storage, and then using different data mining algorithm for an XML document data mining according to the user's interest in knowledge. Web data mining model consists of three logic levels. As shown in figure 1.

Data access layer is to extract and convert the semi-structured Web data, use structured data for representation, build the Multi-level Web database, and preprocess the Web server log data forming the Web log database. We call the Multi-level Web database and the Web log database as the Web database. Data mining layer is the key to realize the system function, using a variety of data mining algorithm and flexible and open form task for the final aim is to provide the effective Web mining solution, and finish all kinds of data mining task.

IV. KEY TECHNOLOGY OF THE MODEL

A. Convert Web documents into the well-structured XML format

The basic idea is to convert the obtained Web page in HTML format into the well-structured XML documents. The main steps:

(1)Through the method of artificial input, we offer the query theme and find out some of the Web pages that accords with a condition, thus these Web sites are the data sources.

(2)Using the Tidy tool[5] to convert the data, eliminate some useless advertising message, filter lots of irrelevant markers from the HTML document, correct the common errors, and generate the good-format equivalent XHTML document.

(3)Finding the reference point of the data in the XHTML document, and using XPath or XSL technology to identify the reference point and extract data. Finally, using the XML documents to save these data, after many data extracting and incorporating these XML file to the external files system for storage.

The process is shown in figure 2.

Following is a pure Java implementation of the HTML-XML converter part of the source code:

```
package org.w3c.tidy;
public class HtmlToXml
{public static void main(String[ ]argv)
    {…
     String file;
     InputStream in;
     String prog="Tidy" :
     Node document;
     Out out=new OutImpl();
      /*normal output stream */
     int argc=argv.1ength+1;
     int argIndex=0;
     Tidy tidy;
     Configuration configuration;
     String arg;
     tidy=new Tidy();
     configuration=tidy.getConfiguration();
    / * read command 1ine *  /
    while(argc>0)
    {
        if(argc>1&& argv[argIndex].startsWith("-"))
        {
            arg=argv[argIndex].  substring(1);
            if(arg.1ength()>0 && arg.charAt(0)== '-')
                arg=arg.substring(1);
            if(arg.equals("asxml") ‖ arg.
equals("asxhtml"))
                configuration.xHTML= true;
                --argc;
                ++arglndex;
                continue;
      configuration.adjust();    / * ensure config is
se1f-consistent */
```

```
                    ......
      / * Internal routine that actually does the parsing.*/
      / * The caller can pass either an InputStream or file
name.*/
          document=tidy.parse(null,file,System.out);
          totalwarnings+= tidy.parseWarnings;
          totalerrors+=tidy.parseErrors;
     ……
     }
  }
```

B. Data extracting

Regardless of the Web page or the XHTML document generated in the previous process, most of them are irrelevant to extract information. Therefore we need to find a specific area in the previous generated XHTML document, in order to extract the need data and avoid those irrelevant information for data mining. The information extraction generally adopt two methods: One way is by absolute path; Another is by anchor point data extraction[6]. Due to the HTML page might change, although absolute path can position in an XML document sections and document components, but in the pages of the position changes, it is more easy to make a mistake. Below is an absolute path list:

/html/body/center/table[5]/tr[2]/td[2]

Based on this situation, we can realize by using an independent of absolute path extraction method—Finding the anchor point which contains in the extracted information. Normally, anchor point is based on the information content of existence, and it has nothing to do with the HTML path. If you want to search a list of "High" text, the absolute path can be written as:

//table[starts-with(normalize-space(.),"High")]

After getting the anchor point, we can create the actual extracting data code in the XSL file. XSL specifies how to search for data from the anchor point, and to use the needed format to form an XML output file. The searched data generally contains in the same < table > element, in this table it also generally includes the keyword of the required information, Eg: the customer purchase records in transaction database. Firstly, we need to the contents of a text node which is the key element node <td>, thus we can find its grandfather node. Because its grandfather nodes may well be the < table > , so we take < td > grandfather node as anchor points. Taking all content which under the grandfather node to loading in an XML document. If only a data extraction, according to the above steps it has been completed. However, the Web data mining is a cycle process, it needs multiple data extracting to form the data source of the data mining, and the result will be combined into XML data files for store.

The following is part of program implementation code:

```
Public static void main(string args[ ]){
try {
 Document
xhtml=XMLHelper.ParseXMLFromURLString("file://wp
.xml");
 Document
xsl=XMLHelpr.ParseXMLFromURLString("file://xsl/
wp.xsl");
// For a given xhtml document on the map on the xsl
transform
 Document
xml=XMLHelper.transformXML(xhtml,xsl);
XMLHelper.outPutXMLToFile("XML"+File.separator
+result.xml");
    }catch(XMLHelperException xmle){
  ......
      }
    }
```

The resulting xml document format as follows:

```
 <?xml version="1.0" encoding="utf-8"?>
// Statement
    <table>  // To anchor the table
        <tr id="1">  // Table row
            <td> Table of contents </td>   // The
contents of the row in column
                    ……
        </tr>
        ……
     </table>
```

C. XML data source in data mining

XML provides a DOM (Document Object Model) and SAX (Simple Application for XML)[7] which are two data access interface. External applications through both interfaces are very easy access to XML documents.

DOM interface is developed by the W3C. This interface defines a series of XML documents used to achieve access and modify data objects, and XML documents into the document tree structure. The object tree is the relationship between elements within XML documents reflect, through them, you can access and modify XML documents of all the data. Applications are also available in this XML document tree structure hierarchical data access. For document information, such as data, the significance of the data and data relationships, the DOM interface can convert it to a tree node or nodes in between. Since all of the XML document tree structure information can be included, which makes random access XML document data has become very convenient. Currently, DOM standard is set at two levels Including DOM Level 1 and DOM Level 2.

DOM interface is a comprehensive analysis to XML, It is necessary to complete all the XML DOM tree into memory, the random access speed is very fast. However, when more complex and huge XML document, you need to take up more memory space and the speed of access to the DOM tree will be greatly reduced.

SAX interface can avoid the defects of the DOM interface, it does not require all the document into memory, and data files using a time-driven sequential access. In SAX interface, When the XML parser encounters a particular event, it will call the appropriate function to handle the event. Of course, SAX interface, only just calls the corresponding function, as for the specific data processing is through execution of the function to complete. SAX is not a W3C standard. It is a

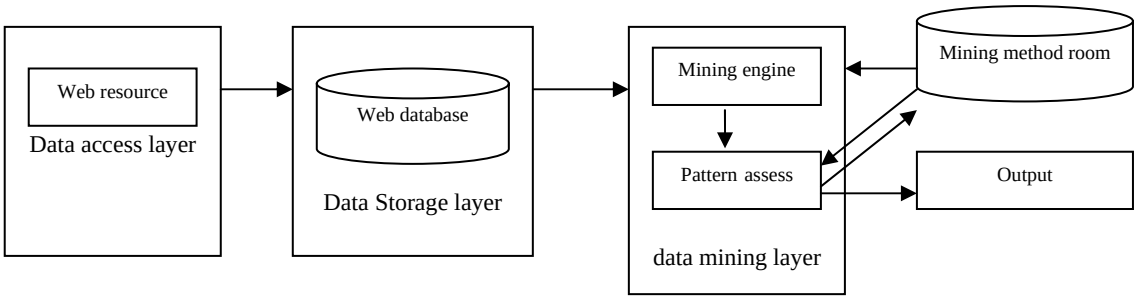


Figure 1. Web data mining model

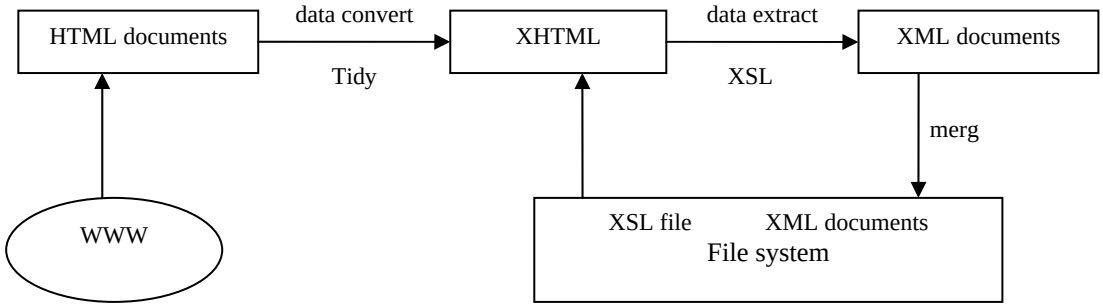


Figure 2. XML data converting model

group from the Internet, people interested in XML technologies out joint research. As an event-based XML programming interface, SAX has been widely recognized by various XML groups.

Microsoft .NET Framework provides support for XML DOM object model, this support is through a series of related classes to achieve. For SAX interface, NET also has a corresponding analog implementation. In the XML data source, provided by means of XML DOM interface, or SAX interface, using traditional data mining Web data mining can be carried out to obtain useful knowledge, Form a knowledge base.

We have used Microsoft VS .NET development tools, implements a number of agricultural supply and demand information of agricultural products on the site of Web data mining. Practice shows that the proposed use of XML and the efficiency of semi-structured data model for Web data mining auxiliary method are effective.

V. CONCLUSION

Web data mining is a new research field, It is different from the traditional data mining. Due to the Web data is an unstructured data, it makes the data mining become very difficult. While XML solves this problem very well. Because it is also a kind of unstructured data model, its appearance makes the data mining based on Web get greatly simplified. This paper ,on the basis of both, offers a Web data mining model which based on XML. The model according to the general flow of Web data mining puts a variety of data mining algorithm and other mining system modules together, and completes the Web data mining. This paper emphatically expounds the key technology how to transform the HTML pages document into XML document, and utilizes traditional data mining methods to complete Web data mining through XML. It has a far-reaching significance in further Web data mining, information retrieval and knowledge discovery.

REFERENCES

[1] Man Li Chun ,Zhu Hong ,Research and discussion of Web Data mining [J].southwest national university journals .2005,31(2):302-306

[2] Femando B A，Miguel G，Rita E．Toward the Design Quality Evaluation of Object-oriented Software Systems[C]．Proceedings of 5th Intemational Conference on Software Quality．1995

[3] Ricardo Baeza-Yates Berthier Ribeiro-Neto．Modera Information Retrieva1．ACM Press．1999

[4] HAN Jing，ZHANG Hong jiang，CAI Qing sheng.Predietion for visiting Path on WEB Journal of software.2002.6:1041-1043

[5] Formatting HTML page of small tools http://www.love0452.com/thread-87070-1-1.html

[6] Jacky w.w.wan ,Gillian Dobble. Mining Association Rules from XML Data Using XQuery. Proeeedings of the second workshop on Australasian information security Data Mining and Web Intelligenee,and Software Internationlisation，NewZealand，2004

[7] Zhang Jian Xi,Wang Hong Guo,Zhao Pei Ying . XML application in WEB data mining technology [J]. Information technology and informatization,2005, Article 5

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15, August 2010, pp. 386-388

Analysis of Personalized information of Library Service Model based on Web2.0

Liu Zhong
Library of Henan Polytechnic University, Henan Polytechnic University, Jiaozuo Henan
E-mail: liuzhong@hpu.edu.cn

***Abstract*—Personalized Information Service is the product of development of network information environment, it is also the direction of Library Information Service Development. This paper describes the concept and the typical Web2.0 technology, Proposed the content of Personalized Library Services in the Web2.0 environment, combined with the Web2.0 related technologies. Summarized personalized information service model Based on RSS, Blog, Folksonomy, Wiki's personalized information service.**

***Index Terms*—Web2.0, Library, Personalized Information Service, Blong Wiki Folksonomy**

I. INTRODUCTION

With the promotion of new technologies of the internet and new applications of existing technology, Internet has grown from Web1.0 Times into Web2.0 Times, Web2.0 is not simply technology or Problem-solving program, It is a set of executable concept system, Practiced the ideal of the network socialization and personalization. The development of library has close relation with the latest computer and information technology. Usually some new computer information technology just appear then will applied in library management and services. Introducing the new ideas and new technology of Web2.0 in the library information service, launched Personalized information service is an effective way to solve difficulties of Multifarious network information and Screening difficult problem, Can provide better service for customers and meet the user's individual demand, Realize the maximum social values of the library information resources.

II. OVERVIEW OF THE WEB2.0

Web2.0 is a new kind of Internet application relative to Web1.0(the Internet mode before 2003),the appearance of it is the revolution for the Internet from the core content to external application, represent the Internet development theory system, there is no clear-cut distinction about the definition of Web2.0 now, The concept is put forward by The famous O 'DaleDougherty Reilly company and MediaLive company's Craig Cline on "brainstorming" session In March 2004, Currently mainly from the following two aspects defined Web2.0[1]: First, the definition from concept. In the era of Web1.0, Large Web portals hold the right to speak. They decided what the people see and hear. While Web2.0 is pay attention to the power of grassroots. Here the grass-roots is not referred to some of the concentrated individuals, But some of the individual which scattered among various groups. Web2. 0 in concept bring to people the transmission is a kind of freedom, equality, and open information exchange. Second, the definition from the core technology. It is based on the basis of the front thoughts. In order to realize this idea caused the technical application.

In a word, more accepted industry definition of Web2.0 is Blogger Doon mentioned in his Web2.0 Doon &to concept "interpretation" "call it Flickr, Craigslist, Web2.0 Linkedin, Tribes, Ryze, Friendster, Del icio. 43Things.com, us, etc.Take Blog、TAG、SNS、RSS、Wiki as the core of social software application. according to the new theory and technology of Six degrees of separation ,xml、ajax realize new generation mode of the internet.

III. TYPICAL TECHNOLOGY AND CHARACTERISTICS OF WEB2.0

A. RSS(Simple information polymerization)

RSS is the English initials acronym of Rich Site Summary or Really Smiple Syndication,in Chinese call it "summary information polymerization".RSS technology is mainly based on XML standards, the content which widely used in Internet is packaging and delivery agreements. In essence, RSS technology is a kind of Information aggregation technology, applied to various news reports, the service push, number or other data inquiry,etc. Currently, is one of the most application of library technology. RSS service could actively push the latest information to readers directly. Make readers can directly updated content without visiting the website, Thereby reducing the cost of retrieval time.

B. Blog

Blog's full name should be Web Log, Later, abbreviation as Blog. In essence, the blog is personal diary, Personal homepage or personal website. Especially on the technical level. Blog site is an easy-to-use web site, You can quickly issued ideas, communicate with others and engaged in other activities. All these is free of charge. Blog can be used as a kind of means announcement of a library information and communicate with readers. Library web site can provide blog space for readers. As secondary function of "book club" or "My Library", Help form the Readers' community.

C. Wiki (Wikipedia)

Wiki is a kind of more cooperation hypertext system writing tools, Wiki site can have many maintenance, Everyone can express their own opinions, perhaps discuss or expand the common theme. Facing the collaborating communities in writing, Wiki support the collaborative writing when facing community, also included a group of

auxiliary tool which support this writing. Could Browse, create and change the Wiki text based on the Web, and the cost of creation, alteration, release is smaller than HTML text, While Wiki system also supports collaborative writing which facing communities, to provide necessary assistance for collaborative writing, finally, Wiki writers naturally constitute a community, Wiki system provide simple communication tools for this community. Compared with other hypertext systems, convenient and open is the characteristics of Wiki, so the Wiki system can help us to share a field of knowledge in a community.

D. Tag

Tag is a more flexible and interesting classification, you can add one or more Tags for each log, each post or each picture,etc. You can see the website content which use the same tags. Thus produced more contact with others, tag reflects the power of groups, enhanced the correlation and the content of the interaction between the users.

The above four are the typical application technology for Web2. 0. In addition, there are other Web2.0 technology also widely used, Such as MSN (social network software), P2P (peer-to-peer network), IM (instant messaging) ,etc.

IV. THE CONNOTATION OF PERSONALIZED SERVICE IN WEB2.0 LIBRARY

Personalized information service is based on the information of the user's information usage behavior, habits, preferences, characteristics and the user's features to meet their individual needs the information content and information systems functions. The personalized information service is not come accidentally, it is the product of the development of network information environment, is also the development trend of information services and the development direction of the library information services.

In past, the library service model is mainly centers on librarian and the information service, the only purpose of the work is beneficial to the librarian who carrying out service work, but rarely consider the active participation of information users. The user is always accept service passively, their information needs is not fully reflect during the service, so their needs would be very difficult to meet. For a long time, the library reader's personality rarely considered in service. The librarian think that readers should adapt to the service provided by the library, therefore, a variety of standards which all libraries provide are the same, the services for every reader are also identical. The service did not cohere with the diversification of the reader and personalized requirements, so their information needs can't satisfy [3].

Currently, a variety of techniques and concepts of Web2.0 become accepted and adopted by more libraries. The Web2.0 emphasizes interaction with readers; enhance the experience of users, directly on the user's requirements. So the user has more spoken right, in the environment, the information can read, writable and interoperable. Internet users have a complete self-information initiative, the production of individual producers of information, but also manage information, and information users to interact. They change from passive acceptor to the owner of the network. in The past, "I provide what, the user acceptance what" was traditional library service mode, but now it is "what users need, what I offer.". The new generation of information users characterized by large number, demand complex, and difficult to analyze. The emergence of new forms and new features make the traditional personalized information services can not satisfy the information users' needs, and the library must face some new challenges.

Based on the above reasons, the traditional personalized information service can't meet the user's information behavior in the new environment. Therefore, the organization of the Library personalized information services must emancipate the mind, expand ideas, innovation and reform. The existing ideas and technical transformation of the Web2.0 must applied in the personalized information services, and continuing to explore Web2.0's personalized information services in the Library.

V. THE LIBRARY PERSONALIZED INFORMATION SERVICE MODEL IN WEB 2.0

A. Personal information push service

The information push service is a kind of new service which appears based on the push technological development, through the RSS technology, between library each stand may share the information content, the user may through a browsing window or read the software have the RSSFeeds information together when does not open the website page and forms own information gateway, but does not need to visit various websites one by one to obtain the website push information [4]. The latest news push service is the RSS technology most widespread domain in the library application, deliver the latest news in the hall or the network to users timely and high effectively, provides the personalized service and "the one-stop" work style for the users, raised the employ rate of the library resources. In the library retrieval system use the RSS technology, unifies the characteristic of the library service, carry on pushing and have custom-made the service towards the library resources information, the literature information, the special booklist, the special literature material, the conference information and so on., enables it to realize user's individual information to have custom-made and push service, promoted the information service level of the library enhancement[5].

B. Personalized interaction service of user

Blong as the typical application of the web2.0, caused the people to pay attention by its unique way. It develops fully page's content using the link, take the diary way, transmit real-time information through the network, was considered the fourth brand-new network exchange way after Email, BBS, ICQ. The Blog application is very widespread, in addition the simple edition, issue and maintenance mode, cause the application of Blog in the library personalization information service to be easy and feasible, one of Blog most remarkable characteristics has good interactive. Information alternately service based on the Blog , Moreover, take the Blog system which obtained free as the information interactive platform. The librarian may issue any information (including writing, picture, audio frequency and video frequency) momentarily. The

user will receive the library renew information through the RSS initiative push, and may register the library website to express the view with regard to the interested information, will carry on the discussion with the librarian and other users, may also give the comment and suggestion to the library service. Librarian understands user's idea through browsing user's commentary and message, then perfect library service. [6]Through Blog, may establishes a good interact exchange platform between the librarian and the reader, consequent, promotes the library personalization information service quality unceasingly.

C. Individuation information for classification qualitative services

The Folksonomy is constituted by Folk and Taxonomy and was proposed by the US senior Internet expert Thomas. It's a compound word and was derived from the most characteristically customized Tag function in social bookmarking service then applied widely in social software. Folk has the meaning of concourse in English while the Taxonomy is classification. It is always translated as Folksonomy, popular classification, free classification, grassroots classification and so on in domestic. The Folksonom, as typically used in Web2.0, do not need to be established, maintained and learned a large table of classifiable system in information organization. The category Tag is a private label of personal understanding based on content and can reveal the content of the information, all of which reflects the user's individuation. The methods, for example crossing the Folksonom, using the Tag tool, filter, rating commend and comment, are all aiming to let the library users participate, share and create cooperated and attract other users in the value-added content, at last create a strong effect of adsorption and the long tail effect. The applied of Folksonom improves the efficiency for users to search information, receive the information service.

D. Personalized information service based on the resource sharing

Resource sharing is one of the most important academic thought in modern information services area. Resource sharing is given a broader meaning and development under the premise of the massive number of information. As the 'same writing' for Wiki which is a network services for lots of people to write, unload and publish. We can use it to building knowledge network system which support the sharing of domain knowledge within a community. It can not only promote the use of existing resources, but also to add new resources for the library. Wiki can be used in the construction of special information base and thematic Database[8], you can also build more knowledge for the subject network system to support the sharing of domain knowledge in the disciplines and to provide bibliographic information for readers to comment on the platforms. Its application has provided a new model for the library's business activities and personalized information services.

VI. CONCLUSIONS

In a word, it is diverse for personalized information service model in the network environment of library. With the popularized of Internet and the variety expanding network functions, the new information service model will continue to emerge.

Only in the way of using new technological to innovate its information service model and provide users with more effective personalized information service, the library can keep up with the information age and obtain more developing space.

REFERENCES

[1] Fan Bingsi, Hu Xiaojing. "Library 2.0: Building the New Library Services". University Libraries. 2006, (1): 2-6.

[2] About RSS[EB/OL]. http: //www. haorss. com /about. asp.

[3] Liu Ningyu. "Based on the Web2.0 for readers management and services mode". Heilongjiang Chronicles. 2009.（21）53-54

[4] Kang Wei, Sun Chengyi. "The contribution of WEB2.0 technology for navigation library building in the academic resources". New Century Library, 2007, (2): 79-81,

[5] Fan Wei, Chen Shunian. "Based on the RSS of book information service concept and implementation". Modern Books Intelligence Technology. 2005, (12): 59-62.

[6] Chen Liangjin. "Base Blog to the library information interaction service research". Library Journal, 2007(2): 48-49

[7] Liu Chen. Folksonom "the Application Analysis Research of Library Network information Service". Modern Information, 2008(2):57-58

[8] Web2.0: Building the New Library. http: //www. ariad-neacuk/issue45/miller/

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15, August 2010, pp. 389-392

Research On Security Architecture MSIS For Defending Insider Threat

Hui Wang [1,2], Dongmei Han [1], and Shufen Liu[2]
[1] College of Computer Science and Technology, Henan Polytechnic University ,Jiaozuo, China
Email:wanghui_jsj@hpu.edu.cn,handongm520@163.com
[2] College of Computer Science and Technology, Jilin University ,Jilin, China
Email: liusf@jlu.edu.cn

***Abstract*—Network threat confronting organizations comes from not only outsider threat, but also insider threat. Nowadays, insider threat is widely recognized as an important issue of security management. However, tools and controls on how to fight against it are still in the research phase. Security architecture for defending insider threat is presented, which is composed of four parts: monitoring platform, secure authentication platform, information security platform and security management system. The first three parts of the architecture are to solve the problem from a technical viewpoint and the last is from a management point of view. It is simple and practicable to prevent and reduce insider threats by the combination of advanced security tools and good management system.**

***Keywords*-Internal Network; Insider Threat; Architecture; Security Management System**

I. INTRODUCTION

At present, the "insider threat" or "insider problem" has received considerable attention, and is cited as the most serious security problem in many studies. It has become a novel and hot research topic [1, 2, 3]. Classification statistics were conducted by American CSI / FBI according to the event source over the years. And the annual cost of losses is shown in Table 1[4]. Statistics show that: although most organizations are implementing effective strategies against external threat, the weakest link in organizational information systems security chain is insider threat. Insider threat is much greater than outsider threat in terms of the loss.

Table1
CSI / FBI annual loss cost survey according to event source

Year	System penetration/$	Insider abuse/$	Unauthorized insider access/$
2005	$841,400	$6,856,450	$31,233,100
2006	$758,000	$1,849,810	$10,617,000
2007	$6,875,000	$2,889,700	$1,042,700
Total	$8,474,400	$11,595,960	$42,892,800

In January 2008, at Societe Generale's second largest bank, a trusted and junior employee, Jerome Kerviel, perpetrated 72 billion worth of loss and fraud, through his knowledge of banking procedures, information systems and theft of coworker's passwords. Apart from Kerviel's actions, failure of control mechanisms leads to this fraud, undoubtedly the largest in the history of banking.

These two examples show that the most serious security breach and the most important economic damage are basically made by the insider threat from organizations. How to prevent and predict insider threat? This paper proposes a integrated and overall security architecture from the point of the combination of technology and management.

II. INSIDER THREAT

Trzeciak (2009) defines insider and insider threat as "An insider is a current or former employee, a contractor or a business partner who has or had authorized access and intentionally exceeded that access in a manner that negatively affected the confidentiality, integrity or availability of the organization's information or information systems'. Insider threat can be defined as the threat to information system security due to the intentional misuse of computer systems by users who are authorized to access those systems and networks [5].Due to the legitimacy and trust the insiders enjoy, this type of crime is difficult to detect and mitigate before the occurrence.

Previously, confidentiality of electronic documents concerned by many companies is focused on external personnel. Technical means, such as intrusion detection, firewall, information encryption, access control mechanisms, are to solve the problem of external protection. However, these controls and tools are designed to fight against outsider threat of organization network, and little progress has occurred in dealing with the insider threat, including insider attack and insider misuse. Because of the lack of knowledge about insider threat, organizations can not take appropriate preventive measures. These all cause the frequency of insider threats higher and higher. Whether intentional or accidental, insider threats will be one of the greatest threats to security. If the network security is unknown or not implemented, Internet users, in practical applications such as surfing unsafe websites, click on a malicious e-mail link, or not to encrypt sensitive data and forth, will continue to unwittingly play the role of safety bomb. As the mobility of business people is more and more, users use a large number of removable storage devices such as U disk, mobile hard drive, writable CD and MP3 players, network connection such as Bluetooth, as well as mobile devices such as laptop, PDA. Insider threat as an example

This research are supported by The Doctor Grant of Henan Polytechnic University(B2010-62).

of mobile devices is shown in figure 1. A serious threat of confidential data leakage to enterprise is posed. The survey of Ministry of Public Security exposed that the ratio attack or virus origin from internal staff increased by 21% over the previous year, and the ratio of involving external personnel decreased by 18%, which reveals most network unit concerned for external defensive considerations which led to the threat from insider rise at the same time. However, the fatal results are usually caused by insider threat.

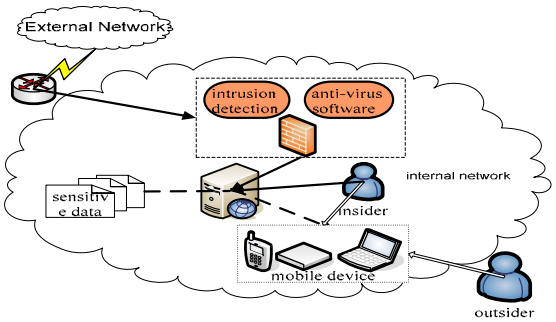


Figure1 Insider Threat--Mobile Devices

(including staff who aren't familiar with computer technology) because of network popularization and software development. Interface of these tools is humane and easy to understand. It is one of reasons that insider threats are mostly caused by internal staff. And internal users generally face database directly and operate directly on the server. Taking advantage of fast network, critical data are stolen or destroyed with ease. Users in the organization have different privileges; secret information lacks of effective control and supervision; it is difficult to manage the staff; system is vulnerable to be attacked by means of passwords and unauthorized operation. These factors cause insider threats increasing more and more.

III. ESTABLISHING INSIDER THREAT DEFENSE SYSTEM

Damage caused by insider threat is obvious. The goal of this paper is to extremely mitigate business damage posed by the insider misuse or the insider attack, endeavor to cease the insider threat initially, and reduce internal risk to a minimum.

In order to prevent internal threats, a relatively secure internal network needs not only advanced and effective security configuration, but also comprehensive management system and experienced security managers

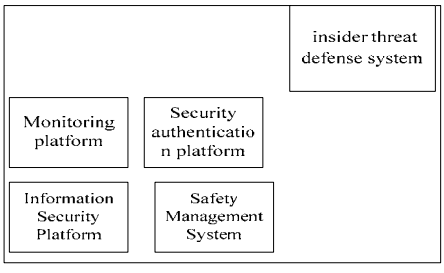


Figure 2 MSIS

[6]. In this paper, a integrated and overall security architecture for an effective internal defense has been proposed combining the results of current research and the concepts of technology and management. Three systems platform and a safety management system are included in the network. This architecture is called MSIS taking the first letter of each part of the composition. It can be shown in Figure 2.

A. Monitoring platform（ MP）

The architecture including a monitor platform MP has been proposed in order to make the internal users on the host and network effectively and prevent violations from internal and enhance their internal security. Organizations must monitor all critical information system activity like servers, software applications and other data resources, Access must be strictly controlled and any suspicious activity must be investigated. MP has a powerful logging system. As shown in literature [7], an improved surveillance method based on complex roles has been proposed in order to monitor the work activities of the users in organizations, applications and operating systems.

Currently, MP launched by software companies is generally composed of three parts: Client, server-side and management-side. Client is the agent installed on the computer software. It is used to collect host data and receive the security policies and directives configured by the administrator from the server-side. Its ultimate aim is to monitor the host behavior. Server-side is installed in a platform with the high performance. It is used to receive various kinds of information sent by the host client. And then the information can be managed and stored. Management-side is usually a web service or other applications. After users logging in, the corresponding management interface can be accessed by managers. Appropriate security policy is configured and issued. Client log can be inquired and analyzed. A variety of statistical information can be counted and managed.

The following functional areas should be included in a comprehensive network of MP: firstly, desktop management and control of host behavioral; secondly, internet behavior management and breaking of illegal host access; thirdly, security management of terminal equipment and storage media; fourthly, remote installation of system patches distribution and software; what's more, monitoring and safety assessment of the host system performance; in the end, monitoring of network equipment.

Although there are many monitoring products in the market and their functions are different. All the questions can not be completely solved. This article points out that scientific management mechanism in internal network and the fast upgrade of system must be included in a perfect MP. And security policy in off-host must be supported and excellent compatibility and multiple security mechanisms must be contained in system deployment.

B. Security authentication platform （SAP）

This paper presents that SAP performs a variety of authentication methods to achieve secure login and

authentication of users。 It is independent from the landing system of the original computer, and has higher security and reliability. It is made of the authentication server, authentication agent and authentication tokens. Authentication server is the authentication engine of the network, which is managed by the security administrator or network administrator. It is mainly used for token issue, the design and implementation of the security policy. The certification agent is a special agent software implementing the authentication server to establish a variety of security policies. The authentication tokens serve the users in the form of hardware, software or smart card and so on, which are used to confirm the user's identity. If a user provides a correct token code, then it can be highly assured that the user is a legitimate user.

A complete SAP is the basis of the security system. It uses the combination of multiple software and hardware certification system, improving the reliability and supporting a variety of standard CA server. It is convenient and has less influence to the original system. At the same time, for all peripheral, input and output ports and operating license management, only authorized persons can achieve authority to operate the computer, and only authorized disk, disk partition, peripherals, mobile storage devices can be used by an authorized person on a authorized computer, and only authorized input and output ports can be used by a person authorized. All these measures lay the foundation for the reliable operation of the security system.

C. Information Security Platform （ISP）

In the ISP, Compulsory encryption to information over a network and Control of all network traffic were introduced in this article. That could effectively circumvent malicious listeners, unauthorized external connections and illegal access.

Communication protocol for computer networks is designed without considering its security and it is a completely open protocol. That makes it easy to be intercepted at random in the course of data transmission and exchange. To ensure information security within the network, security issues about important data must be solved in communication processes between any two machines in the LAN. The ISP proposed in this article makes mandatory encryption for network transmission come true and the communication key between any two computers is not the same. That effectively prevents the network behavior of malicious listener. At the same time, if host in the internal network gets access to the external network illegally through Modem, ADSL dial-up or dual card and other methods, they can not communicate with each other because of different data encapsulation. This effectively prevents the illegal behavior about access to the external network. Computers to the internal network from the external network, whether accessing to the internal network directly through the exchange of equipment or connecting to an internal computer through direct network connections, can not communicate with others, which effectively prevents the occurrence of illegal access.

D. Security management system （SMS）

A prefect SMS is essential to fight against insider threat of enterprises. This paper considers that security administrators should be able to keep abreast of the latest developments about network security and implement real-time monitoring of user behavior on the network. They should protect network equipment and the security of online information. It is also required that they can foresee network threats and take appropriate responses. At the same time, they should endeavor to cease the insider threat initially, and reduce internal risk to a minimum.

In addition, from the perspective of network security, enterprises take measures to manage employees. They should identify data that need to be protected, keep in touch with employees and provide security education everywhere. Firstly, leaders must recognize the importance of network security. Only in this way, can staff recognize it. Then some appropriate policies and regulations may be developed, so that enterprises can adhere to the principle that "there shall be laws to abide by and evidence to investigate, everyone who is meritorious should be reworded, everyone who is wrong should be punished." Only in that way can employees promote safety awareness and keep the internal network without damage.

Organizations must monitor all critical information system activity like servers, software applications and other data resources. Access must be strictly controlled and any suspicious activity must be investigated.

IV CONCLUSION

How to reduce insider threat? The use of advanced technology is required, but the establishment of insider threat for security architecture is essential. The advantage of this architecture is that it proposes an integrated approach on how to combine technology and management. However, details of the various platforms and advanced technologies aren't explained more and the factors including people and environmental issues are not analyzed accurately. From an overall point of view, in later research, many cooperative controls about technique, environment and people should be designed to be ordered and synchronous. At the same time, inter-linkages of various controls and their priority sequence and control principles should be fully considered.

REFERENCES

[1] GB. Magklaras,S. M. Furnell, "A preliminary model of end user sophistication for insider threat prediction in IT systems"[J], Computers and Security, 2005, vol. 24(5), pp. 371-380.

[2] M. Kemp, "Barbarians inside the gates: Addressing internal security threats" [J], Network Security, 2005, vol. 2005(6), pp. 11-13.

[3] Y. Yu, J. C. Chiueh, "Display-only file server: A solution against information theft due to insider attack" [C], Washington, DC, United States, 2004, pp. 31-39.

[4] RichardsonR. “2003 CSI/FBI computer crime and security survey” [J]. Computer Security Joumal 2003, 19(2): 21-40.

[5] Schultz E. “A Framework for Understanding and Predicting Insider Attacks” [J]. Computer and Security, 2002, 21(6):526-531.

[6] Hui wang, shu-fen liu, and yin-jia zhang, “Insider threat analysis and solution probe of for information system” [J], Jilin University Technology (Engineering Science), 2006, vol. 36(5), pp. 809-813.

[7] Park Joon S, Ho ShuyuanMary, “Composite role-based monitoring (CRBM) for countering insider threats” [J].Springer-Verlag Gm bH, 2004, 3073: 201-213.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15, August 2010, pp. 393-395

Application of Hybrid Filter in CT Image Processing Based on Visualization Toolkit

Chen Zhen[1], Li Guoli[2]
[1] School of Computer Science and Technology, Zhejiang University of Technology, Hangzhou, China
Email: son666666@gmail.com
[2] School of Information Engineering, Zhejiang University of Technology, Hangzhou, China
Email: guolili@zjut.edu.cn

***Abstract*—In order to improve the detail information of disease area on CT image, based on Visualization Toolkit, a hybrid filter combined Gaussian Smoothing Filter with improved Spatial Template Convolution Filter is proposed. The test example shows that the method has better capability on noise suppression and detail preservation.**

***Index Terms*—image processing, hybrid filter, Gaussian smoothing filter, spatial template convolution filter, visualization toolkit**

I. INTRODUCTION

With the development of medical science, CT image is playing an increasingly important role in medical diagnosis [1]. It is the basis for a number of disease diagnoses and treatment. CT images often contain a lot of noise because the patient's position moving or using metal markers when CT images were taken. It is a difficult question using the filter to remove noise while preserving the edge information [2]. The general de-noising filter loses image edge information when removes noise, which affects the diagnostic result. The smoothing filter reduces the noise, but makes images become blurred. The sharpening filter highlights the edge information of image, but the noise removal effect is not obvious [3].

This paper proposes a hybrid filter in CT processing based on VTK (Visualization Toolkit) [4] [5]. The hybrid method makes full use of the advantages of each filter and avoids their disadvantages. Test results show that the hybrid filter has the good ability of noise suppression and detail preservation, can improve the accuracy of disease organ sketch.

II. METHODS

For the purpose of deal with the radial noise caused by the metal markers on brain CT images, Gaussian smoothing filter is used to extract the low-frequency part firstly. And then the Spatial Template Convolution Filter is used to retain the high-frequency part of the image.

Gaussian Smoothing Filter

The Space Weighted Average Filters can be used for the pixel and weight selection problem. Gaussian Smoothing Filter is one of them. Gaussian smoothing filter is a linear smoothing filter selecting right value according to the shape of Gaussian function (Normal Distribution Function). It is efficiency for removing the normal distribution noise [6] [7] [8].

One-dimensional zero-mean Gaussian functions:

$$g(x) = e^{-\frac{x^2}{2\sigma^2}}. \tag{1}$$

The Gaussian distribution parameter σ determines the width of the Gaussian filter. Two-dimensional zero-mean Gaussian function is often used for image processing:

$$G(x,y) = Ae^{-\frac{x^2+y^2}{2\sigma^2}} = Ae^{-\frac{r^2}{2\sigma^2}}. \tag{2}$$

Sampling and quantifying the continuous Gaussian distribution, making the template normalized, the discrete template is obtained:

$$G^3 = \frac{1}{16}\begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix}. \tag{3}$$

VTK Implementation:

```
vtkImageGaussianSmooth * gauss
                = vtkImageGaussianSmooth::New();
gauss->SetInputConnection(reader-
>GetOutputPort());
gauss->SetDimensionality(2);
gauss->SetRadiusFactors (2,2,0) ;
```

Gaussian filter solves the problem of Space distance-weighted average, does not consider the changes in the pixel gradient reflecting the local features such as edges, etc.

Improved Spatial Template Convolution Filter

Spatial filter does convolution operations by template image in the image area. It is a process of multiplication and summation the image pixel gray value with the coefficient matrix. The corresponding coefficient matrix is called template. The space filtering process by 3×3 template is shown from Figure 1 to Figure 3 [9] [10].

Figure 1 is the input image, S0 is the current pixel (x, y); Figure 2 is a 3×3 template when we use in the template convolution processing; Figure 3 is the output image, R (x, y) is the output after template convolution processing:

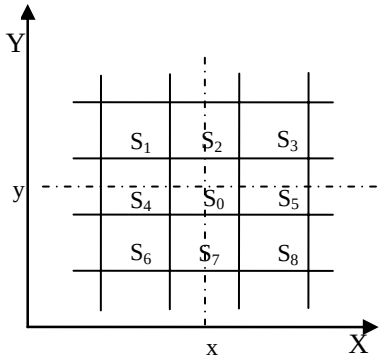


Figure 1.The input image.

k_1	k_2	k_3
k_4	k_0	k_5
k_6	k_7	k_8

Figure 2.The 3×3 template.

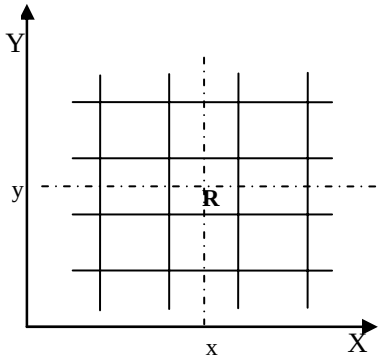


Figure 3.The output image.

$$R(x,y)=\sum_{i=0}^{8} k_i s_i \quad . \tag{4}$$

The results of template operator relate not only with pixel gray values, but also its neighborhood pixel gray values. Commonly used in edge detection convolution core templates are:

$$\begin{pmatrix} -1 & 0 & -1 \\ 0 & 4 & 0 \\ -1 & 0 & -1 \end{pmatrix} \begin{pmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{pmatrix} \begin{pmatrix} -1 & -1 & -1 \\ -1 & 9 & -1 \\ -1 & -1 & -1 \end{pmatrix} \begin{pmatrix} -1 & -2 & -1 \\ -2 & 4 & -2 \\ -1 & -2 & -1 \end{pmatrix}$$

Improved Spatial Template Convolution Filter using the output image pixels g_i after Gaussian smoothing filter replace s_i on the above formula in order to retain the details information produced by Spatial Template Convolution filter. The improved formula is shown as following:

$$R(x,y)=\sum_{i=0}^{8} k_i g_i \quad . \tag{5}$$

VTK Implementation:

```
const double kernel[9]={1,1,1,1,-8,1,1,1,1};
vtkImageConvolve
*conv=vtkImageConvolve::New();
conv->SetInputConnection(guass->GetOutputPort());
conv->SetKernel3x3  (kernel) ;
```

III. Results and Discussion

Figure 4 is a brain CT Image. Figure 5 is the disease area of nasopharyngeal which has external interference caused by metal markers outside human face.

Figure 6 is the result of Spatial Template Convolution Filter. Figure 7 is the disease area. Figure 7 shows that the boundary information of the image is clearer, and the

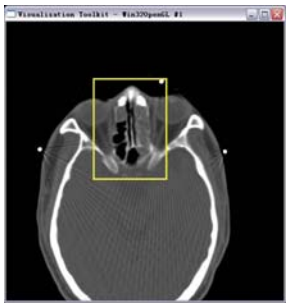


Figure 4.Brain imaging image.

Figure 5.The disease area in Figure 4.

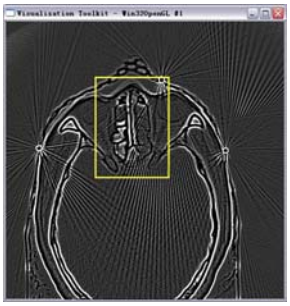


Figure 6.The result of Spatial Template Convolution Filter.

Figure 7.The disease area in Figure 6.

Figure 8.The result of Gaussian Smoothing Filter.

Figure 9.The disease area in Figure 8.

Figure 10.The result of Improved Spatial Template Convolution Filter.

Figure 11.The disease area in Figure 10.

impact of noise on the image is more severe with jagged border also.

Figure 8 is the result of Gaussian Smoothing Filter. Figure 9 is the disease area in Figure 8. From Figure 9 we can find that the image is blurred after smoothing filtered, but seems smoother.

Figure 10 is the result of Improved Spatial Template Convolution Filter. Figure 11 is the disease area in Figure 10. Comparing Figure 11 with Figure 5, it is obvious that the impact of noise on the image is reduced, and the edge information of image is clearer also.

IV. Conclusion

De-noising and edge detection are common problems and key issues in medical image processing. In this paper, the result of a hybrid filter shows that the method has better capability on noise suppression and detail preservation.

Acknowledgment

The work is supported by National 973 Planned Project (2006CB708307), the National Natural Science Foundation of China (60872112 and 10805012), the Natural Science Foundation of Zhejiang Province (Z207588), and the College Science Research Projects of Anhui Province (KJ2008B268).

References

[1] BU Jin-mei, XIAO Yi, LONG Mei, NI Li, LI Hong-yang, “Application of the computer in medical picture processing,” China Medical Education Technology, vol.15, no.4, pp.203-204, Dec 2001.

[2] YU Hua, WU Xiao-ming, CEN Ren-jing, YUAN Zhi-run, DENG Jun, “Filter of noises for DSA using Multi-SE morphological pyramid,” Journal of Jinan University (Natural Science), vol.21, no.1, pp.65-68, Feb 2000.

[3] RUAN Qiu-qi, “Digital image processing,” Electronic Industry Press, Beijing, pp.204-210, 2004.

[4] Xenophon Papademetris, “An Introduction to Programming for Medical Image Analysis with the Visualization Toolkit,” BioImage Suite, 2006.

[5] William J. Schroeder, “The VTK User’s Guide,” Kitware, Inc 1998-2001.

[6] XIE Qin-lan, “Adaptive Gaussian smoothing filter for image denoising,” Computer Engineering And Applications, vol.46, no.16, p.183, 2009.

[7] LI Bi-cheng, PENG Tian-qiang, PENG Bo, “Intelligent Image Processing,” Electronic Industry Press, Beijing, 2004.

[8] RUAN Qiu-qi, “Digital image processing (The 2nd Edition),”Electronic Industry Press, Beijing, 2007.

[9] ZHANG Guang-lie, ZHENG Nan-ning, LIANG Feng, WU yong, “Spatial Template Convolution Filtering for Video Processing Algorithm and Their VLSI Implementation,” Microelectronics & Computer, vol.20, no.2, p.27, 2003.

[10] ZHANG Yu-Jin, “Image processing and analysis,” Tsinghua University Press, Beijing, 1999.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15, August 2010, pp. 396-399

Improved Sparse Multi-path Channel Estimation via Modified Orthogonal Matching Pursuit

Jing Lu[1], Rui Wang[1], An-Min Huang[2]
1. School of Computer Science and Technology/ Henan Polytechnic University, Jiaozuo, China
wangrui@hpu.edu.cn
2. Department of Electronic Engineering University of Electronic Science and Technology, Chengdu, China

***Abstract*—Impulse response of sparse multi-path channel (SMPC) can be recovered from a short training sequence since most entries of SMPC are zeros. Though the ordinary orthogonal matching pursuit (OMP) algorithm provides a very fast implementation of SMPC estimation, it suffers from the coherence of atoms in dictionary, especially in the case of SMPC with a large delay spread and short training sequence. In this paper, a modified OMP method is proposed and a sensing dictionary is designed adaptively to improve the performance of the OMP algorithm. Numeral experiments illustrate that the proposed algorithm based on adaptive sensing dictionary outperforms the ordinary OMP algorithm.**

***Index Terms*—sparse multi-path channel (SMPC); generalized orthogonal matching pursuit (OMP); sensing dictionary; sparse approximation.**

I. INTRODUCTION

The problem of channel estimation has been studied extensively and some methods have been proposed in the literatures. Conventional methods for channel estimation were based on least-square (LS) algorithm. Unfortunately, all entries of the solution obtained by these methods were non-zeros and it was wrong in the case of sparse multi-path channel (SMPC). SMPC is frequently encountered in wireless communication applications and has only a small portion of entries is significantly different from zero. Taking advantage of the sparsity, impulse response of SMPC can be recovered from relatively small number of received data and training data. However, finding the sparsest solution is an NP-Hard combinatorial problem and massive works have been down to develop suboptimal methods for this problem.

Relax methods [1] and greedy algorithms [2, 3] are the most popular methods for finding the sparse solution. In particular, greedy algorithms, such as matching pursuit (MP) [2] and orthogonal matching pursuit (OMP) [3] can provide a very fast implementation of sparse approximation [4]. Some methods for sparse channel estimation have been proposed based on MP [5, 6]. However, according to the sufficient condition developed by Tropp [7], both MP and OMP suffer from highly coherence of redundant dictionary, especially in the case of SMPC with either large time delay spread or relatively small number of training data and received data. Recently, a modified OMP algorithm was developed to improve the performance of the ordinary OMP algorithm in the case of highly coherent dictionary by introducing a sensing dictionary [8]. However, this algorithm only considered the noiseless situation and the sensing dictionary is non-adaptively designed, which is independent of the received data. In this paper, a novel OMP algorithm is proposed to improve the performance of the ordinary OMP algorithm. An adaptively designed sensing dictionary is constructed and posterior information is utilized efficiently to prevent false atoms from being selected due to highly coherence between atoms in the ordinary dictionary. Numeral experiments illustrate that the performance of the proposed algorithm based on adaptive sensing dictionary is much better than that of the ordinary OMP algorithm.

The rest sections are organized as follows. In Section II, the sparse multi-path channel model is presented and the coherence between atoms is formulated. The adaptive approach to design sensing dictionary is given in Section III. Finally, we compare the performance of the proposed algorithm with other algorithms via simulations over wireless Gauss channel in Section Ⅳ and conclusions are given in Section Ⅴ.

II. PROBLEM FORMULATION

Let's transmit the training sequences $s(n)$, $n = 0,1,\cdots,N-1$, through a stationary multi-path sparse channel. The training sequence symbols $s(n)$ for $n < 0$ can be obtained from the previous estimates or for the first arriving frame they are assumed to be zero [5]. The received base-band signal samples can be modeled as

$$r_t = \sum_{i=0}^{L-1} s(t-i)h_i + e_t \,, \tag{1}$$

where $t = 0,1,\cdots,N-1$, h_i is the channel impulse response with length L and e_t is additive white Gaussian noise with zero mean and variance σ_e^2. Denote the power of training sequence and the received signal by σ_s^2 and σ_r^2, respectively. In the vector form, we have

$$\mathbf{r} = \mathbf{Sh} + \mathbf{e} \,, \tag{2}$$

where $\mathbf{h}=\begin{bmatrix} h_0 & h_1 & \cdots & h_{L-1} \end{bmatrix}^T$, $\mathbf{r}=\begin{bmatrix} r_0 & r_1 & \cdots & r_{N-1} \end{bmatrix}^T$, $\mathbf{e}=\begin{bmatrix} e_0 & e_1 & \cdots & e_{N-1} \end{bmatrix}^T$, T denotes transposition and $\mathbf{S}$ is the known training matrix given by

$$\mathbf{S}=\begin{bmatrix} s(0) & s(-1) & ... & s(-L+1) \\ s(1) & s(0) & ... & s(1) \\ \vdots & \vdots & \ddots & \vdots \\ s(N-1) & s(N-2) & ... & s(N-L) \end{bmatrix} = \begin{bmatrix} \mathbf{s}_0 & \mathbf{s}_1 & \cdots & \mathbf{s}_{L-1} \end{bmatrix} \quad (3)$$

Denote the number of nonzero entries of $\mathbf{h}$ as K. The channel $\mathbf{h}$ is sparse if $K \ll L$ holds. In the context of sparse analysis, $\mathbf{S}$ is called as dictionary and the column vector $\mathbf{s}_i$ $(i=0,1,\cdots,L-1)$ as atom. As a result of short training sequence, which improves throughput efficiency for the systems where transmitted packet length is short, the dictionary is highly redundant. In other word, the dimension of the received base-band signal vector $\mathbf{r}$ is much smaller than the number of atoms in the dictionary, i.e., $N \ll L$.

The ordinary OMP algorithm iteratively selects the atom that correlates most strongly with the residual signal. At each step *k*, the best atom $\mathbf{s}_{m_k}$ is selected by solving the simple optimization

$$m_k = \arg\max_{0\le i\le L-1} \hat{h}_i^{(k)}, \quad (4)$$

where

$$\hat{\mathbf{h}}^{(k)} = \begin{bmatrix} \hat{h}_0^{(k)} & \hat{h}_1^{(k)} & \cdots & \hat{h}_{L-1}^{(k)} \end{bmatrix}^T = \left|\mathbf{S}^T \mathbf{g}_{k-1}\right|, \quad (5)$$

where $k=1,2,\cdots,K$. We have $\mathbf{g}_0=\mathbf{r}$ for initialization and $\mathbf{g}_k=\mathbf{P}_k\mathbf{r}$ for $k=1,2,\cdots,K-1$, where

$$\mathbf{P}_k = \mathbf{I}_M - \hat{\mathbf{A}}^{(k)}\left(\hat{\mathbf{A}}^{(k)T}\hat{\mathbf{A}}^{(k)}\right)^{-1}\hat{\mathbf{A}}^{(k)T}, \quad (6)$$

$$\hat{\mathbf{A}}^{(k)} = \begin{bmatrix} \mathbf{s}_{m_1} & \mathbf{s}_{m_2} & \cdots & \mathbf{s}_{m_k} \end{bmatrix}, \quad (7)$$

and $\mathbf{I}_M$ is an identity matrix.

To illustrate the effect of coherence between atoms, e.g., at the initialization step, we express the sparse channel estimation as

$$\hat{\mathbf{h}}^{(1)} = \left|\mathbf{S}^T\mathbf{g}_0\right| = \left|\mathbf{S}^T\mathbf{r}\right| = \left|\mathbf{S}^T(\mathbf{S}\mathbf{h}+\mathbf{e})\right|, \quad (8)$$

or

$$\hat{h}_i^{(1)} = \left|\sum_{l=0}^{L-1}\mathbf{s}_i^T\mathbf{s}_l h_l + \mathbf{s}_i^T\mathbf{e}\right|. \quad (9)$$

for $i=0,1,\cdots,L-1$. If $h_l=0$, the coherence $\mathbf{s}_i^T\mathbf{s}_l$ can not affect the estimated value of $\hat{h}_i^{(1)}$. However, if $h_l\neq 0$, $\mathbf{s}_i^T\mathbf{s}_l$ will draw the estimated value of $\hat{h}_i^{(1)}$ away from its correct value h_i. As a result, we may either choose a false atom when $h_i=0$ or omit a correct atom when $h_i\neq 0$ at this step if the coherence is large enough. Here, the problem is how to mitigate the effect of the coherence on the performance of OMP algorithm.

III. THE PROPOSED ALGORITHM

In order to identify the correct atoms in the case of high coherence, we resort to the modified OMP based on a sensing dictionary $\mathbf{W}$, and use $\hat{\mathbf{h}}^{(k)}=\left|\mathbf{W}^T\mathbf{g}_{k-1}\right|$ rather than $\hat{\mathbf{h}}^{(k)}=\left|\mathbf{S}^T\mathbf{g}_{k-1}\right|$ in (5). Obviously, the ordinary OMP is a special case of the general OMP with $\mathbf{W}=\mathbf{S}$. Given the received signal $\mathbf{r}$, the probability of appearance in the reconstruction of $\mathbf{r}$ is different for different atom [9]. Therefore, the adaptive sensing vector is taken as the solution to the following optimization

$$\min_{\mathbf{w}_i} \mathbf{w}_i^T\mathbf{S}\mathbf{U}^{(k)}\mathbf{S}^T\mathbf{w}_i \quad (10)$$

$$s.t.\ \ \mathbf{s}_i^H\mathbf{w}_i = 1, \quad (11)$$

where $i=0,1,...,L-1$, $\mathbf{U}^{(k)}=\mathrm{diag}(\left|\hat{\mathbf{h}}^{(k)}\right|^{\rho})$, $\hat{\mathbf{h}}^{(k)}=\left|(\mathbf{W}^{(k)})^T\mathbf{r}\right|$ and $\rho>0$. Similarly, the closed-form solution can be given by

$$\mathbf{w}_i = \mathbf{D}_i\mathbf{s}_i, \quad (12)$$

where

$$\mathbf{D}_i = \frac{1}{\mathbf{s}_i^T\left(\mathbf{S}\mathbf{U}^{(k)}\mathbf{S}^T+\beta\mathbf{I}_M\right)^{-1}\mathbf{s}_i}\left(\mathbf{S}\mathbf{U}^{(k)}\mathbf{S}^T+\beta\mathbf{I}_M\right)^{-1}, \quad (13)$$

for $i=0,1,\cdots,L-1$, and β is a positive regularization parameter. We here take the correlation between the received vector (or the residual vector) and each atom in the dictionary as an approximate measure of this probability. Because $\mathbf{U}^{(k)}$ is calculated from the sensing dictionary itself, we must set an initial sensing dictionary such as $\mathbf{W}=\mathbf{S}$.

The sensing dictionary given by (13) is the adaptive function as a result of the adaptive minimum coherence optimization with the distortionless response constraint. It

is easy to see that the sensing dictionary given by the non-adaptive design method [8], which is completely determined by the dictionary $\mathbf{S}$ and independent of the received signal, corresponds to a special case of (13) with $\mathbf{U}^{(k)} = \mathbf{I}_L$ ($\mathbf{I}_L$ is an identity matrix) at each step.

Accordingly, the modified OMP algorithm based on adaptive sensing dictionary can be summarized as follows.

(1) Initialization: $\mathbf{g}_0 = \mathbf{r}$, $\mathbf{W} = \mathbf{S}$, $k = 1$, $\beta > 0$ and $\rho > 0$;

(2) For $i = 0,1,\cdots,L-1$, repeat the following process for J times:

$$\hat{\mathbf{h}}^{(k)} = \left|\mathbf{W}^T \mathbf{g}_{k-1}\right|, \quad \mathbf{U}^{(k)} = \mathrm{diag}(\left|\hat{\mathbf{h}}^{(k)}\right|^{\rho}),$$

$$\mathbf{D}_i = \frac{1}{\mathbf{s}_i^T \left(\mathbf{S}\mathbf{U}^{(k)}\mathbf{S}^T + \beta \mathbf{I}_M\right)^{-1} \mathbf{s}_i} \left(\mathbf{S}\mathbf{U}^{(k)}\mathbf{S}^T + \beta \mathbf{I}_M\right)^{-1},$$

$$\mathbf{w}_i = \mathbf{D}_i \mathbf{s}_i, \quad \mathbf{W} = \left[\mathbf{w}_0 \quad \mathbf{w}_1 \quad \cdots \quad \mathbf{w}_{L-1}\right];$$

(3) $\hat{\mathbf{h}}^{(k)} = \left[\hat{h}_0^{(k)} \quad \hat{h}_1^{(k)} \quad \cdots \quad \hat{h}_{L-1}^{(k)}\right]^T = \left|\mathbf{W}^T \mathbf{g}_{k-1}\right|$

$$m_k = \arg \max_{0 \le i \le L-1} \hat{h}_i^{(k)},$$

$$\hat{\mathbf{A}}^{(k)} = \left[\mathbf{s}_{m_1} \quad \mathbf{s}_{m_2} \quad \cdots \quad \mathbf{s}_{m_k}\right],$$

$$\mathbf{P}_k = \mathbf{I}_M - \hat{\mathbf{A}}^{(k)} \left(\hat{\mathbf{A}}^{(k)T} \hat{\mathbf{A}}^{(k)}\right)^{-1} \hat{\mathbf{A}}^{(k)T}, \quad \mathbf{g}_{k+1} = \mathbf{P}_k \mathbf{r}.$$

(4) $k = k+1$, go to (2) and repeat until $k = K$.

Finally, the position of the nonzero entries of SMPC is detected by $\left[m_1 \quad m_2 \quad \cdots \quad m_K\right]$, and the corresponding nonzero values are estimated as $\left(\hat{\mathbf{A}}^{(k)T} \hat{\mathbf{A}}^{(k)}\right)^{-1} \hat{\mathbf{A}}^{(k)T} \mathbf{r}$. To reduce the computation cost, the sensing dictionary can be calculated only for $k = 1$ and used at the subsequent steps

IV. Simulation Results

To gain some insights into the performance of the proposed algorithm, we carry out some experiments of SMPC estimation. The nonzero entries of SMPC are drawn randomly from a uniform distribution on $[-1,-0.2] \cup [0.2,1]$ and the number of nonzero entries is K=5. The position of nonzero entry of $\mathbf{h}$ is generated randomly. The channel length is set as L=50 or 100, the length of training sequence is N=30, and the signal to noise ratio (SNR) is 10 dB. The other involved parameters used in the algorithms are set to $\rho = 3$, $\beta = 0.1$ and $J = 10$, which may be further optimized to obtain better performance. Simulation results are obtained over 10000 independent Monte-Carlo trials.

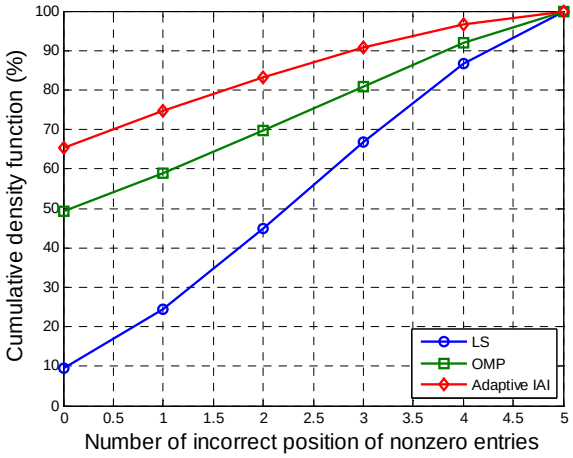


Figure 1. Cumulative density function of the number of incorrect position of nonzero entries with N=30, L=50.

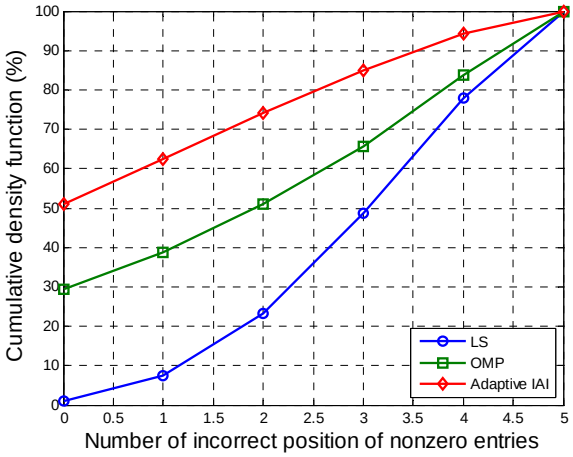


Figure 2. Cumulative density function of the number of incorrect position of nonzero entries with N=30, L=100.

We compare performance of the modified OMP algorithm based on adaptive sensing dictionary with that of the least squares method (LS) and the ordinary OMP algorithm. We compare the ability of these algorithms to detect the nonzero entries of SMPC. The cumulative density functions (CDF) of the number of incorrectly detected nonzero components for the channel length values of 50 and 100 are shown in figures 1 and 2, respectively. From these CDF functions, we see that the proposed algorithm based on adaptive sensing dictionary gives more accurate detection of nonzero entries of SMPC than other algorithms. These results show that the modified OMP algorithm based on adaptive sensing dictionary significantly outperforms the other methods, especially in the case of more redundant dictionary of the latter (L=100).

V. Conclusion

In this paper, we propose a modified OMP algorithm by introducing an adaptive sensing dictionary. This algorithm significantly improves the performance of sparse multi-path channels estimation especially in the case of sparser and longer SMPC. The numeral experiments indicate that the proposed algorithm

outperforms both the ordinary OMP algorithm and the modified OMP algorithm based on non-adaptive sensing dictionary.

REFERENCES

[1] S. S. Chen, D. L. Donoho, and M. A Saunders, "Atomic decomposition by Basis Pursuit," *SIAM Journal on Scientific Computing*, vol.20, no.1, pp.33–61, 1999.

[2] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," IEEE Tran. on ASSP, vol. 41, no. 12, pp. 3397-3415, Dec. 1993.

[3] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal Matching Pursuit: Recursive function approximations to wavelet decomposition," In Proceeding of the 27th Annual Asilomar Conference on Signals Systems and Computers, pp.40-44, Nov. 1993.

[4] R. Gribonval and S. Krstulovic, "MPTK: Matching Pursuit made tractable," In Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP'06), pp.496-499, May 2006.

[5] S. F. Cotter and B. D. Rao, "Sparse channel estimation via Matching Pursuit with application to equalization", IEEE Tran. on Commun., vol.50, no.3, pp.374-377, Mar. 2002.

[6] S. Kim and RA Iltis, "A matching-pursuit/GSIC-based algorithm for DS-CDMA sparse-channel estimation", IEEE Signal Process. Lett., vol. 11, no. 1, PP. 12-15, 2004.

[7] J. Tropp, "Greed is good: Algorithmic results for sparse approximation," IEEE Trans. Inf. Theory, vol.50, pp.2231-2242, Oct. 2004.

[8] K. Schnass and P. Vandergheynst, "Dictionary Preconditioning for Greedy Algorithms," IEEE Trans. Inf. Theory, vol.56, no.5, pp.1994-2002, May 2008.

[9] E. O. Divorra, L. Granai, and P. Vandergheynst, "On the use of a priori information for sparse signal approximations," IEEE Trans. Inf. Theory, 54, (9), pp.3468-3452, Sept. 2006

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15, August 2010, pp. 400-404

Automatic Identification of Parallel Structure Based on Conditional Random Field

Wang Dongbo, Zhu Danhao, Su Xinning, Xie Jing
Department of Information Management, Nanjing University, Nanjing, China
Email: wangdongbo0102@gmail.com, yangminggaol@163.com

***Abstract*—Based on conditional random field (CRF), the article explores the identification of parallel structure by making use of Tsinghua University 973 Treebank. The CRF tag set is defined as T={B，F，G，I，M，E，S}. The CRF template is a complex feature template that contains 18 features composed of word and POS and simple linguistic feature template that contains 25 features composed of word, POS and coordinate conjunction. The best F-measure scale in the parallel structure identification reaches 83.23% and 83.71% based on complex feature template and simple linguistic feature template respectively in the open test.**

***Index Terms*—parallel structure,conditional random field(CRF), feature template, tag set, Tsinghua Treebank**

I. INTRODUCTION

Parallel structure that has more than two centers is regarded as an inner center-structure, and each center has the same syntactic and semantic function as the whole structure. Parallel structure can be divided into two types, i.e. one with coordinate conjunction and the other without coordinate conjunction from the formal point of view, for example "高等植物和高等动物以及人类" and "生产实践". The identification of parallel structure is favorable for constructing large scale Chinese treebank, which is of much use for the researches of machine translation and information extraction[1]. If the parallel structure of "改革/vn 开放/vn 和/c 现代化/vn 建设/vn" is identified in advance from the sentence of "在/p 这/r 一/m 年/q 中/f ，/w 中国/ns 的/u 【改革/vn 开放/vn 和/c 现代化/vn 建设/vn】 继续/v 向前/v 迈进/v。", the whole sentence will be parsed[2] easily. Linguistic researchers can take advantage of the identification results from the statistical point of view.

II. LITERARY REVIEW

In previous studies, researchers have analyzed and identified the parallel structure from various perspectives. Agarwal, Boggess(1992)[3] identifies conjuncts of coordinate conjunctions appearing in text which has been labelled with syntactic and semantic tags. The algorithm is tested on a 10,000 word chapter of the Merck Veterinary Manual. The best accuracy rate of the algorithm on the 'Eye and Ear' chapter is 81.6%. Akitoshi Okumura and Kazunori Muraki(1994)[4] propose an English coordinate structure analysis model, which provides top-down scope information of the correct syntactic structure by taking advantage of the symmetric patterns of the parallelism, and the accuracy of the model in the MT analysis system is 75%. Kurohashi, S. and Nagao, M(1994)[5] presents a syntactic analysis method that first detects conjunctive structures in a sentence by checking parallelism of two series of words and then analyzes the dependency structure of the sentence with the help of the information about the conjunctive structures. 150 Japanese sentences are analyzed to illustrate the effectiveness of this method. Zhou Qiang(2002)[6] identifies the parallel structure, learning the way of identifying the Japanese parallel structure, and comes to the conclusion that the errors of identifying Chinese parallel structures are serious. Zhan Weidong(1999)[7] points out that the identification of unmarked parallel structures is very difficult by analyzing the ambiguity types of parallel structure. Sun Honglin(2001)[8] improves the efficiency of identifying the parallel structure by identifying the parallel structure boundary based on a simple probability model. Wu Yunfang(2003)[9] comprehensively analyzes the parallel structures by taking Chinese information processing as the starting point. Wang Dongbo, Chen Xiaohe and Nian Hongdong(2008)[10] identifies the Coordination with Overt Conjunctions based CRF, but the Coordination without conjunction is not identified. Miao Yanjun, Li Junhui and Zhou Guodong(2009)[11] combines maximum entropy model and several novel rules to automatically identify boundaries of coordinate structure. Although the best F scale is 78.1%, the coordinate structure without conjunction is not identified. The article explores the identification of parallel structure which includes parallel structure with coordinate conjunction and parallel structure without coordinate conjunction by conditional random field based on Tsinghua University 973 Treebank, and it is proved that CRF is very practical.

III. AUTOMATIC IDENTIFICATION OF PARALLEL STRUCTURE

A. Conditional Random Fields Description

Conditional Random Field (CRF), a recently introduced conditioned probabilistic model for labeling and segmenting sequential data, is an undirected graph model which calculates the conditional probability over output nodes given the input nodes. To be more specific, if X and Y are labeled data sequence and sequence label random variable and X and Y are joint distribution, observation sequence and label sequence will form a conditional model which is p (Y|X). Definition: G=(V,E) is an undirected graph, if $Y = (Y_v)_{v \in V}$ follows

Markovian which means $p(Y_v \mid X, Y_w, v \neq w) = p(Y_v \mid X, Y_w, v \sim w) \quad v \sim w$ expresses that w is G adjacent note), (X， Y) is Conditional Random Field. Under the condition of observation sequence $X=(X_1, X_2,\ldots\ldots, X_n)$ and labeled sequence $Y=(Y_1,Y_2,\ldots\ldots,Y_n)$[12], Y is a tree (a chain structure in the simplest condition). Thus, based on the basic theory of Conditional Random Field, the joint probability of X and Y label sequence can be worked out by the following formula:

$$P_\theta(y \mid x) \propto \exp\left(\sum_{e \in E,k} \lambda_k f_k(e, y|_e, x) + \sum_{v \in V,k} \mu_k g_k(v, y|_v, x)\right)$$

X is data sequence, and *y* is label sequence, and *v* is vertex set, and *e* is limbic set, and *k* is feature number. The identifying of parallel structure is just like that the sequences of word and POS (part of speech) select label and identify the boundary by the CRF. Since the length of parallel structure may be long or short, the tag set of CRF is relatively complex.

B. Corpus Preprocessing

The experiments are based on the tool of CRF++ which is developed by C++[13] .When CRF++ is used, train and test files must contain many tokens, each of which contains many columns separated by space or tab and is written on a line. Columns are separated by tab when the corpus is trained and tested. The whole token sequence forms a sentence which is separated by blank line. The train and test corpus is like the following.

Table 1. Example of train and test corpus

Word	POS	Tag
中国	nS	S
古代	t	S
财政	n	S
为	vC	S
“	w	B
度支	n	F
”	w	G
、	w	I
“	w	M
国用	n	M
”	w	M
、	w	M
“	w	M
岁计	n	M
”	w	M
、	w	M
“	w	M
国计	n	M
”	w	E
。	w	S

“中国 ns S” is treated as a token that includes three lines which are word, POS and the tag set of parallel structure respectively.

The basic format of the template is *%x[row,col]*, which is used to determine the token in which row is used to determine the relative row and line is used to determine the absolute column in the input data. Feature templates are divided into the Unigram template which is efficient when the corpus is trained and tested and Bigram template which is easy to ensure the test result consistency based on possible string. The flowing is the example of parallel structure train or test corpus and feature template.

Train or Test Corpus	**Feature Template**	
财政 n S	U030:%x[-1,0]	词源
许毅 nP S	U040:%x[0,0]	与
词源 n B	U01020:%x[-1,1]	n
与 c F	U01030:%x[0,1]	c
起源 n E	U511:%x[0,0]/%x[0,1]	与/c

The identification of parallel structure makes use of the Unigram template from the efficiency of train and test consideration. The train and test corpus is Tsinghua University 973 Treebank. The basic information about the Treebank is in the table 2 and table 3.

Table 2. Basic data of Tsinghua Treebank

Style	Files	Sentence	Words	Chinese character	Word length
Literature		16335	340208	415040	20.83
News	154	6877	173942	246757	25.29
Acdemic	15	5589	158780	240289	28.41
Application	195	3169	66586	97924	21.01
Sum	503	31970	739516	1000010	23.13

Table 3. Distributed data of sentence length in Tsinghua Treebank

	Simple sentence			Complex sentence		
Literature	Sentences	Words	Average length	Sentence	Words	Average length
News	9692	102895	10.62	6643	237313	35.72
Acdemic	3025	34023	11.25	3852	139919	36.32
Application	2021	24204	11.98	3568	134576	37.72
Sum	16608	178068	10.72	15362	561458	35.90

The parallel structures in the Tsinghua Treebank are marked by the label “xx－LH”. The parallel structure of “[np-LH 分配/vN 、/、 再分配/n]” in the sentence of “[zj-XX [dj-ZW 财政/n [vp-PO 是/vC [np-LH [dj-ZW 国家/n [vp-ZZ [pp-JB 为/p [vp-PO 实现/v [np-DZ 其/rB 职能/n]]] ，/， [vp-PO 参与/v [np-DZ [np-DZ [np-LH [np-DZ 社会/n 产品/n] 和/c 国民收入/n] [np-LH 分配/vN 、/、 再分配/n]] 活动/n]]]] 及其/c [np-DZ

形成/v 的/u [np-DZ 分配/vN 关系/n]]]]] 。/。]" is labeled by "np-LH".

The article uses the following formula to calculate the Conditional Random Fields tag set.

$$L_k = \frac{1}{N}\sum_{i=k}^{k} iN_k \quad (k>2)$$

$\boldsymbol{L_k}$ ($i>=k$) is the mean weighted parallel structure length ,and $\boldsymbol{N_k}$ is the parallel structure frequency whose length is k, and k is the longest parallel structure length in the corpus, and N stands for parallel structure total number in the corpus in the formula. The article shows that CRF tag set is 7-word tag set which is T={B，F，G，I，M，E ， S}, based on above the formula and detailed experiments. In the tag set, B represents the beginning word of parallel structure, F represents the second word of the structure, G represents the third word of the structure, I represents the fourth word of the structure, M represents the fifth or more word of the structure, E represents the ending word of the structure, and S represents parallel structure outside word.

C. Feature Selection and Feature Template Definition

Feature is the key problem during the parallel structure identification by CRF, for whether the feature is good or not will affect the identification results. The identification in this article is divided into complex features identification and simple linguistic features identifiction. Complex features consist of word and part of speech which are marked as W and P respectively on the basis of complex feature template. The word observing window is 7 which is {-3, -2, -1, 0, 1, 2, 3}, and the pos observing window is 5 which is {-2, -1, 0, 1, 2}. In the observing window, 0, -1 and 1 respectively represents current location, prior location and next location. Based on observing window and experimental results, this article puts forward that the complex features model has 18 features which are { W-3, W-2, W-1, W, W+1, W+2,W+3, W-1/W, W/W+1, W-1/W+1, P-2, P-1, P, P+1, P+2, P-1/P, P/P+1, W/P }.The complex feature template is the following.

Unigram
U010: %x [-3, 0]
U020: %x [-2, 0]
U030: %x [-1, 0]
U040: %x [0, 0]
U050: %x [1, 0]
U060: %x [2, 0]
U061: %x [3, 0]
U0300: %x [-1, 0]/%x[0, 0]
U0400: %x [0, 0]/%x[1, 0]
U01000: %x [-1, 0]/%x[1, 0]
U01010: %x [-2, 1]
U01020: %x [-1, 1]
U01030: %x [0, 1]
U01040: %x [1, 1]
U01050: %x [2, 1]
U01060: %x [-1, 1]/%x[0, 1]
U01070: %x [0, 1]/%x[1, 1]
U511: %x [0, 0]/%x[0, 1]
Bigram
B

Simple linguistics features consist of word and part of speech and the coordinate conjunction which are marked as W, P and Y/N respectively on the basis of simple linguistic feature template. The word observing window is 7 which is {-3, -2, -1, 0, 1, 2, 3}, and the pos observing window is 5 which is {-2, -1, 0, 1, 2}. The coordinate conjunction observing window is 5 which is {-2, -1, 0, 1, 2}. Based on observing window and experimental results, this article puts forward that the simple linguistic feature model has 25 features which are { W-3, W-2, W-1, W, W+1, W+2,W+3, W-1/W, W/W+1, W-1/W+1, P-2, P-1, P, P+1, P+2, P-1/P, P/P+1, Y/N-2, Y/N -1, Y/N, Y/N+1, Y/N+2, (Y/N-1)/ Y/N, Y/N/ (Y/N+1), W/P }.The simple linguistic feature template is the following.

Unigram
U010: %x [-3, 0]
U020: %x [-2, 0]
U030: %x [-1, 0]
U040: %x [0, 0]
U050: %x [1, 0]
U060: %x [2, 0]
U061: %x [3, 0]
U0300: %x [-1, 0]/%x[0, 0]
U0400: %x [0, 0]/%x[1, 0]
U01000: %x [-1, 0]/%x[1, 0]
U01010: %x [-2, 1]
U01020: %x [-1, 1]
U01030: %x [0, 1]
U01040: %x [1, 1]
U01050: %x [2, 1]
U01060: %x [-1,1]/%x[0,1]
U01070: %x [0, 1]/%x[1,1]
U02010: %x [-2, 2]
U02020: %x [-1, 2]
U02030: %x [0, 2]
U02040: %x [1, 2]
U02050: %x [2, 2]
U02060: %x [-1, 2]/%x[0, 2]
U02070: %x [0, 2]/%x[1, 2]
U511: %x [0, 0]/%x[0, 1]
Bigram
B

The sample of train and test corpus based on simple linguistic template is as in the tale 4.

Table 4 . Sample corpus based on simple linguistics template

Word	POS	Tag	Coordinate conjunction
中国	nS	S	N
古代	t	S	N
财政	n	S	N
为	vC	S	N
“	w	B	N
度支	n	F	N
”	w	G	N
、	w	I	Y
“	w	M	N
国用	n	M	N
”	w	M	N
、	w	M	Y
“	w	M	N
岁计	n	M	N
”	w	M	N
、	w	M	Y
“	w	M	N
国计	n	M	N
”	w	E	N
。	w	S	N

D. Automatic Identification of Parallel Structure Framework

The automatic identification of parallel structure is composed of training part in which features are extracted on the basis of feature template and the parameters of template are obtained by GIS algorithm and test part in which the result of open test is obtained. The following is the framework:

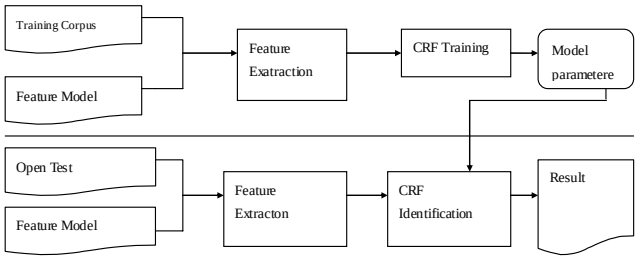


Figure 1. Automatic identification of parallel structure framework

IV. RESULTS

The test experiment is an open test making use of Tsinghua University 973 Treebank whose scale reaches 100 million Chinese charactres. There are 27137 parallel structures, not including nesting parallel structures in the tree bank. The Treebank is divided into 10 parts, 9 of which are used to train and 1 part is used to test, so as to ensure the experiments fair and responsible. There are 10 train and test experiments used 18 complex feature template, and 10 train and test experiment used simple linguistic feature template. The results of the open test are measured by precision, recall and F scale. Experiment Environment: Operation system: Windows XP, CPU: Intel Core2 Duo, Frequency: 2.40GHz,Memory: 2GB. The results of 10 experiments used complex feature template are the following.

Table 5. Open test result based on complex feature template

Train scale		Precision	Recall	F-measure
Open Test	2-10	82.29%	80.44%	81.35%
	1,3-10	81.49%	79.86%	80.67%
	1,2,4-10	83.56%	81.23%	82.38%
	1-3,5-10	79.63%	80.12%	79.87%
	1-4,6-10	81.22%	83.21%	82.20%
	1-5,7-10	85.27%	78.52%	81.76%
	1-6,8-10	83.22%	83.24%	83.23%
	1-7,9-10	79.53%	81.27%	80.39%
	1-8,10	82.35%	82.34%	82.34%
	1-9	80.79%	78.96%	79.86%

The best F-measure reaching 83.23% is to test the 7 part. The results of 10 experiments used simple linguistic feature template are the following.

Table 6. Open test result based on simple linguistic feature template

Train scale		Precision	Recall	F-measure
Open Test	2-10	82.89%	80.54%	81.70%
	1,3-10	82.49%	79.96%	81.21%
	1,2,4-10	83.76%	81.83%	82.78%
	1-3,5-10	81.12%	80.42%	81.12%
	1-4,6-10	83.71%	82.65%	83.71%
	1-5,7-10	78.92%	81.98%	78.92%
	1-6,8-10	83.54%	83.41%	83.54%
	1-7,9-10	81.37%	80.45%	81.37%
	1-8,10	82.44%	82.54%	82.44%
	1-9	78.97%	79.87%	78.97%

The best F-measure reaching 83.23% is also to test the 7 part.

Compared with the complex feature template, the result based on simple linguistic feature template is better, because the feature of coordinate conjunction is be conducive to identify the boundaries of parallel structure.

V. CONCLUSION

The article studies the identification of the parallel structure based on CRF and Tsinghua University 973 Tree-Bank. The best F scale in the identification of parallel structure reaches 82.38% in the open test. In the future, linguistic knowledge can be added into the feature template in order to improve precision and recall. And the

nesting parallel structures can be identified in order to identify overall parallel structure.

ACKNOWLEDGMENT

The authors wish to thank Professor Chen Xiaohe and Doctor Li Bin. This work was supported in part by a grant from the Research of Knowledge Mining Technology and application Based on Intelligent Information Process (Grant No. 08JJD870225) which is supported by the Foundation from Ministry of Education of China and the research of Automatic Acquisition of English–Chinese Parallel Pairs from Websites (Grant No. 2010CW02) which is Supported by the Scientific Research Foundation of Graduate School of Nanjing University.

REFERENCES

[1] Fei Sha, Fernando Pereira (2003) 'Shallow Parsing with Conditional Random Fields'. In Proceedings of Human Language Technology Conference and North American Chapter of the Association for Computational Linguistics (HLT-NAACL):135-136.

[2] Church.K. Astochastic parts program and noun phrase parser for unrestricted text[C]. In: Proceedings of the 2nd Conference on Applied Natural Language Processing. Austin: Association for Computational Linguistics, 1988, pp: 136-143.

[3] Agarwal, R. and Boggess L.A simple but useful approach to conjunct identification[C]. In: Proceedings of 30th Annual Meeting of Asscosiation for Computational Linguistics.1992, pp15-20.

[4] Okumura, A. and Muraki, K. Symmetric pattern matching analysis for English coordinate structures[C].In: Proceedings of the 4th Conference on Applied Natural Language Processing.1994, pp:41.

[5] Kurohashi, S. and Nagao, M. A syntactic analysis method of long Japanese sentences based on the detection of conjunctive structures[J]. Computational Linguistics, 1994, Vol.20,No.4, pp:530.

[6] Zhou Qiang ,"Phrase Bracketing and Annotating on Chinese Language Corpus". Beijing: Peking University, 2002.

[7] Zhan Weidong, "A Study of Constructing Rules of Phrases in Contemporary Chinese for Chinese Information Processing". Beijing: Peking University, 1999.

[8] Sun Honglin,"Aanalyzing the Chunking in the Unconfined Chinese Text". Beijing: Peking University, 2003.

[9] Wu Yunfang (2003) 'Study on Chinese Coordination for Chinese Information Processing' .Beijing: Peking University.

[10] Wang Dongbo,Chen Xiaohe and Nian Hongdon, "Automatic Identification of Coordination with Overt Conjunctions Based on Conditional Random Fields", JOURNAL OF CHINESE INFORMATION PROCESSING, Vol.22, No.6, 2008, pp.3-4.

[11] Miao Yanjun, Li Junhui and Zhou Guodong, "Automatic identification of coordinate structure based on statistics and rules", Application Research of Computer, Vol.26, No.9, pp.3403-3404.

[12] J. Hammersley, & P.Clifford (1971) 'Markov fields on finite graphs and lattices'. Unpublished manuscript.

[13] The features of CRF++[EB/OL].[2009-12-06]. http://crfpp.sourceforge.net/#features.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15, August 2010, pp. 405-408

A Handoff Method Based on AAA for MIPv6

Jia Zong-pu[1], Zhang Jing[2]
[1] Computer Science and Technology Department, He Nan Polytechnic University Jiao Zuo, China
Email: jiazp@hpu.edu.cn
[2] Computer Science and Technology Department, He Nan Polytechnic University Jiao Zuo, China
Email: zhangjing8754@hotmial.com

***Abstract*—In the era of commercial demand increased day by day, the mobile IP protocol combined with AAA (Authentication, Authorization and Accounting) technology is widely used in authentication, authorization and billing issues. However, compared to single mobile IP switch protocol, because MIPv6-AAA model need achieve AAA user's authentication and authorization in the process of switch, so it will generate more switch time delay, and also have security issues. Therefore, this article give a new MIPv6 switch method, it is when MN switch in the inner-domain, do not need the authentication of home domain, and reduce the switch time; but when switch in the inter-domain it will set mobile node agent (MNA) to save original MN information temporarily, to avoid the registration process failed, and increase the security. This solution achieved MIPv6-AAA model optimization through improve these two areas.**

***Index Terms*—MIPv6, AAA, handoff, MN, agent**

I. INTRODUCTION

As computer and communication technologies developed, people have more and more requirement for the network services. Traditional fixed access Internet mode can not afford people's requirement; they need wireless internet services. Mobile IP protocol can combine with any link layer technology, and support the vertical switch, make the user can continue access the network when they are moving, and this is considered as the best solution for the mobility problems of network layer. Currently, large-scale increased in mobile users, and the Internet for business applications is also become popular. For this, IETF (Internet Engineering Task Force) make AAA (Authentication, Authorization and Accounting) combined with mobile IP technology, focus on solving the user's authentication, authorization and accounting issues, provide security for mobile IP achieve large-scale commercial business.

At present, there are many researches on combine the AAA with mobile IP, reference [1] described a solution to design the layer structure and set facilities, give the normalizing process of AAA certification and MIPv6 registration process, and the process of establish local SA authenticate, and also compared with the existing solution, pointing out the advantages of the solution. From the performance analysis we can see that, MN's movement character is the important parameter which affecting the performance. But this solution did not consider how to use mobile switching rate, dwell time and other parameters to describe the MN movement characteristics, and guide AAA structure become layer and dynamic adjustment.

Reference [2] give a structure of combine mobile IPv6 with AAA based on WLAN, use RADIUS as the protocol of AAA, but RADIUS just can support IPv4, so under the situation of MIPv6, there has problem that AAAH and AAAL use RADIUS protocol to communicate, reference [2] said use NAT-PT to solve the problem of transmit IPv4 packet though IPv6 network, but when they use this mechanism the system become instability.

The system in [6] is under MIPv6 they use netfilter structure of Linux operation system to implement the function of authentication, authorization and billing of Diameter AAA, and use IPSec6 to catch stream of IPv6, and then use proper AH/ESP process module to deal with it. In this solution even it realize access control and safety communication, but it can cause communicate efficiency reduced and time delay increased and more bad effects.

Reference [11] extends the RFC4285 authentication mechanism of Mobile IPv6, it use common AAA authentication platform, give a solution suit layer mobile IPv6 and mobile IPv6 authentication, and it also achieved by software. However, in the solution of preconfigured NAI and the key stored in the file or database as clear text, did not provide data security; authentication option provides data integrity and authentication, but did not provide confidentiality.

In this article it gives a new MIPv6-AAA switch method based on AAA, this solution shows that when mobile node switched in the same AAA administration domain and different sub network, the authentication process do not need though home domain; when mobile node switched in different AAA administration domains, set up MNA to store related information of mobile node MN temporary, in order to make registration and authentication process safety and reliability.

II. RELATED BACKGROUND

A. MIPv6(Mobile IPv6)

Mobile IPv6 is the improved protocol of mobility support for IPv6. The basic aim of its design is let the connection of the transport layer and higher levels not changed with the IP address changes, the mobile node should be always reached by the user [3]. Mobile IPv6 includes three parts: mobile node (MN), home agent (HA)

This project was supported by the Open Foundation of the Key National Defense Science and Technology Laboratory of Education Ministry in JiLin University (No. 421060701421).

and correspondent node (CN). MN has one permit IP address HoA(home of address) in home network. When MN move to foreign network it will have one temporary transfer address CoA(care-of address), after this MN need to complete mobile registration with HA, MN will send binding update (BU) message to tell HA the CoA, and then HA respond to the previous BU though binding acknowledge (BA) message. When CN communicate with MN, because it didn't know MN had moved, so it send data packet to HoA as terminal address, the data package was caught by the HA of MN, and HA transfer the data package to MN of foreign network with tunnel. After MN realize the data come from CN and transferred by HA, it will send BU message to CN and tell the current CoA, after this the rest data packages send to MN will send to CoA directly as terminal address.

B. AAA

Authentication, Authorization and Accounting (AAA) is an important mechanism to ensure security of network and rational use of resources, especially for the Internet provider's point, it is the key point to ensure the normal operation of network. The use of all kinds of resources on network, need to be managed by the AAA. Authentication, authorization and accounting system together to make the network system to accurately recorded the usage of network resource for a particular user. In this way it can effectively safeguarding the rights of legitimate users, and also can protect the operation of network system security and reliable [4]. The AAA architecture was shown in Fig. 1.

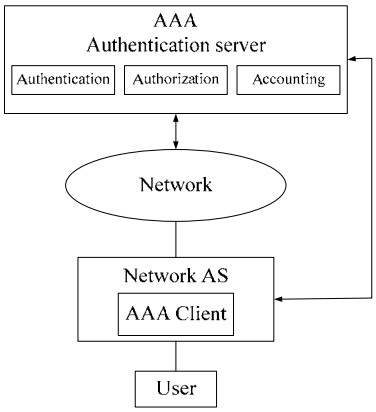


Figure 1. The architecture of AAA

C. The AAA structure under Mobile IPv6 environment

There are two AAA protocols, which are remote authentication dial in user service (RADIUS) protocol and Diameter protocol [5]. Currently, the Diameter protocol was most used. Diameter protocol is a protocol stack [6], which includes the basic protocol and the extend application protocols, such as mobile IP protocol (MIP), Network Access Services protocol (NASREQ), multimedia protocols (IMS), Extensible Authentication Protocol (EAP) and SIP protocols and so on. In the basic protocol, it defines some common functions, such as message format, message transfer mechanism and so on; in the application extension, based on the application detail extend the basic protocol [7]. Basic Diameter protocol must be combined with extend application to use, and provides basic AAA functions for mobile IPv6 in the extend application of mobile IPv6.

Among the Diameter authentications based on MIPv6, in Mobile IP, it includes the mobile node MN, the home agent HA, foreign agent (FA) and other functional entities, except these it also joined the foreign AAA server, home AAA and server AAAF AAAH, the application model shown in Fig. 2 below.

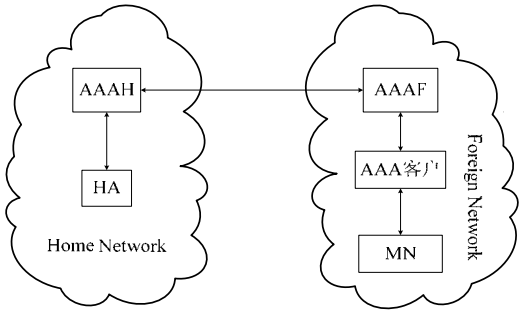


Figure 2. The application model of diameter based on MIPv6

According to the region of AAA server it defined management domain, when the MN switched between different administrative domains of AAA server, it called inter-domain switch; when the MN switched in the same AAA server administrative domain but within different FA subnets, it called inner-domain switch. And also satisfy the following assumptions [7]: (1) the mobile user's identity use NAI [8] (Network Access Identifier) for the only sign, the format of NAI is user@realm, which realm presents the administrative domain where MN located; (2) in long terms between mobile users and AAAH share one key; (3) the communication between AAAF and AAAH is safe; (4) all CN have consensus on the use of public key and symmetric key encryption mechanism.

MIPv6-AAA provides solution for the authentication, authorization, registration and key distribution and other issues of mobile IP, provide a reliable guarantee for large-scale implementation of mobile IP.The authenticate registration process was shown in Fig. 3, and the specific message exchange description see [9].

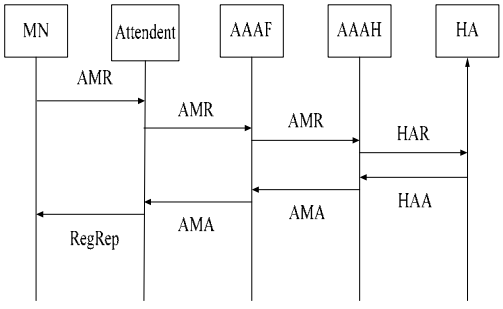


Figure 3. Basic model of authentication and registration process for MIPv6-AAA

Among these, Attendant is the entrance of access domain AAA system, provide and register access domain

address; AMR (AAA mobile node request): mobile node requests; HAR (home agent MIPv6 request): request of MIPv6 home agent; HAA (home agent MIPv6 answer): home agent MIPv6 response; AMA (AAA mobile node answer) mobile node response; RegRep (registration reply): Registration Response.

III. IMPROVED SOLUTIONS

MIP-AAA basic infrastructure provides the integration method of Mobile IP and AAA authentication, but when the mobile node switched in this model, it should complete to register mobile IP, and also should complete the user's authentication and authorization by the AAA. Therefore, MIP-AAA has more delay in switching.

If foreign region is far from home region, the transmission time of transmit message will consume a long time, and the main time delay of authentication process took place on the message exchange between foreign region and home region. One part of the improved solution is the authentication processes do not go though the home region [10] when mobile node switched between different subnets of the same AAA control region. In addition, according to the related data shows that, in the normal movement, 69% of movements occurred in the same region. Therefore, this solution can effectively reduce the switch time delay.

On the other hand, when mobile nodes switched between different AAA administrative domain, MN need to send both authentication request and registration request. In general, the two requests should received responses at the same time. After the process of registration completed, MN can use a new transfer address to receive the data packet transferred by the HA, at this time the MN identity address information has changed, if the process of authentication occur error or delays, it needs to require re-authentication process of MN, and also needs the original MN identity address information, but this time the information has changed, so it can not complete the authentication. And another part of this improved solution is to set a new data structure: the mobile node agent (MNA), use MNA to temporary store related information of the mobile node MN, and then it can guarantee the process of registration and authentication safety and reliable.

A. The MN handoff analysis inner-domain

When switch happened in the inner-domain, the authentication process will no longer go though the home domain. The message flow chart is shown in Fig. 4.

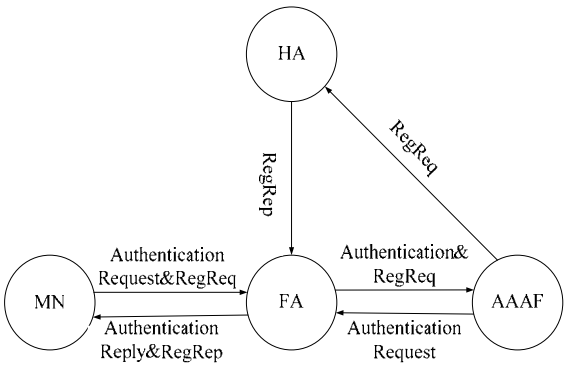


Figure 4. Authentication and handoff process for home network

1) MN→FA: MN sends registration request message and certification message to the FA, and judge whether the switch of MN taken place inside domain or not by the realm value of NAI;

2) FA→AAAF: FA continue to send transfer message 1) to AAAF;

3) After AAAF received message 2), verify the identity of MN, and separate the authentication process and registration process:

a) AAAF→FA: AAAF send authentication responds to FA;

b) AAAF→HA: AAAF send registration request message to HA;

4) FA→MN: FA transfer the authentication responds, then the process of authentication finished. MN can use the resources of foreign network, enjoy the service provide by FA;

5) HA→FA: After HA received the message 3b), it directly give the registration responds message back to FA;

6) FA→MN: FA transfer the registration responds message, then the process of registration finished. MN can use the new transfer address to receive the data packet transferred by HA.

B. MN Handoff Analysis Inter-domain

When MN roaming to a new administrative domain, just AAAH has the full detail information of MN, so the process of switch inter-domain need though home domain, and the registration model shown in Fig. 5:

1) MN→FA: MN send registration request message and authentication message to FA, and judge the realm value of NAI has changed or not. If it changed, start the inter-domain switch;

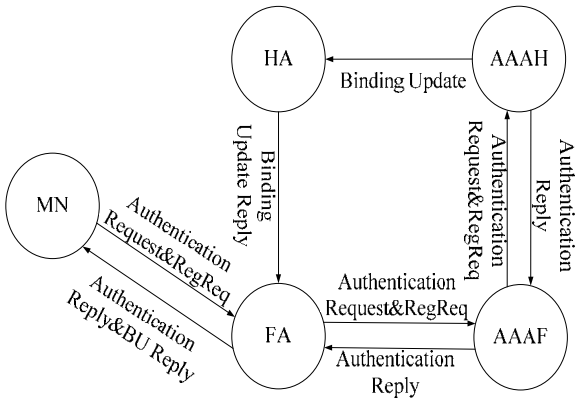


Figure 5. Authentication and handoff process for home network

2) FA→AAAF: After FA received the message 1), it generates the MNA of mobile node, set the legal tag in the original MNA to FALSE, then sends authentication requests to AAAF;

3) AAAF→AAAH: If AAAF can not authenticate, then transfer the message to AAAH;

4) AAAH:

a) AAAH→AAAF： After AAAH successfully authenticate the identity of MN, then send the

authentication responds to AAAF;

b) AAAH→HA: AAAH send binding update message to HA;

5) AAAF→FA: AAAF received the authentication responds message from AAAH, then continue transfer to FA and tell MN authentication has succeed;

6) HA→FA: HA send binding update confirm message to FA;

7) FA→MN: FA sends authentication responds and confirmed binding update to MN, and also set the legal tab on MNA in registry to TURE, notice MN the new mobile node agent MNA have produced;

Till now authentication process and registration process have all finished, MN can use foreign network resources, enjoy FA service, and can use new transfer address to receive the data packet transferred by HA.

IV. CONCLUSION

In this solution, when mobile node switch happened in the inner-domain, because of the authentication process do not go though home domain, after AAAF authenticate the MN identity directly back to FA, reduced the message transmit and process time go and back to home domain, and greatly increase the switch speed; when switch happened in the inter-domain, because set the MNA of mobile node, and make it temporarily keep the original information of MN in case of use, then it can guarantee the switch process safety and reliably.

The next stage is: in this solution it doesn't provide any security for the process of switch inner-domain; when the switch happens in the inter-domain, the data integrity and authentication of the configure of new data structure MNA need to improved, all these are need improved.

REFERENCES

[1] W. S. Xiao, Y. J. Zang, and Z. C. Li, "Hierarchical AAA in mobile IPv6 networks," Journal on Communications, vol. 27, Feb. 2006, pp. 50-55.

[2] R I Chen, R C Wang, and H C Chao, "Mobile IPv6 and AAA architecture based on WLAN[A]," Proc. of the 2004 International Symposium on Applications and the Internet Workshops, 2004.

[3] G. M. Wang, "Security Issues and Solutions on IPv6 Mobile," Journal of University of Electronic Science and Technology of China, vol. 36, Dec. 2007, pp. 1417-1419.

[4] P. Chen, and J. G Yu, "Access authentication in MIPv6 based on hierarchical AAA," Journal of Network Security Technology and Application, May.2009, pp. 32-35.

[5] C R igney, A Rubens, and S W illens, "Remote authentication dial in user service (RADIUS)," Science, RFC 2865, Jun. 2000.

[6] Z. P. Lan, F. L. Jin, and Z. S. Wang, "Study on AAA and security system based on MIPv6," Computer Engineering and Design,vol. 30,Mar. 2009, pp. 3778-3779.

[7] T. Lin, D. Tang, Y. Zhang, H. B. Zhao, and Z. Q. Hou, "Research and implementation of mobile IPv6 fast handoff with AAA functions," Mini-Micro Systems, vol. 26, Jul. 2007, pp. 1125-1129.

[8] A Boba, and Beadles, "The network access identifier," Science, RFC 2486,Jan. 1999.

[9] M Cappiello, AFloris, and L Velt ri, "Mobility amongst heterogeneous networks with AAA support," Proc. IEEE International Conference on Communications, 2002, pp. 2064-2069.

[10] D. Ma, D. K. He, Y. Zheng, and W. F. Zhang, "A fast anthentication and registration scheme for AAA-based Mobile IP," Journal of the China Railway Society, vol. 30, Feb.2008, pp. 98-103.

[11] H. Chen, H. C. Zhou, Y. J. Qin, and S. D. Zhang, "Design and implementation of hierarchical mobile IPv6 authentication based on NAI," Computer Engineering and Applications, vol. 43, 2007, pp. 125-128.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15, August 2010, pp. 409-413

Secure Access Authentication for Media Independent Information Service

Guangsong Li[1, 2], Qi Jiang[1], Xi Chen[1], and Jianfeng Ma[1]
[1]Key Laboratory of Computer Networks and Information Security (Ministry of Education),
Xidian University, Xi'an, Shaanxi, 710071 China
[2]Zhengzhou Information Science and Technology Institute, Zhengzhou, Henan, 450002, China
Email: guangsongli@yahoo.com.cn, jiangqixdu@gmail.com, xichen@gmail.com, jfma@mail.xidian.edu.cn

***Abstract*—To optimize vertical handover in heterogeneous networks, IEEE 802.21 standard defines Media Independent Handover (MIH) services. The MIH services can be a new target to attackers, which will be the main concerns for equipment vendors and service providers. In this paper, we focus specifically on security of Media Independent Information Service (MIIS), and present a new access authentication scheme for MIIS based on Kerberos. Security and performance of the protocol are also analyzed in this paper.**

***Index Terms*—heterogeneous network, media independent information service, handover, authentication**

I. INTRODUCTION

Recent advances in wireless communication technologies have resulted in the evolution of various wireless networks, such as cellular network, wireless local area network, ad hoc network, personal communications networks etc. Communication in next generation networks will use multiple access technologies, creating a heterogeneous network environment [1]. Real-time multimedia services such as voice over IP and interactive streaming becomes more and more popular in current wireless networks, so ubiquitous roaming support for real-time multimedia traffic in an access independent manner becomes increasingly important. Seamless mobility can be achieved by enabling mobile terminals to conduct seamless handovers across diverse access networks, that is, seamlessly transfer and continue their ongoing sessions from one access network to another. Vertical handover in the heterogeneous networks is one of the major challenges for seamless mobility with ubiquitous connectivity, since each access network may have different mobility, quality of service and security requirements [2]. Moreover, real-time applications have stringent performance requirements on end-to-end delay and packet loss.

In order to optimize vertical handover in heterogeneous networks, IEEE 802.21 working group have designed a framework [3] by providing mobile users with information useful for making handover decisions. Examples of the information are the presence of neighboring networks, the type of their links, their characteristics and the services supported. The heart of the framework is the Media Independent Handover Function (MIHF) which provides abstracted services to higher layers and vice versa by means of a unified interface. This is accomplished by defining a set of services, the Media Independent Handover (MIH) services, which consist of media independent information service, media independent event service, and media independent command service. The Media Independent Information Service (MIIS) specifies information about nearby networks useful for handover decisions and the query/response mechanism that allows mobile nodes to get that information. Users get that information from one or more information servers supporting MIH. The Information Server (denoted as IS) may be located in the visited domains or in the users' home domain, i.e., the domain of the service provider that holds information about the users' authentication and authorization profiles.

MIH messages are exchanged over various wireless media between mobile nodes and access networks. Thus the MIH services can be a new target to attackers, which will be the main concerns for equipment vendors and service providers [4]. Some typical threats about MIIS are: identity spoofing, tampering, information disclosure, denial of service, etc. However, security mechanisms are not within the scope of the IEEE 802.21 standard. Security of MIH protocol currently relies on security of underlying transport protocols without a mechanism to authenticate peer MIH entities. Because IEEE 802.21 provides services that affect network resources, cost and user experience, MIH level security will be an important factor to network providers that want to deploy these MIH services in their networks. Nevertheless, there are very few security mechanisms for MIH services in the literature.

IEEE 802.21a task group was set up to address security issues of MIH services. The task of the group is [5]: (i) to reduce the latency during authentication and key establishment for handovers between heterogeneous access networks that support IEEE 802.21; (ii) to provide data integrity, confidentiality, replay protection and data origin authentication to MIH protocol exchanges and enable authorization for MIH services. The technical requirements document [6] of the group describes use scenarios and requirements for security signaling optimization during vertical handover and MIH protocol security. The scope of document [7] is to propose some solutions based on the requirements described in [6].

Won et al. propose a new secure MIH message transport solution, referred as MIHSec [8]. The idea of

MIHSec is to utilize the Master Shared Key (MSK) generated by the L2 authentication procedure, for generating the MIH keys. MIHsec method though has a good performance for MIH message transportation, it introduces other issues. First, it is closely integrated with L2 authentication, thus it is not media independent. Second, the MSK needs to be securely delivered to IS by AR (access router), which means a secure association should be settled a-priori between each AR and IS. So the scheme does not posses scalability. Finally, in MIHsec protocol, the AR that sends MSK to the IS may know the key for MIH messages encryption, which lowers the level of security.

In this paper, a Kerberos based access authentication scheme is proposed for MIIS. The rest of this paper is organized as follows. In Section 2 we present our new approach in detail. Section 3 includes the security and performance analysis. Finally, conclusions are drawn in Section 4.

II. KERBEROS BASED ACCESS AUTHENTICATION FOR MIH INFORMATION SERVICE

The MIIS message exchanges are critical to handover decision phase. The IS needs both to protect itself from attack, and to provide mobile clients provable trust, in order that they can exchange the information securely and make their handover decisions without fear of malicious inaccuracies or mischief. This section focuses on a new proposal for access authentication of MIIS. The scenario we considered is that the access control for information service is applied through an access authentication controller, namely an AS.

A. Network Model

There are some application servers (S1, S2) in core network, which provide application services like, voice over IP, video conference, games, etc. When MN passes the network access authentication, it establishes connection with a Point of Attachment (PoA). MN may request a type of application service through certain PoS. In order to support mobile user to handover seamlessly between heterogeneous networks, an IS is deployed to provide information about neighbor networks. We assume all MNs should register with AS and subscribe some services they needed when network initialization. There exists a security association between any MN and AS, namely a shared key.

Here, MIIS is taken as a service at application layer. It is assumed that MNs have not secure associations with application servers directly. In this scenario of many application servers existing, Kerberos is a natural choice for secure access of services because of its single sign on characteristic. Additionally, to avoid multiple and repeated authentications of MNs, it is proposed that the function of Kerberos AS be implemented on the network access AS. Suppose AS and TGS have already built a secure association, i.e. they share a key to protect communications between them. We also assume that all application servers, (S1, S2, and so on, including IS) have shared some keys with the TGS respectively. For example, there is a long term key $k_{TGS\text{-}IS}$ shared between IS and TGS for secure connection or authentication. Suppose *prf* ( ) is a secure key derivation function, and *h*( ) is a secure hash function.

B. MIIS Access Authentication Based on Kerberos

Since MNs have not secure associations with application servers (including IS), the access control of application services is applied through AS.

When a MN wants to access the network, a media specific mutual authentication procedure (e.g., 802.11i, 802.16e authentication mechanism) must be executed between the MN and AS. If MN passes the procedure successfully, a shared Master Key (MK) and some session related keys are set up between MN and AS, and MN is permitted to connect with the network. The MK will be saved by both of MN and AS for later use. In our protocol, we will exploit MK for Kerberos authentication. Let k_{tra} = *prf* (MK, TGS, MN, "TICKET.REQUEST. AUTHENTICATION"), where TGS, MN are the identifiers of TGS and MN respectively, and the last parameter of *prf* ( ) is key description. AS computes k_{tra} after MN passes the network access authentication, and it delivers identifier of MN, services list MN subscribed, and k_{tra} to TGS securely. k_{tra} will be used for authentication of MN when requiring for service tickets. TGS sets up an entry for MN, which consists of identifier of MN, the key k_{tra}, a field of sequence number, and the service list of MN. The sequence number is initialized to 0 and it is strictly monotone increasing to ensure freshness of request message of MN.

After MN connects with the network, it should contact IS to get information about neighbor networks. To this end, MN must obtain IS service ticket. Then mutual authentication is performed between MN and IS using the service ticket. The flow chart of our protocol is as follows, in which flow (1) and (2) describe service ticket request and response, flow (3) to (5) describe mutual authentication between MN and IS, and flow (6) denotes the protected MIIS response message.

(1) IS service ticket request (MN→TGS)

MN computes the ticket request authentication key k_{tra} in the same way as AS. Then MN sends a service ticket request message (ST_REQ) to TGS for IS. The message content of ST_REQ is as the following, {STReq, MN, TGS, IS, n_M, MAC_M}, where STReq denotes identifier of the request, MN and IS denote identifiers of the mobile node and information server respectively, and n_M is the sequence number generated by MN, MAC_M is a message authentication code derived from the equation MAC_M= $h(k_{tra}$, STReq, MN, IS, $n_M)$.

(2) IS service ticket response (TGS→MN)

Upon receiving the ST_REQ message from MN, TGS extracts MN, n_M from the message, and finds the item related to MN in its database, namely the entry (MN, k_{tra}, sequence number, service list). If n_M is equal to or smaller than the sequence number in the entry, the request message is dropped because of staleness. Otherwise, TGS computes the value $h(k_{tra}$, STReq, MN, IS, $n_M)$ using k_{tra}. If the value matches with MAC_M, TGS believes the message is really originated from MN. Then TGS checks

the service list of MN to find whether it has subscribed service of IS. If MN has not subscribed the service of IS, TGS will respond a reject message to MN. Otherwise, a service ticket ST will be generated for MN.TGS chooses a random number as the service authentication key $k_{MN\text{-}IS}$ for mutual authentication between MN and IS. And TGS also derives a key encryption key $k_s=prf(k_{tra}$, MN, TGS, n_M , "KEY.ENCRYPTION.KEY"), which will be used for secure transmission of $k_{MN\text{-}IS}$ to MN. The service ticket is as follows: ST={MN, IS, [MN, IS, $k_{MN\text{-}IS}]k_{TGS\text{-}IS}$},where [MN, IS, $k_{MN\_IS}]k_{TGS\text{-}IS}$ denotes cipertext encrypted with the key $k_{TGS\text{-}IS}$ shared between TGS and IS.

TGS needs to generate a service ticket response (ST_RES) message. The ST_RES message consists of the following items, {STRes, MN, TGS, n_M, ST, [MN, IS, $k_{MN\_IS}]k_s$, MAC_T}, where STRes denotes identifier of the response, MN and IS denote corresponding identifiers, and n_M is the sequence number generated by MN, MAC_T is a message authentication code derived from the equation $MAC_T=h(k_{tra}$, STRes, MN, TGS, n_M, ST, [MN, IS, $k_{MN\_IS}]k_s$). ST_RES message is transmitted to MN by TGS. And the sequence number related to MN in database of TGS is updated to n_M.

(3) IS service request (MN→IS)

When MN receives the ST_RES message from TGS, MN first check the sequence number to see if it is the number MN just sent. Then it calculates the value $h(k_{tra}$, STRes, MN, TGS, n_M, ST, [MN, IS, $k_{MN\_IS}]k_s$), and compares the value with MAC_T in the ST_RES message. If the two values are identical, MN believes the message is generated by TGS. MN computes the key k_s using $prf(k_{tra}$, MN, TGS, n_M), and decrypts the ciphertext [MN, IS, $k_{MN\_IS}]k_s$ to get the key $k_{MN\text{-}IS}$.

Now MN is able to contact with IS for MIIS. As for the authentication of this phase, we use a challenge-response mechanism to avoid maintaining records and states about MNs on IS. MN needs to send an information service request message (IS_REQ) to IS. The message formats of IS_REQ is as the following, {ISReq, MN, IS, ST, r_M}, where ISReq denotes identifier of the request, MN and IS denote the corresponding identifiers respectively, ST is the service ticket generated by TGS, and r_M is a random number chosen by MN.

(4) IS challenge (IS→MN)

After IS receives the IS_REQ message, it decrypts ST using the key $k_{TGS\text{-}IS}$ shared with TGS to obtain the service authentication key $k_{MN\text{-}IS}$. It also gets the identifiers in the service ticket to determine whether the ticket is for MN and IS. Then IS chooses a random number r_I as a challenge number. IS also needs to generate a message authentication code $MAC_I = h(k_{MN\text{-}IS}$, MN, IS, r_M, r_I) and transmits r_I, MAC_I to MN.

(5) MN response (MN→IS)

When MN receives the message from IS, it computes $h(k_{MN\text{-}IS}$, MN, IS, r_M, r_I) and compares the value to MAC_I. If the two values match with each other, MN is confirmed that the peer is the exact information server. MN then derives two service session keys, *isk* for integrity and *esk* for encryption, from two equations $prf(k_{MN\text{-}IS}$, r_M, r_I, "SERVICE.SESSION.KEY.INTEGRITY") and $prf(k_{MN\text{-}IS}$, r_M, r_I, "SERVICE.SESSION.KEY.ENCRYPTION") respectively. MN has to generate a response res_M about r_I using the equation $res_M=h(k_{MN\text{-}IS}$, MN, IS, r_I, $r_M)$. res_M is transmitted to IS as a response to the challenge r_I.

(6) Protected IS service response

Upon receiving the message, IS computes $h(k_{MN\text{-}IS}$, MN, IS, r_I, $r_M)$, and compares it to the value of res_M. If the two values are identical, IS believes the requestor is a valid client. IS then computes service session key *isk* and *esk* as MN does. Then IS can provide MN with the information it needs protected by the service session key.

For accessing services other than the MIH information service, the user needs to obtain the corresponding service ticket from TGS. The user then sends a request message directly to the application server which implements the authentication process as depicted above. Based on the user credentials, the application server authenticates the user, which means that it checks user's service ticket and decide whether to grant access or not according to the authentication phase. The application service and the user can use the service ticket, for setting up IPSec security at IP level, or simply use the shared secret key resulting upon successful authentication, to perform symmetric-cryptography based security at application level.

III. Security and Performance Analysis

A. Security Analysis

We assume that the cryptography suites employed in our protocol are all secure, such as pseudo-random number generator, key deriving function, message authentication code and symmetric key encryption algorithm. Based on the above assumptions, we analyze the security of the proposed scheme. Our access authentication protocol satisfies the following security properties.

Mutual Authentication of Peers

The mutual authentication between TGS and MN is ensured via the mechanism of strictly increasing sequence number and message authentication code. Each time MN requiring a service ticket from TGS, MN generates a new sequence number larger than before, and computes a MAC value of the ST_REQ message using k_{tra} shared with TGS. TGS will check the sequence number in ST_REQ to identify the freshness of the message, and it will also verify the MAC value to check the message integrity. In this phase, all stale and forged messages will be thrown away. If MN passes the procedure, TGS sends to MN a ST_RES message with a MAC value derived from k_{tra}, the new sequence number and other items. The MAC value can be used to verify origin of the ST_RES message. In addition, the sequence number field will be updated after the ST_REQ message is verified successfully.

When MN requests for MIIS, it uses the service authentication key $k_{MN\text{-}IS}$ in service ticket to perform mutual authentication with IS. In this phase, challenge-response and MAC mechanisms are utilized. In the light

of our protocol, IS is able to decrypt service ticket to get $k_{MN\text{-}IS}$ using the long term shared key with TGS. Both MN and IS compute responses (MACs) to the challenge number of the opposite peer using $k_{MN\text{-}IS}$.

Key Freshness and Control

The service session key *isk* and *esk* are computed from a function of $k_{MN\text{-}IS}$, r_M, and r_I, where r_M and r_I originating from the MN and IS respectively. The freshness of random numbers r_M and r_I evidently assures the freshness of service session key. It also can be seen that MN and IS can not independently control the choice of the session key. No one else can impersonate MN or IS to generate r_M and r_I for mutual authentication between them. Therefore, the service session key *isk* and *esk* are fresh, random and independent.

Forward Secrecy

The random nonce is unpredictable for any party except MN or IS. Even if the intruder attacks secret information MN and IS, he can not compromise the past random numbers and the past session keys. Hence the scheme has the property of perfect forward secrecy.

Known Key Security

Since each run of authentication protocol produces different random numbers r_M and r_I which are used to set up service session keys, the non-correlation of random numbers ensures the intruder can not obtain any current session key even if the past session keys are exposed.

Resistance to Replay Attack

Replay attack involves the passive capture of data and its subsequent retransmission to produce an unauthorized effect. An intruder records message flows and then it can retransmit an old request message to trick TGS or IS for false authentication. This replay attack can be prevented as follows. At the phase of service ticket request, the strict increasing sequence number is used to identify a stale request message to prevent replay attack. At the phase of information service request, challenge-response mechanism makes the messages be fresh and unpredictability. Replay attack can be easily detected in this phase.

Resistance to Man in the Middle Attack

A man in the middle attack is that an attacker is able to read, insert, and modify messages between two parties without either party knowing that the link between them has been compromised. Attacker, as a middle-man between the mobile user and IS, cannot derive the correct $k_{MN\text{-}IS}$. Thus the malicious middle-man cannot establish the secure association on behalf of the legitimate MN or IS. Notice that the MAC values are sent in messages flow (3) and flow (5), which guarantees that the identifiers of MN and IS, and the random numbers are bounded together for a particular session. In other words, the attackers cannot pick MN and IS individually from different authentication sessions and combine them to construct a valid message in one authentication session.

All the characters described above have ensured that our protocol satisfies the basic security requirement and is robust to the existing attacks.

B. Performance Analysis

Protocol performance has become an increasingly important topic in wireless computing and networking environments. It is always desirable to make an authentication protocol more efficient. Our protocol may be quite efficient, since it requires only symmetric key operations and a few rounds of message exchanges during access authentication process. We list type and number of computation operations required by MN, TGS and IS in Table 1. The computational cost of our protocol is very reasonable, especially for the mobile node. The computation operations in our protocol are negligible compared to any strong public-key authentication.

In the proposals of 802.21a task group [7], EAP framework is suggested to fulfill mutual authentication between peers for centralized MIH service. EAP-TLS is a typical and widely applied authentication protocol in EAP protocol family [9]. We take it as an example for analysis. The computation operations required by MN, Pos of IS, and AS are given in table I. Compared with our scheme, that method is a rather complex and high-cost process using public key certificates. It adds too much load to mobile nodes (costing much time and energy).

As to communication performance, we can see that in the first phase (service ticket request), only a 2-way handshake is implemented between MN and TGS. It fulfils tasks of data origin authentication and service ticket distribution, since it utilizes a strict increasing sequence number mechanism. In the second phase (information service request), mutual authentication between MN and IS are carried out through a 3-way handshake procedure, meanwhile session key agreement is completed. Nevertheless in 802.21a proposals, a full EAP-TLS procedure requires 8 message flows between MN and AS for their mutual authentication, afterwards it has to perform mutual authentication between PoS of IS and MN (at least 3 message flows). The whole process needs so many message flows that it costs too much bandwidth and time. Thus our protocol performs better than proposal of 802.21a task group in communication performance.

TABLE I.

COMPUTATION OPERATIONS IN 802.21A AND OUR SCHEME

Computation Operations	802.21a	Ours
Sequence number (MN/IS/ASorTGS)	0/0/0	1/0/1
Random number (MN/IS/ASorTGS)	1/1/0	1/1/1
MAC generation (MN/IS/ASorTGS)	1/0/1	4/2/2
DH operation (MN/IS/ASorTGS)	2/2/0	0/0/0
Key derivation (MN/IS/ASorTGS)	2/1/1	3/2/1
Symmetric encryption (MN/IS/ASorTGS)	0/0/1	0/0/2
Symmetric decryption (MN/IS/ASorTGS)	0/1/0	1/1/0
Certificate validation (MN/IS/ASorTGS)	1/0/1	0/0/0

IV. CONCLUSIONS

The IEEE 802.21 standard aims at optimizing handovers among heterogeneous wireless networks. In this paper, we propose a protocol for access authentication of MIIS defined in the 802.21 standard. We adopt a modified version of Kerberos featuring of sequence-number based service ticket distribution and challenge-response based service access authentication. It performs Kerberos authentication without relying on time synchronization, which makes authentication between MN and IS more freely. The security and performance analysis show that the proposed scheme has excellent performance. The new solution has the advantages of lightweight computation, less communication cost, and easy implementation. In fact, our work can be applied to offer integrated authentication and authorization functionalities for any type of application service.

REFERENCES

[1] N. Nasser, A. Hasswa, and H. Hassanein, "Handoffs in Fourth Generation Heterogeneous Networks," IEEE Commun. Mag., vol. 44, No. 10, pp. 96-103, 2006.

[2] G. Karopoulos, G. Kambourakis, and S. Gritzalis, "Survey of Secure Handoff Optimization Schemes for Multimedia Services over All-IP Wireless Heterogeneous Networks," IEEE Communications Surveys & Tutorials, vol. 9, No. 3, pp. 18-28, 2007.

[3] IEEE 802.21 standard, "IEEE Standard for Local and Metropolitan Area Networks-Part 21: Media Independent Handover Services," January 2009.

[4] Y. Ohba, "Five Criteria for Security Extensions to Media Independent Handover Services," http://www.ieee802.org/21/802_21a_5C.pdf, accessed June 2010.

[5] 802.21a PAR, "Amendment for Security Extensions to Media Independent Handover Services and Protocol," http://www.ieee802.org/21/802_21a_Par.pdf,accessed June 2010.

[6] S. Das, M. Meylemans, and Y. Ohba, et al. "IEEE 802.21 Security SG Technical Report," https://mentor.ieee.org/802.21/documents, accessed June 2010.

[7] S. Das, A. Dutta, and T. Kodama, "Proactive Authentication and MIH Security," https://mentor.ieee.org/802.21/documents, accessed June 2010.

[8] J. Won, M. Vadapalli, C. Cho, and V. Leung, "Secure Media Independent Handover Message Transport in Heterogeneous Networks," EURASIP Journal on Wireless Communications and Networking, http://www.hindawi.com/journals/wcn/2009/716480.html, accessed June 2010.

[9] RFC 2716, "PPP EAP TLS Authentication Protocol," October, 1999.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15, August 2010, pp. 414-416

Application of Sequence Alignment Method to Product Assortment and Shelf Space Allocation

Peiqian Liu, Hairu Guo, Weipeng An
School of Computer Science and Technology/Henan Polytechnic University, Jiaozuo, China
liupeiqian@hpu.edu.cn

***Abstract*— In retailing, decisions about product assortment and shelf space allocation have a significant effect on customers' purchasing decisions. Traditionally, researchers usually employed the space elasticity to optimize product assortment and space allocation models. However, the large number of parameters requiring to estimate and the non-linear nature of space elasticity can reduce the efficacy of the space elasticity based models. Instead of space elasticity, this paper utilizes a data mining approach, Sequence Alignment Method (SAM), to resolve the product assortment and allocation problems in retailing. In our approach, the SAM is applied to explore the relationships between product categories and is compared to association rule mining. Experimental results show that SAM achieves better quality than that of association rule mining and can generate very useful information to shelf space management.**

***Index Terms*—Data mining; Shelf space management; SAM; Product taxonomy**

I. INTRODUCTION

Shelf space is an important resource for retail stores since a great quantity of products compete the limited shelf space for display. Retailers need frequently make decisions about which products to display (assortment) and how much shelf space to allocate these products (allocation)[1]. Product assortment and shelf space allocation are two important issues in retailing which can affect the customers' purchasing decisions. Through the proficient shelf space management, retailers can improve return on inventory and consumer's satisfaction, and therefore increase sales and margin profit [2].

Traditionally, researchers apply the space elasticities to determine which products to stock and how much shelf space to display these products. However, there are two major limitations that reduce the effectiveness of the space elasticity [3]. First, due to the non-linear nature of space elasticity, the space elasticity based models are very complicated, and the specific solution approach is developed for each model. Additionally, it is necessary to estimate a large number of parameters by using the space elasticity.

With the rapid development of information technology, transaction data can be easily collected through the point of sale (POS) system. The relationships between products hidden in transaction data can be discovered through data mining to assist product assortment and shelf space allocation [4]. It is not necessary to conduct a series of experiments to estimate a great quantity of parameters in space elasticities.

In this paper, we propose a data mining approach, called Sequence Alignment Method (SAM), to make decisions about which products to stock, how much shelf space allocated to the stocked products and where to display them. SAM analyzes customers' shopping behaviors from transaction data to obtain relationships between product categories. As a dimensionality reduction technique, we employ a product taxonomy. In this taxonomy, similar products are identified and grouped together using the product taxonomy so as to build the customer profiles and to search for the neighbors in the reduced dimensional space. The products and categories frequently bought together can be displayed together. Finally, experiment shows SAM is better quality than association rule mining.

II. METHODOLOGY

The proposed procedure of shelf space management is divided into three phases: product classification, neighborhood formation, and top-N products generation.

A. product classification

In this phase, the marketing manager or domain expert categorizes all the products by specifying the level of product aggregation on the product taxonomy. A product taxonomy is practically represented as a tree that classifies a set of products at a low level into a more general product at a higher level. The leaves of the tree denote the product instances, Stock Keeping Units (SKUs) in retail jargon, and non-leaf nodes denote product classes obtained by combining several nodes at a lower level into one parent node. The root node labeled by All denotes the most general product class. Fig.1 shows an example of product classification, where class nodes are denoted by shaded regions.

In this example, class(SKU00) = Outerwear and class(SKU10) = Shoes, etc.

Recent data mining research has shown that data mining algorithms usually produce the best results when product-related transactions are evenly occurred [5].

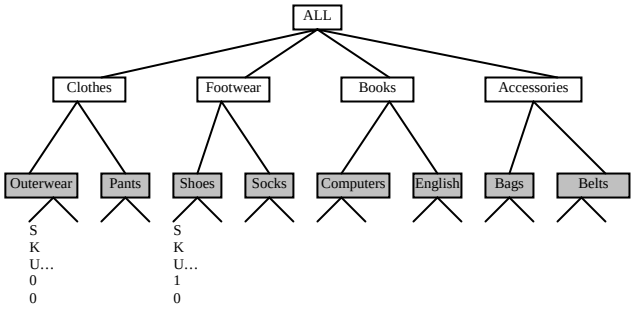


Figure 1. Example of product classification.

B. *neighborhood formation*

This phase performs computing the similarity between customers and, based on that, forming a neighborhood between a target customer and a number of like-minded customers.

A sequence is a number of elements arranged or coming one after the other in succession. In general, the distance (or similarity) between sequences is reflected by the number of operations necessary to convert one sequence into the other. As a result, SAM distance measure is represented by a score. The higher/lower the score, the more/less effort it takes to equalize the sequences and the less/more similar sequences are. In addition, SAM scores for the Insertion and the deletion operations to unique elements of source (first) and target (second) sequences during equalization process, not scores the reordering operations to common elements. Common elements appear in both compared sequences whereas unique elements appear in either one of them. Finally, SAM represents the minimum cost for equalizing two sequences.

In this paper, SAM distance measure between two sequences S_1 and S_2 is calculated using the following formula :

$$d_{SAM}(S_1, S_2) = \sum_{i=1}^{n} w_i \left| A_{i1} - A_{i2} \right| \qquad (1)$$

where d_{SAM} is the distance between two sequences S_1 and S_2; n is the length of S_1 or S_2 after equalization; w_i the weight value for the deletion operations or insertion operations on the ith element, a positive constant not equal to 0, determined by the researcher ($w_i > 0$); A_{ij} is the shopping sum on the ith element in S_j(j=1,2).

To illustrate SAM, consider the following sequences S_1 (source sequence) and S_2 (target sequence). Both sequences represent a purchased list extracted from the transaction database. Each element in the list is represented by a pair of number (C_i, A_i), where C_i represents the class ID, and A_i the shopping sum for C_i.

S_1: {(1,4), (4,1), (7,9), (8,9)}

S_2: {(1,1), (2,4), (3,1), (4,9), (8,9), (7,4)}

First, the maximum number of similar elements having the same class ID is defined. Then, in order to equalize S_1 with S_2; unique elements (2,0) and (3,0) are inserted into S_1 which gives the following sequences (a 0 means class 2 or 3 are not purchased):

S_1: {(1,4), (2,0), (3,0), (4,1), (7,9), (8,9)}

S_2: {(1,1), (2,4), (3,1), (4,9), (8,9), (7,4)}

The equalization process continues with reordering common class 7 (or 8) in S_1 or S_2:

S_1: {(1,4), (2,0), (3,0), (4,1), (7,9), (8,9)}

S_2: {(1,1), (2,4), (3,1), (4,9), (7,4), (8,9)}

Finally, equalizing s1 with s2 took 2 insertion, (If we assign 2 to w_i for insertion and deletion, and 1 to w_i for otherwise) which gives us:

$d_{SAM}(S_1, S_2) = (4-1)+2\times(4-0)+ 2\times(1-0)+(9-1) +(9-4)+(9-9)=26$.

After pair wise distances between sequences are calculated using SAM, A distance matrix is build for holding distance scores between each sequence pair on the diagonal. Because this study is focused on SAM, no special attention is paid to the clustering method. Therefore, a simple hierarchical clustering algorithm like Ward [6] is used to form neighborhood. In order to define an optimal solution for the number of neighborhood, r^2 index is used. r^2 is one of the most commonly used stop criteria [7], equals the proportion of variation and ranges in value from 0 to 1.

C. *Top-N products generation*

The final phase is to ultimately derive the top-N products from the neighborhood of customers. For each neighborhood, we produce a list of N products that the neighborhood is most likely to purchase. In this paper, we adopt the highest likely-to-buy rate (HLR) for generating a product list for a given neighborhood. The HLR method chooses products with the highest likely-to-buy rate of all neighbors. Formally, the likely-to-buy rate of the neighborhood i for a product j $LTB_{i,j}$is defined below:

$$LTB_{i,j} = \frac{\sum_{j \in neighborood(i)} r^{ij}}{\sum r^{i}} \qquad (2)$$

Where r_{ij} is the total number of occurrences of purchases of a neighborhood i for a product j; and r_i the total number of occurrences of all purchased of a neighborhood i.

III. EXPERIMENTAL EVALUATION

A. *Experimental result*

The proposed data mining based procedure for product assortment and allocation is implemented with an example of a retail store. The database includes product data, customer data and transaction records. There are 3060 product items, which are divided into 32 categories. 10 percent of the transaction records was set as training data and 90 percent was set as test data. For neighborhood formation, r^2 index is used based on SAM distance matrix. In our research, 6 neighborhoods are formed with r^2 equal to 0.64.

Finally, top-N product list are generated for each neighborhood;

$N_1 = \{C_{11}, C_{13}, C_{16}, C_{18}, C_{20}, C_{27}, C_{30}\}$

$N_2 = \{C_2 , C_5 , C_9 , C_{14}, C_{15}, C_{28}, C_{29}\}$

$N_3 = \{C_1 , C_4 , C_8 , C_{19}, C_{21}, C_{24}, C_{26}\}$

$N_4 = \{C_6 , C_{12} , C_{17}, C_{22}, C_{23}, C_{25}, C_{31}\}$

$N_5 = \{C_3 , C_7 , C_{10}, C_{11}, C_{18}, C_{32}, C_{21}\}$

Where N_i denotes neighborhood i, and C_i represents the class ID of products. So classes in the same N_i should be displayed as near as possible in a store.

B. Evaluation metrics

To evaluate the quality of the method, recall and precision have been widely used in relative research. Another widely used combination metric called F1 metric [8] that gives equal weight to both recall and precision was employed for our evaluation. It is computed as follows:

$$F1 = \frac{2 \times recall \times precision}{recall + precision} \qquad (3)$$

Finally, we compared the quality of SAM with that of association rule mining. Fig.2 shows our experimental results. It can be observed from Fig.2 that SAM works better than the association rule mining, achieving an average improvement of 38%.

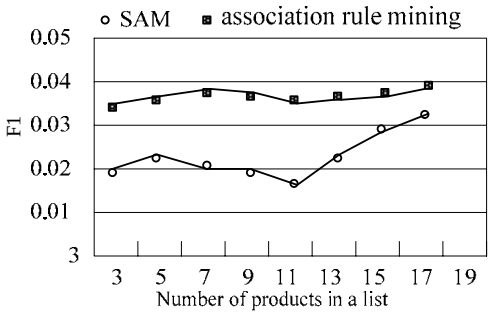


Figure 2. Quality comparison of SAM and association rule miming

IV. CONCLUSION

With the rapid development of information technology, retailers have put a huge amount of transaction data in storage, and they potentially can be used to support shelf space management. This paper develops a data mining based approach SAM to simultaneously make decisions about Product Assortment and shelf space allocation. Firstly, the marketing manager categorizes all the products on the product taxonomy. Secondly, SAM is used to compute the similarity between customers and, based on that, forming a neighborhood between a number of like-minded customers. Finally, top-N product list are generated for each neighborhood. The top-N products can be displayed as near as possible in a store. Experimental results shows that SAM works better than the association rule mining.

REFERENCES

[1] Borin, N. and Farris, P.W, A sensitivity analysis of retailer shelf management models. Journal of Retailing, 1995, 71(2), pp.153–171.

[2] Yang, M-H., and Chen, W.-C., A study on shelf space allocation and management, International Journal of Production Economics, 1999, pp 60–61, 309–317.

[3] Hwang, H., and Choi, B., A model for shelf space allocation and inventory control considering location and inventory level effects on demand, International Journal of Production Economics, 2005,.97(2), pp.185–195.

[4] Mu-Chen Chen and Chia-Ping Lin, A data mining approach to product assortment and shelf space allocation , Expert Systems with Applications, 2007, pp.976–986

[5] Berry,J.A.,and Linoff, G., Data mining techniques: For marketing, sales, and customer support, 1997, New York: Wiley.

[6] Hair, J.F., Andersen, and R.E., Multivariate Data Analysis. Prentice Hall, 1998, New Jersey

[7] Mobasher, B. and Dai, H., Discovery of aggregate usage profiles for webpersona lization, 2006, WebKDD Workshop, Boston, pp.120-126.

[8] Sarwar, B., Karypis, G., Konstan, J. A., & Riedl, J. (2001). Item-based collaborative filtering recommendation algorithm, Proceedings of the 10th International World Wide Web Conference, pp.285–295.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15, August 2010, pp. 417-421

Research on Assembly and Fault-tolerant of Interface Component in Distributed Human-computer Interactive System

Ming-chuan Zhang, Hong-yi Wang, Shi-bao Sun, Qingtao Wu, Guan-feng Li
College of Electronic and Information Engineering Henan University of Science and Technology, Luoyang, China
zhlzmc@163.com, haustwhy@gmail.com, sunshibao@126.com, wqt8921@126.com, haustlgf@gmail.com

***Abstract*—In order to realize the human-computer interactive interface fault-tolerant, dynamic reconfiguration and uniform control of interface switching of each human-computer interactive node in distributed system, a rapid developing model is proposed to build human-computer interactive system recurring to an assistant developing tool designed self-owned. Subsequently, a fault-tolerant scheduling model is presented as well as corresponding scheduling algorithms based on the rapid developing model. This model is made up of Center Scheduling Component and many human-computer interactive nodes. Each human-computer interactive node includes a Local Scheduling Component, an Interface Generating Component and some human-computer interactive interface components. Finally, feasibility of the model and correctness of the algorithms are proved by the simulation experiment.**

***Index Terms*—distributed system, interface component, human-computer interactive system, component assembly, scheduling model**

I. INTRODUCTION

With the development of computer technology, users have higher expectations for computer system, especially to the human-computer interactive system (HCIS). Not only do users require that the human-computer interactive interfaces (HCIIs) are beautiful, easy to use, response sensitive, but also require HCIS possess reliability, ability to resist trouble and dynamic reconfiguration etc. The main purposes this paper studies are to solve three problems in distributed system, which are the fault-tolerant (FT), the dynamic reconfiguration and the uniform control of the human-computer interactive interface components.

Recent years, the dynamic reconfiguration and FT problems are paid attention to by more and more researchers in distributed system. Remote dynamic component configuration is discussed in reference, which greatly improves system flexibility using configuration files [1]. The dynamic deployment and re-configuration of pervasive service components in a self-controlled manner are researched. In particular, a service component self-deployment algorithm using partitioning techniques and a simple service re-configuration algorithm are proposed and evaluated. The effectiveness of the proposed mechanisms is proved by the experiment results [2]. A new method of QoS-aware and dynamic configuration for Web services composition is presented to improve the adaptive capacity to both the QoS variability of component services and the failure-prone environment [3]. There are two topics which are researched in reference [4]. First, it describes optimizations applied to an implementation of the OMG's Deployment and Configuration of Components specification that enable performance trade-offs between QoS aspects of DRE systems. Second, it compares the performance of several dynamic and static configuration mechanisms to help guide the selection of suitable configuration mechanisms based on specific DRE system requirements. Two methods of dynamic reconfiguration are introduced. First, using configuration file, this method belongs to static configuration ways. Second, utilizing configuration operation in the program, this method belongs to dynamic reconfiguration ways, which can adapt to some configuration situations that can't be estimated in advance. The software architecture supporting dynamic reconfiguration is studied in reference [5]. It is solved with the graph-oriented programming method, which realizes dynamic reconfiguration and the description of software architecture based on components in the uniform way. Certain problems of components dynamic reconfiguration are researched in references [3-5], and some achievements are got. However, there are some difficulties in practical application. Furthermore, systemic fault-tolerant has not been considered.

In references [6-8], FT and dynamic reconfiguration are studied in distributed system. Among them, the FT and dynamic reconfiguration are realized in different technology. However, the FT and dynamic reconfiguration have not been considered in the high level, as well as uniform control of distributed system. Component frameworks provide the strategy for the development and deployment of complex multiphysics applications to satisfy the need [9]. In order to build dynamically adaptable applications, the service-oriented component models supporting the dynamic availability of components at run-time are researched as well as offering the possibility in reference [10].

This paper considers dynamic reconfiguration and FT of human-computer interactive interface components, as well as uniform control of interface switching of each node synthetically. A rapid developing model, a fault-

The project numbers are NSF of China No.70671035 and PKSTP of Henan Province China No.082102210076.

Ming-chuan Zhang is with the Electronic & Information Engineering College, Henan University of Science and Technology, Luoyang 471003, China (e-mail:zhlzmc@163.com)

tolerant scheduling model and corresponding scheduling algorithms are presented to resolve the problems.

II. THE ASSEMBLY MODEL OF INTERFACE COMPONENT

There are two kinds of components in the human-computer interactive system. One is framework component. The other is business component (interface component). There is at least a component that is called main framework for one application. The main framework denotes the business logic relation of the application. The business components and framework components are assembled by the main framework to realize the application business logic. The assembly architecture of component and framework is shown as Figure 1. Because the main framework is the first entity that is assembled and the function of other components is to cooperate with the main framework, its actual function is a component container. The component assembly architecture is shown as Figure 2.

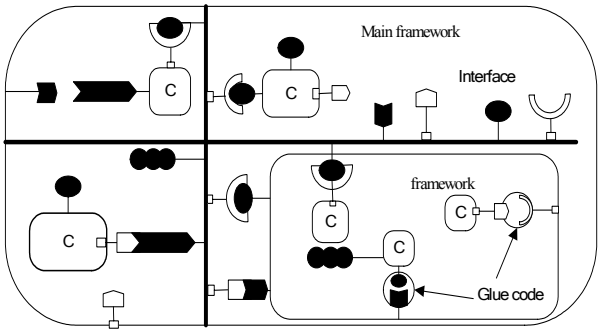


Figure 1. the assembly of component and framework

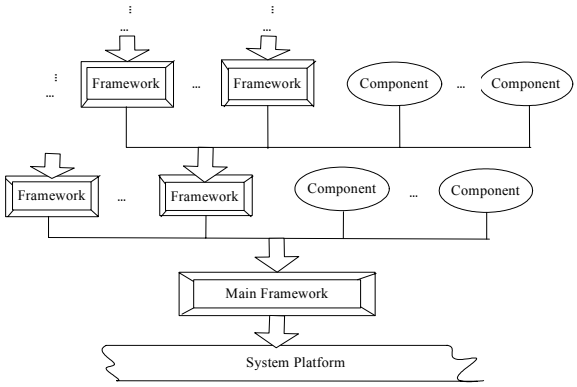


Figure 2. the architecture based on component assembly

The assembly characteristic of the business component and framework component are illuminated as follows.

(1) Business components can only be assembled to framework components. Business components can not assemble business components.

(2) Framework components can be assembled to framework components. That is to say, simple business logic combines with other business logic to gain complicated business logic.

(3) The workflow of application is determined by framework assembly architecture. The main framework calls framework component and business component. The higher framework calls lower framework and business component. The system function is realized by iterative. The calling flow can be reversed.

(4) The converter, mapping and glue for no matching interfaces are done by framework component.

III. THE RAPID DEVELOPING MODEL OF HCIS

The human-computer interactive system can be developed rapidly recurring to an assistant developing tool which is similar to the software of drawing and designed self-owned. Each component in the component library has its style which can be used directly to build the interface style of the target system. For the components nonexistent in component library, the interface style can be designed by the assistant developing tool.

When designing a human-computer interactive system, the style of human-computer interface is gained recurring to the assistant developing tool, as well as the designing result file. The target system can be build based on the file. The rapid developing model is shown as Figure 3 and the developing steps are shown as follows.

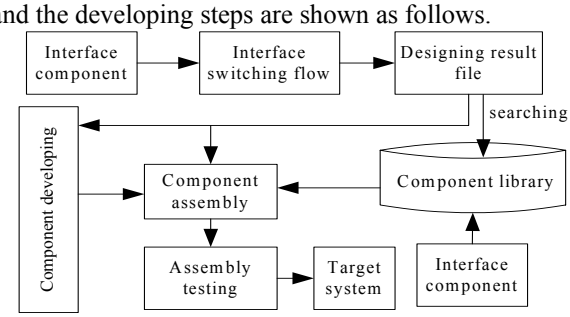


Figure 3. the rapid developing model of HCIS

(1) Determining components: abstracting the requirement and determining system framework model to gain the components which build the target system based on the requirement analysis.

(2) Determining the flow of interface switching: analyzing application workflow to gain the flow of interface switching.

(3) Gain designing result: generating interface designing result file which includes the ID of framework component, the ID of business component, the ID of the component needing developing and interface switching information based on the step one and step two.

(4) Component developing: developing the new components whose ID is generated in step three and putting them into the component library.

(5) Component assembly: assembling the various components to gain target system according to the designing result file which is generated in step three.

(6) Assembly testing: testing the target system, if the target system satisfies the need, the designing process is accomplished. If not, the process will be iterative

IV. A FAULT- TOLERANT SCHEDULING MODEL

Several definitions used in the fault-tolerant scheduling model and scheduling theorem are introduced.

Definition 1: An Interface Scheduling Task (IST) refers to a kind of prearrange scheme to the ICs

scheduling. An Interface Scheduling Request (ISR) refers to an applying command which makes the IST begin to carry out. IST is a static concept, and ISR is a dynamic concept.

Definition 2: The Interface Scheduling Task Set (ISTS) is a 2-tuple, ISTS = {IST_{nft}, IST_{ft}}. Among them, IST_{nft} means the tasks set without FT requirements. If the scheduling of ISR of IST_{nft} fails, nothing will be done. IST_{ft} means the tasks set with FT requirements. If the scheduling of ISR of IST_{ft} fails, the ISR will be executed repeatedly by the system.

Definition 3: The Set IST_{nft} is a 4-tuple, IST_{nft} = <SN, DN, CS, P>. Among them, SN refers the node which sends ISR. DN refers the node which ICs scheduling happens in. CS is the lists of ICs which are needed showing. P refers the priority of ISR.

Definition 4: The Set IST_{ft} is a 5-tuple, IST_{ft} = <SN, DN, Dl, CS, P>. Among them, SN refers the node which sends ISR. DN refers the node which ICs scheduling happens in. DL is the deadline of ISR finished. CS is the lists of ICs which are needed showing. P refers the priority of ISR.

Definition 5: The SN and DN of an IST is the same node, the task is called Local Scheduling Task (LST), and the corresponding ISR is called Local Scheduling Request (LSR). Otherwise the task is called Uniform Scheduling Task (UST), and the corresponding ISR is called Uniform Scheduling Request (USR).

Definition 6: CS is a list, CS = {Componentk|0<k<=N, Componentk □ ICs}. There, N is the maximal number of ICs which a display can show synchronously.

Theorem 1: The IST in set IST_{ft} can all be succeeded in carrying out, if and only if for each ISR of each IST, the interval from beginning that the ISR executes successfully to DL is greater than the interval that the scheduling course needs. This is described as (1).

$$\forall IST \in IST_{ft},\ IST \rightarrow sucess \Leftrightarrow$$
$$\forall IST \in IST_{ft},\ \forall ISR \in IST,\ DL - ISRST \geq ET \qquad (1)$$

Among them, → refers executing, success refers executing successful, ISRST refers the beginning time of the ISR executing successful, ET refers the time that executing scheduling needs.

This model is made up of Center Scheduling Component and many human-computer interactive nodes. Each human-computer interactive node includes a Local Scheduling Processing Component, an Interface Generating Component and some human-computer interactive interface components.

The architecture of the fault-tolerant scheduling model is shown as Figure 4. This model is made up by a Center Scheduling Component (CSC) and many human-computer interactive nodes. Each human-computer interactive node includes a Local Scheduling Component (LSC), an Interface Generating Component (IGC) and some interface components (ICs). When the system is structured, a LSC, an IGC, and some ICs are disposed on each human-computer interactive node based on the logical function of the node (finished work of the node self) and the fault-tolerant requirement for other nodes (substituting other nodes breaking down).

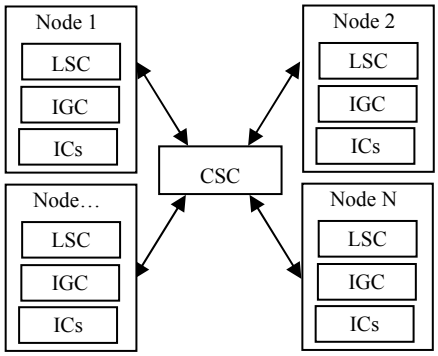


Figure 4. the fault-tolerant scheduling model

The scheduling process of ICs is illustrated using example of node 1 as follows. According to requirements, the operator of node 1 sends an ISR to LSC of node 1. If the ISR is LSR and belongs to set IST_{nft}, the LSC deals with the ISR, and sends the result Interface Forming Information (IFI) to IGC of node 1. The IGC carries out ICs addressing, framework generating and ICs loading. Finally, the display is shown. Otherwise, the ISR is sent to CSC by the LSC. The CSC deals with the ISR, and sends the IFI to the IGC of the node which the ISR specifies. The IGC finishes the work as the IGC of node 1.

In addition, CSC detects the running state of each node. Suppose there is a fault in node 1. According to the load of other nodes, reliability requirement and priority of the running ICs of node 1 at that time, CSC produces one or more scheduling request, which are carried to make other nodes substitute the entire or partial functions of node 1. At the same time, a message is sent to show that there is a fault in node 1, which needs to be maintained. This course is called systemic FT. Moreover, if the operator thinks the scheduling result of above description is unsuitable, he can send ISR to reschedule. This course is called factitious FT.

V. SCHEDULLING ALGORITHM

The main algorithms used in the model are introduced in this section.

A. The Algorithm of LSC

The main function of LSC is to receive ISR. If the ISR is a LSR without FT requirements, it was processed by the LSC. Otherwise, it was sent to CSC. The algorithm of LSC is described as algorithm 1.

Algorithm 1

Step1: receive ISR; /* ISR∈IST, IST∈ISTS */
Step2: if (IST∈IST_{nft}) goto Step3; else goto Step4;
Step3: if (IST.DN == IST.SN) goto Step5;
else goto Step4;
Step4: send ISR to CSC; goto Step7;
Step5: process ISR, and get the IFI;
Step6: send the IFI to IGC;
Step7: end.

B. The Algorithm of CSC

The CSC has three functions. First, receiving ISR sent by LSC and pushing it in the appropriate position of ISR queue according to the priority of ISR. Second, processing the request of the front of the queue and sending the IFI to the IGC of DN of ISR. Third, detecting the state for each node and realizing systemic fault-tolerant. If there is a fault in a certain node, according to the fault-tolerant requirements of the node and systemic load, the entire or partial functions of the node are substituted by other nodes. The algorithms of CSC are described as algorithm 2 and algorithm 3.

Algorithm 2

Step1: get the front request, processing to get the IFI;
Step2: if ($IST \in IST_{nft}$) goto Step4;
Step3: get startTime; /*startTime refers the time of the ISR starting to execute */
if (DL - startTime >= ET)
send the IFI to IGC of IST.DN;
else goto Setp7;
goto Step5;
Step4: send the IFI to IGC of IST.DN; goto Step6;
Step5: receive the return (value)of IGC of IST.DN;
if (!value) goto Step3;
Step6: scheduling success; goto Step8;
Step7: scheduling failure;
Setp8: end.

Algorithm 3

Step1: detect Ni; /*Ni ∈ the set of HCIN */
if (! malfunction) loop(timeout);
Step2: if(all of ICs running in Ni without FT)
goto Step6;
Step3: CSC selects one or more HCIN, recording DNi. Then, produce FT Information, process to get IFI, and send it to DNi;
Step4: get startTime; /*startTime refers the time of the ISR starting to execute */
if (DL - startTime >= ET)
send the IFI to IGC of DNi;
else goto Setp7;
Step5: receive the return (value)of IGC of DNi;
if (!value) goto Step4;
Step6: notice Ni malfunction, and schedule the ICs with FT showing in DNi; goto Step8;
Step7: notice Ni failure, and scheduling failure;
Step8: end;

C. The Algorithm of IGC

The functions of IGC are ICs addressing, frame generating and ICs loading according to the IFI of CSC or LSC sending. After that, send the display result to CSC. The algorithm is described as algorithm 4.

Algorithm 4

Step1: receive the IFI from CSC or LSC;
Step2: addressed ICs according to the IFI;
Step3: generate frame according to the IFI;
Step4: load ICs;
Step5: send the display result to CSC.
Step6: end;

VI. EXPERIMENT ANALYSIS

The architecture of simulation experiment is made up of Information Process System, Application System, Data-Command Agency and Human-computer Interactive System. The software platform is CORBA. Among them, the Information Process System produces and processes the simulation necessary data. The Data-Command Agency manages the data and command uniformly. The architecture of simulation system is shown as Figure 5.

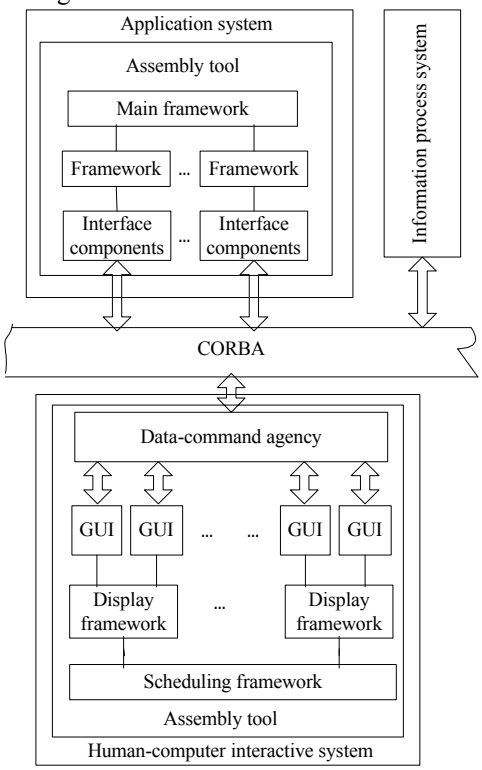


Figure 5. the architecture of simulation system

The fault-tolerant scheduling model of the Human-computer interactive subsystem is determined according to the designing result file which is generated recurring to the assistant developing tool. In the model, the CSC cooperates with LSC to accomplish the fault-tolerant scheduling function. The LSC receives ISR, processing it or sending it to CSC according to the fault-tolerant requirement. The CSC receives and processes ISR sent by LSC to generate interface forming information. In the mean time, it detects the states of each node to generate the fault-tolerant information. The fault-tolerant scheduling model is shown as Figure 6.

The Human-computer interactive subsystem configuration is shown as follows in simulation system.

(1) Hardware: PIII800 processor; 256M memory; 10/100M Ethernet.

(2) Software: Linux OS; CORBA;

The results of the experiment are described as follows.

A. The system running normal

The ISR is LST: The scheduling result is correct. The time from the ISR sent to the scheduling finished is 300ms.

The ISR is UST: The scheduling result is correct. The time from the ISR sent to the scheduling finished is 500ms.

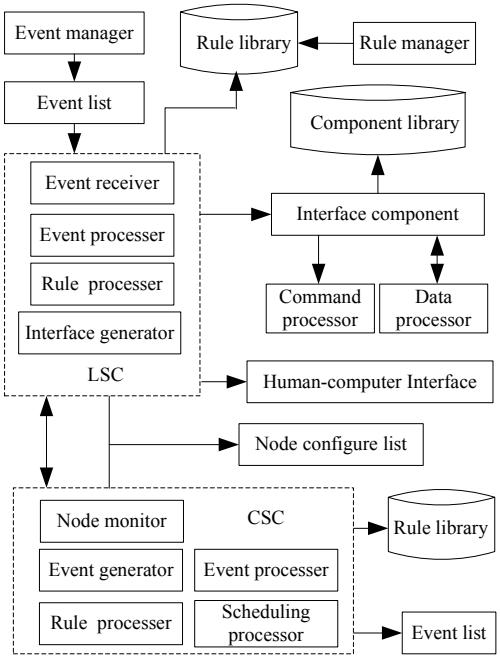


Figure 6. the fault-tolerant scheduling model

B. One node malfunction

FT: The FT result is correct. The time from the malfunction happening to the processing finished is 2-2.5s. This time is impacted mainly by the time of the malfunction detecting.

ICs Scheduling: The scheduling result is correct. The scheduling time is the same as (A).

C. Two nodes malfunction

FT: Commonly, the FT result is correct. The time from the malfunction happening to the processing finished is 2-3s. This time is impacted mainly by the time of the malfunction detecting. When there are many ICs with FT requirements in fault nodes, and the priority of the ICs running in other two nodes is more prior, partial functions of the fault nodes are substituted possibly.

ICs Scheduling: The scheduling result is correct. The scheduling time is the same as (A).

VII. CONCLUSION

Three achievements are done in this paper. First, a fault-tolerant scheduling model is presented which can realize interface component tolerant and uniform switching control in distributed human-computer interactive system. Furthermore, it can realize dynamic reconfiguration and upgrading online of interface components. Second, a new method and technique is provided to design human-computer interactive system rapidly. Third, the feasibility of the model and its algorithms are proved by the simulation experiment.

There are still some problems remained to study such as the granularity division of the interface components, the mathematic model of effect evaluating of scheduling, the dynamic link and composition of interface components, which will be researched in the future work .

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China under Grant No.70671035 and the Key Scientific & Technological Project of Henan Province China under the grant No.082102210076.

REFERENCES

[1] Lu, Liu; Zongyong, Li; Ruibo, Li. “Improving information system flexibility through remote dynamic component configuration”. International Conference on Service Systems and Service Management(ICSSSM 06). Oct.2006, pp. 461-466.

[2] Ou, Shumao; Liu, Dongsheng; Yang, Kun. “Dynamic algorithms for self-deployment and self-configuration of pervasive service components”. Intelligent and Software Intensive Systems (CISIS 09) IEEE Press, Mar. 2009, pp. 525-530.

[3] Yang, Huaizhou, Li Zengzhi. “Research on QoS-aware and dynamic configuration of web services composition system,” Journal of Xi'an Jiaotong University, vol. 44, Feb.2010, pp.25-30.

[4] Venkita Subramonian, Gan Deng, Christopher Gill. “The design and performance of component middleware for QoS-enabled deployment and configuration of DRE systems,” Journal of Systems and Software, Vol.80, May 2007, pp.668-677.

[5] Wu Haomin, Cao Min. “Description of Software Architecture Supporting Dynamic Reconfiguration and Abstract Programming”. Computer Engineering & Applications. Oct 2004, pp.94-98.

[6] S M Ellis. “Dynamic software reconfiguration for fault-tolerant real-time avionic systems”. Microprocessors and Microsystems. Vol.21, 1997, pp.29-39.

[7] Cao Min, Wu Gengfeng. “Architectural Level support for Dynamic Reconfiguration and Fault Tolerance in Component-Based Distributed Software”. Computer Engineering & Applications. Jun 2004, pp.100-104.

[8] Song Yi, Liu Yunchao. “A Graph-oriented Approach for Dynamic Reconfiguration and Fault Tolerance in Distributed Software”. Computer Applications. Vol.23, Dec 2003, pp.37-41.

[9] Brandt Steven, Allen Gabrielle, Eastman Matthew. “Dynamic deployment of a component framework with the Ubiqis system,” International Conference on the Applications of Digital Information and Web Technologies(ICADIWT 09), Aug.2009, pp.68-74.

[10] Lazar, I. Parv, B. Motogna, S. “A platform independent component model for dynamic execution environments,” the 10th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC 08), Sep.2008, pp 257-264.

[11] Espiritu Jose F, Coit David W. “Component replacement analysis for complex electricity distribution configurations,” IIE Annual Conference and Expo 2008, May.2008, pp.170-175 .

[12] Liu Yunsheng, Xia Jiali. “Scheduling Real-time Transactions Based on Functional Alternation Characteristics”. Chinese Journal of Computers. Vol.26, No.2, Feb 2003, pp.250-256.

[13] KaDae Ahn, Jong Kim, SungJe Hong. “Fault-tolerant real-time scheduling using passive replicas”. FTCS, Vol27, 1997, pp.98-103.

[14] Philippe Aniorte. “A distributed adaptable software architecture derived from a component model”. Computer Standards & Interfaces. Vol.25, 2003, pp.275-282

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15, August 2010, pp. 422-425

Research of Cooperation of IPSec and Firewall

Yang Li-shen[1], Ren Zheng-wei[2]
[1] School of Computer Science & Technology Henan Polytechnic University ,Jiaozuo, China
Email: yangls@hpu.edu.cn
[2] School of Computer Science & Technology Henan Polytechnic University ,Jiaozuo, China
Email: rzw801@163.com

***Abstract*—IPSec provides security services at the IP layer and ensures the packets transmitted safely in Internet by authenticating and encrypting. As IPSec encapsulates some important information of packets, it can not cooperate efficiently with packets filter firewall, which filters packets according to protocol and port. For the cooperation question of IPSec and firewall, this paper proposes the solution that handles security problems on protocol head and datagram separately, combines this layered approach with the key agreement way, and lets the firewall involved in the key agreement phase of IPSec , make the encrypted data packets pass, thus solving the compatibility operation problem.**

***Index Terms*—IPv6, IPSec, IKE, safe connection(SA), packets filter firewall**

I. INTRODUCTION

With the rapid development of Internet, the existing IPv4 network reveals gradually many problems. The address space is shortage. Network quality of service is difficult to guarantee. Network bandwidth is constrained. Support of mobility is limited. It is so difficult to solve the problems of network security.IPv4 has been difficult to cope with the rapid development of information networks. With more and more challenges, it has been inevitable history that IP agreement carries out the transition of from IPv4 to IPv6 [1]. IPv6 support IPSec forcibly, it implements authentication based on network layer, integrity and confidentiality. Firewall is mainly used to protect the internal network, while IPSec is mainly used to protect the security of data transported on the network, it will be more conducive to combine the two technologies to protect data security of the whole network [2]. The most commonly used firewall is designed for IPv4. This paper studies on the cooperation of IPSec and Firewall.

II. IPV6 ADVANTAGES

The current Internet protocol, version 4, known as IPv4, poses several problems such as impending exhaustion of its address space, configuration and complexities due to rapid growth of the Internet and emerging new technologies. As a result, IETF developed the next generation IP, called IPv6.In addition to the expansion of address space, improve network speed and enhanced security features, ipv6 also can reduce the router cost by clustering mechanisms, reducing the need to maintain the routing table. The state and stateless address allocation can reduce the network administrator to maintain IP address of the workload; better QoS in dealing with streaming media data can ensure that the information flow[3]. Extended protocol allows the sender to add packet information to enhance network flexibility.Ipv6 have other technical advantages.

III. IPSEC PROTOCOL

When IETF established IPv6 standards, security protocol of IPSec in network layer is introduced, and becomes an integral part of protocol stacks. IPSec introduces authentication and encryption mechanisms to ensure the integrity of data packets and confidentiality, provides network layer-based authentication, therefore, achieves safety of network layer, guarantees end-to-end communications security service provided includes access control, no link integrity, data sources authentication ,confidentiality. As the network's rapid development, the traditional IPv4 has brought security problems to get worse. Network attacks, information leaks and other incidents have occurred from time to time. To this end, IETF has developed IPSec protocol to protect IP Communications. IPSec is an optional extension for IPv4，but it is a necessary component for IPv6. Its main function is to provide encryption, authentication and other security service in the network layer, which provides two security mechanisms.

IPSec security framework includes AH, ESP, IKE, the security associations and other related components. AH is authentication header and is used to authenticate the identity of network packet. ESP is Encapsulating Security Payload and is used to encrypt the contents of network transmission. IKE is Internet Key Exchange and is used to manage key.

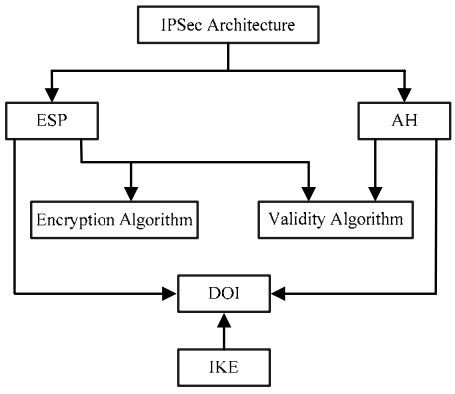


Figure 1 IPSec Architecture

A. AH

AH supports MD5-96, SHA1-96 authenticated encryption algorithm. It mainly provides authentication of information packets, and detection of data integrity, but also has the “anti-replay” feature. AH message format consists of five fixed-length field and variable-length authentication data domains. Five fixed-length fields are the next header field, payload length of the field, reserved fields, SPI and serial number field. The next header field in which the value of the transmission mode is the value of the protected upper layer protocol, such as TCP or UDP. In tunnel mode, if it is IPv6 package, the value is 41. The location of AH header depends on the AH operating mode [4]. AH have two operating modes: transport mode and tunnel mode. As shown in the following figures.

IP header	AH header	data

Figure 2 AH transport mode

new IP header	AH header	older IP header	data

Figure 3 AH tunnel mode

B. ESP

Encapsulating Security Payload (ESP) is a header which is inserted inside IP packet. Encapsulating Security Payload supports DES-CBC, 3DES-CBC, RC5, CAST-128 and other encryption algorithm. It provides encryption for the IP layer, and authentication for data sources. In fact, ESP provides similar services as AH, but adds data confidentiality and the confidentiality service of limited data stream. Confidentiality services encrypt the relevant parts of IP data packets by practical cryptography. Confidentiality of the data stream is provided by confidential services in the tunnel mode. ESP packet format is formed by the four fixed-length fields and three variable-length domains [5]. ESP's position in the data packet depends on the operation mode of ESP. There are also transmission mode and tunnel mode, as shown below.

IP header	ESP header	data	ESP end	ESP authentication

Figure 4 ESP transport mode

New IP header	ESP header	Older IP header	data	ESP end	ESP authentication

Figure 5 ESP tunnel modes

C. Internet Key Exchange (IKE)

IKE is a generic Internet Key Exchange protocol, which is an agreement by the other three protocols (ISAKMP, Oakley and SKEME) .IKE negotiate cryptographic algorithms which are used by AH, and ESP protocols, and put keys of algorithm required on the right place. IKE protocol which uses key, data encryption and other way provides authentication for both of communications. IKE determines which service should be used for different business stream, and for the services negotiations and consultations needed to SA and other key. IKE makes different business streams determine what services should be used, and negotiate key for the services need, and consultation SA. It negotiates SA, verifies entities and exchanges of keys through ISAKMP packets, the ISAKMP exchange and payload. IKE includes key exchange and key distribution, up to four keys: AH and ESP have both sending and receiving the keys. IKE completes SA consultations through two stages of exchange [6]. The first stage is used to create the IKE SA, and the second phase is used to create IPSEC SA on the basis of IKE SA protection.

D. SA

Security association is the foundation of IPSec; it can be divided into protocol header SA and data area SA. Protocol header SA processes the part of the protocol header, and the data area SA processes the data part. The different SA is distinguished through the SPI (security parameter index). SA tracks the details of IPSec session, which is negotiated between the two nodes. AH finishes authentication of the two nodes, and ESP completes the encryption between two nodes [7].

E. The impact of IPSec on IPv6 network

In terms of security features, IPSec can ensure data integrity and confidentiality in the authentication, so that communication security, access control, privacy have been strengthened, but it can not be guaranteed for availability and non-repudiation. Common attacks for IPv4 network include IP fraud, replay attack, reflection attack, middle attacks, application layer attacks and so on. Because IPSec improve the security of data transmission in IPv6, most of the attack acts will be relieved about data authentication, data integrity and data confidentiality. Some will be inhibited including IP fraud, replay attack, network attacks and other forms of attack. Other attack acts will be continually existed about resource depletion and the protocol defects, including denial of service attacks, application layer attacks and so on.

IV. FIREWALL

A firewall is a defense system which makes the local network and the external network isolated. That allows Local Area Network (LAN) and the Internet or other external network between each isolation, limited network visits to protect internal network. In addition to network management, set rules about access and be accessed , cut off access which is prohibited, the firewall in the computer systems also need to analysis and filter out the data package, to monitor and record content and activities of the information through the firewall [2]. It also can detect and alarm the attack acts from the network. These

are the basic functions of the firewall. Whether hardware firewall or software one should have these five basic functions. Firewall works in the network layer or application layer. Firewall working in the network layer filters the information of transmission mainly through packet filtering technology, which selects the data in the network exports (routers) .Just only those packets that meet the conditions are allowed to pass, and others are abandoned. Network layer firewalls can allow the host and the servicer which are authorized, directly access to the internal network. You can also filter a specified port and Internet address information of the internal users, and limit the internal network to access the external network [8]. The application layer firewall is to control applications access. It is essentially an application gateway, also called a proxy server (Proxy Server). When users use a TCP / IP applications, the proxy server will ask users to provide external network host name. If the users answer and provide the correct user identity and authentication information, the proxy server establishes connections between internal network and Internet hosts, and acts as a relay for two communication entities.

V. COOPERATION OF IPSEC AND FIREWALL

IPv6 is the next generation IP protocol. It solves problems of the current IP address shortage, introduces encryption and authentication mechanisms, and implements authentication on the network layer. It is ensured for the integrity and the confidentiality of data packets, so you can say that IPv6 achieves the security of network layer. Though IPv6 has good security features, it can not replace the firewall. The current firewall generally provides the filter form of the physical layer, and lacks authentication and encryption mechanisms [2]. If it add authentication and encryption mechanisms, the firewall will be more perfect. IPSec is used to determine authentication mechanism and encryption mechanism of a specific connection, which can farthest hide the contents of the packet for the intermediate nodes, so it protects the data from attack. Firewall is to focus on the connection based on security policy and between the specific networks, which filters the packet under the upper layer protocol and port to prevent unlawful incursion on the internal network [8].

IPv6 uses encryption option, the data is encrypted transmission. AS IPSec encryption provides the end to end protection, and encryption algorithm can be optional, the key is not public. Because firewall is unable to decrypt the packet, firewall can not know TCP / UDP port. If the firewall releases all of the encrypted packets, it will no longer be able to restrict the port that the external users can access, and it also can not prohibit external users' illegal access. The chief cause of conflict between IPSec and firewalls is that the firewall need to access packet header information and protocol header of the transport layer, and may be modified; but IPSec is to be encryption or authentication for the entire packet including protocol header, preventing firewalls from doing their jobs properly. So it's inevitable that the protocols part and the data one will be processed separately.

In order to make the firewall access or modify protocol header information in the IP packet, this paper adopts to deal with protocol header and data part on the IP packet a safe disposal separately. The data portion deals with a safe disposal between hosts, which need a secure communication, each other. That is to say, it applies encryption and other safe disposals in the sender, and also applies encryption and other safe disposals equally in the receiver [9]. It is completely transparent for the intermediate nodes in the network, such as firewalls, routers, dealing with safe disposal of data related. It is in contrast to end process mode of data parts, the safe disposal of protocol header adopts subsection mode, which is the protocol header in the transmission process are protected by stages. Between the two firewalls, between the receiving host and its firewall, deal with the same safe handling equally. This three equal handling is mutual independence, each producing a certain stage of transport security. The firewall is involved in the IPSec key agreement phase, and records the security association and each key, which is established the connection through the firewall. In other words, the firewall instead of the internal host and external host exchange a session key and establish the tunnel. In the same time, It is established other encrypted channel between the internal host and the firewall instead of external host. Once the SA began negotiations, the firewall involved in the consultation process between the two hosts .It records the outcome of the SA consultations, and sends the negotiated SA to the firewall at the same time. Then the firewall can easily decrypt the data, and do the corresponding safe disposal for the packet header through the received SA. At last, the firewall makes the data through the security check encrypt and transmit to the receiver of the data packets. This means that IPSec can achieve the end to end protection; the firewall can also get the TCP / UDP port which the messages use, and implements correctly the filtering function [10]. This solution can allow IPSec and firewall to get what they want, and solves the conflict between IPSec and the firewall.

VI. CONCLUSION

This paper presents a solution to the conflict between IPSec and firewall. It is to process separately the transport layer protocol header and user data, using a different security association SA, so we can solve the problem above. The program is completely transparent to the outside of the firewall node, and easy implemented, ensuring network security.

ACKNOWLEDGMENT

The authors would like to thank Institute of Computer Science and Technology in Henan Polytechnic University for their sponsoring to the subject and all the numbers helpful for my paper.

REFERENCES

[1] ZHANG Yu-fang, XIONG Zhong-yang, LAI Su, "Design and Implementation of IPv6 Transparent Firewall with Virus Filtering", Computer Science, Volume 36,Issue 4,Apr. 2009,pp.108-111.

[2] LI Li, YUAN Xin-zhi, ZHENG Chao-mei, WU Fang-yu,"Research and implementation of cooperation of IPSec and firewall based on IPv6", Computer Engineering and Design, Volume 16, Issue 27, Aug.2006,pp.2970-2972.

[3] S Dearing, R Hinden,"Internet Protocol, Version 6 (IPv6) pacification", RFC1883, December, 1996.

[4] Kent S.,"IP Authentication Header", RFC 4302, December 2005.

[5] Kent S.,"IP Encapsulating Security Payload (ESP)", RFC 4303, December 2005.

[6] Kaufman C.,"Internet Key Exchange (IKEv2) Protocol", RFC 4306, December 2005.

[7] Kent s, K seo,"Security Architecture for the Internet Protocol", RFC 4301, December 2005

[8] LIU Zhan-wen, ZHAO Yu-guo, WANG Ji-cheng,"Linux-based IPv6 bridge Firewall Technology", Network & Computer Security , June 2008,pp.66-67,70.

[9] HE Hong-lei,"Research on IPv6 Firewall", Network & Computer Security,No.7,2009,pp.58-59,69.

[10] SUN Yong, ZHANG Heng, MA Yan, WEN Xiang-ming,"IPv6-based Interaction System with Intrusion Detection and Firewall", Computer Engineering,Vol..34.No.11,June.20

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15, August 2010, pp. 426-430

An improved method for classifying XML documents based on structure and content

Zhang Na [1], Zhang Dongzhan [1], Yu Ye [1] and Duan Jiangjiao [1]
[1] Computer Science Department, Xiamen University,Xiamen, China
Email: zdz@xmu.edu.cn

***Abstract*—As more and more structured or semi-structured data is stored and exchanged in XML format, XML mining becomes increasingly important, especially the study of classification of XML documents becomes more widely. Considering the disadvantage of the current classification of XML documents that based on structure and content, this paper presents an improved method called NM-Similarity computing similarity measure, which maintains an high accuracy rate when XML documents are similar in structure but different in content. This method is applied in KNN (K-Nearest Neighbor) method for classification. The structure similarity between two XML documents is computed by using Euclidean distance, and the content similarity is computed by using Cosine measure. A better result can be seen when classifying XML documents which focus on content more (that is, XML documents are created from the same DTD and the structure is similar) and it is more effective on classifying XML documents. The experiments prove that when XML documents are similar in structure but different in content, NM-Similarity in this paper provides a significant improvement in improving classification accuracy rate.**

***Index Terms*—Euclidean distance, Edit distance, XML, Classification, KNN, Cosine measure**

I. INTRODUCTION

XML (Extensible Markup Language) has become the standard language for data transmission and exchanging on the web. XML can not only store data, but also can store the structure and semantic information, and it has common data processing capability, besides, it can present structured, semi-structured or element structure data. As more and more structured or semi-structured data stored and exchanged in XML format, XML data mining becomes increasingly important, especially, researching on the classification of XML documents becomes more widely[1,2].

According to the nature of XML documents, the structure of XML documents has many models when they are classified, such as the tree based model, the map based model, the path based model and so on, and the core issues of the XML structure analysis is the structure similarity of XML documents. When the XML document is considered as a tag tree, the existing structure similarity of XML documents mainly consists of edit distance method, the match method of path and the analysis method of Time Series. Besides structure, the contents of the XML document play a big role in classifying XML documents, but the text classification cannot distinguish the differences involved the structure of XML documents, so it has already been researched deeply from two perspectives—the structure and the content.

This paper studies and discusses the method of XML classification. It also discusses the description of implementation of each section in classification process. In this paper, differing from other methods, the improved method considers from both structure and content to compute the similarity of two XML documents by using Cosine measure. It shows a good result when operating the high-dimensional vector like the content of documents.

The rest of the paper is organized as follows: Section 2 reviews related work on XML document classification. In Section 3, this paper presents an improved method called NM-Similarity computing similarity measure and used in KNN method. Section 4 mainly discusses the experimental evaluation of the approaches and the comparison of other methods. Section 5, concludes the paper.

II. RELATED WORK

At the present time, XML document classification mainly has three methods, including structure-based, content-based, based on the structure and content. In the following, some of the research work dedicated to XML document classification is briefly summarized.

In [8], a classifier called XRules is developed, the training phase uses a database of structures with known classes to build a classification model and the testing phase takes as inputting a database of structures with unknown classes using the classification model to predict their classes. XRules only reflects regular structural patterns of each class.

In [9], it proposes a new method of classification based on hierarchical structure for XML document, paying more attention to the unique structural information of XML document. TF-IDF is used to generate general feature set for the non-structural information of XML document, generate hierarchical feature set for the importance of each hierarchy of XML document and the knowledge feature set is generated by the domain knowledge, then classifies XML documents by combining three feature set.

In [10], it explores the application of clustering methods for grouping structurally similar XML documents. This paper applies clustering algorithms using distances that estimate the similarity between those trees in terms of the hierarchical relationships of their nodes.

This work is supported by Chinese National Natural Science Foundation (50604012)

But this method only reflects the aspect of structure and ignores the importance of content of XML documents, so it influences the accuracy of classification to a certain extent, and when XML documents are very large, the cost that computes the edit distance between two documents is very high.

In [11], a method simplifies the representation of XML document tree using repeated pruning and nested pruning, then computing the edit distance between two XML documents simplified. This paper uses hierarchical clustering to classify the XML documents according to the edit distance. But it does not take the content of the XML tree into account.

In [12], a method is defined for computing the distance between any two XML documents in terms of their structure. The lower this distance is, the more similar the two documents are in terms of structure, and the more likely they are to have been created from the same DTD. But when two documents created from the same DTD can have radically different structures, the result is not good.

In [13], a similar work to that described in [11] is developed. The method, however, is based on the KNN algorithm that relies on an edit distance measure. The approach explores both the content and the structure of XML documents for determining similarity among them. But it is more focus on the differences of structure of XML documents and it is less effective on classifying XML documents whose structure is more similar than content.

III. NM-SIMILARITY: DOCUMENT SIMILARITY MEASURE

As stated in the above, in [13], the method is more focused on the difference of structure, and it is less effective on classifying XML documents whose structure is more similar than content and the XML documents are created from different DTD. Therefore this paper proposes a new method called NM-Similarity on the basis of the method computing similarity between two XML documents in [13], and it is effective on classifying XML documents which are created from the same DTD. NM-Similarity explores both the content and the structure of XML documents for determining similarity. Steps for computing each course appear below.

NM-Similarity is used to get the similarity of two XML documents which is a combination of the content similarity value and the structure similarity value. The structure similarity between the two XML documents is computed by using the Euclidean distance, and the content similarity is computed by using Cosine measure. The document similarity between two XML documents, d_x and d_y, is defined as (equation 1):

$$docSim(d_x, d_y) = (conSim(d_x, d_y) \cdot \lambda) + (strSim(d_x, d_y) \cdot (1-\lambda)) \quad (1)$$

$\lambda \in (0, 1)$, and it depends on the importance of the content similarity and structure similarity.

A. Structure Similarity Measure

Fundamental content of an XML document contains declaration, comment, tag, element, attribute and text. The XML data model is a dendrogram with notes; tree node is the element in the XML document and the notes of node is attributed in the XML document. The structural information of the tree such as element names, data types, parents, children, etc. can be used to compute the structural similarity between two XML documents. To simplify the structure matching process, only the names, which are the most important property of the elements, are used for structure matching.

An XML document can be considered as a tree-based and it is a collection of distinct paths. The structural distance between XML documents can be calculated by these paths. Let D is a dataset of XML documents $\{d_1, d_2, \ldots, d_n\}$, P is a set of distinct paths $\{ p_1, p_2, \ldots, p_f \}$. p_i is a path in P, which contains all element names from the root element to the leaf element. The leaf element is an element that contains the textual content. d_i is the structure of a document, which is modeled as a vector $\{ p_{i,1}, p_{i,2}, \ldots, p_{i,,f} \}$, where each element of the vector represents the frequency of a path in P that appears in the document. Given two documents, d_x and d_y , and their corresponding vectors, $\{ p_{x,1}, p_{x,2}, \ldots, p_{x,f} \}$ and $\{ p_{y,1}, p_{y,2}, \ldots, p_{y,f} \}$ respectively. The distance between the two documents is computed using the Euclidean distance (equation 2).

$$strSim(d_x, d_y) = \sqrt{\sum_{i=1}^{f} \left(P_{x,i} - P_{y,i} \right)^2} \quad (2)$$

$\mathbf{d_1}$	$\mathbf{d_2}$
<R>	**<R>**
<S_1> t_1,t_2, t_3	**<S_1> t_1,t_4**
<S_2> t_4,t_3, t_6	**<S_2> t_2,t_8**
<$S_{2.1}$> t_2,t_7, t_8	**<$S_{2.1}$> t_2,t_9**
<$S_{2.2}$> t_3,t_7	**<$S_{2.1}$> t_5,t_7,t_8**
<S_3> t_5,t_4, t_7	**<S_3> t_4, t_7**

Figure 1. Two XML documents.

As shown in figure 1, the set D contains two XML documents $\{d_1,d_2\}$,element names in the documents are shown as embraced within brackets, <R> is the root element and <S_i> is the internal element of leaf element. The content of a document is denoted by T {t1, t2,…, tm}. The structure of a document is extracted and represented as vector. The structures of all the documents in the dataset can be put together as a document-path matrix, $Y_{n\times f}$, where f is the number of distinct paths in P and n is the number of documents in D, as shown in table 1. Each cell in matrix Y is the frequency of a distinct path appearing in a document.

B. Content Similarity Measure

As shown in figure 1, a set of distinct terms $\{t_1, t_2,\ldots, t_m\}$, denoted by T, is extracted from the dataset D. A term, t_i, is a keyword that appears in the textual content of the elements in the XML document after punctuation mark and stop-word removal and stemming, as shown in figure 2 and figure 3. Let $X_{n\times m}$ is a document-term in T, where m is the number of terms in T and n is the

number of documents in dataset D, is constructed as shown in table 2.

TABLE I.
TWO VECTORS REPRESENTING THE STRUCTURE INFORMATION OF THE DOCUMENTS

document/path	R/ S_1	R/ S_2/ $S_{2.1}$	R/ S_2/ $S_{2.2}$	R/ S_3
$\mathbf{d_1}$	1/4	1/4	1/4	1/4
$\mathbf{d_2}$	1/4	2/4	0	1/4

TABLE II.
TWO VECTORS REPRESENTING THE CONTENT INFORMATION OF THE DOCUMENTS

document/ term	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9
$\mathbf{d_1}$	1/9	2/9	3/9	2/9	1/9	1/9	3/9	1/9	0
$\mathbf{d_2}$	1/9	2/9	0	2/9	1/9	0	2/9	2/9	1/9

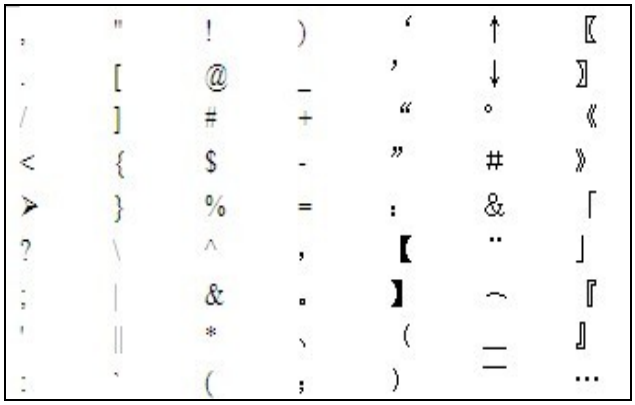


Figure 2. Punctuation Mark Set

a	about	above	Must	may	next
across	after	afterwards	neither	never	no
again	against	all	Nor	Not	nothing
almost	alone	already	Now	of	off
also	among	an	other	otherwise	our
and	another	any	Out	over	own
anyone	anything	anywhere	Per	perhaps	same
are	around	as	Seem	several	she
at	be	because	should	since	so
become	been	before	Some	someone	something
behide	besides	between	sometimes	somewhere	still
beyond	both	but	Such	to	together
By	can	could	Too	towards	than
Down	during	each	Then	that	the
enough	etc	even	their	them	therefore
Ever	every	everyone	This	those	though
Everything	everywhere	few	through	thus	under
first	for	from	until	up	us
further	had	has	Very	via	was
Have	he	her	We	well	were
Here	him	his	What	whatever	when
How	however	last	whole	whose	why
latter	least	less	where	will	with
Many	may	me	without	would	
might	more	most			

Figure 3. Stop-word

The content of a document, di, is modeled as a vector {ti,1, ti,2, …, ti,m }, where each element of the vector represents the frequency of a term in T that appears in the document. Given two vectors, dx and dy, and their corresponding vectors, { dx,1, dx,2, …, dx,f } and { dy,1, dy,2, …, dy,f } respectively. As the cosine measure and metric is inversely proportional to the Euclidean distance, content similarity between two XML documents is measured as (equation 3):

$$ConSim(d_x,d_y)=1-\cos(d_x,d_y)=\frac{d_x^t\cdot d_y}{\|d_x\|\|d_y\|}=\frac{\sum_{i=1}^{f}d_{x,i}\cdot d_{y,i}}{\sqrt{\sum_{i=1}^{f}d_{x,i}^2}\cdot\sqrt{\sum_{i=1}^{f}d_{y,i}^2}} \quad (3)$$

d_x^t is the transposition of the vector d_x, $\|d_x\|$ is Euclid number of the vector d_x, and d_y is Euclid number of the vector d_y.

C. Application of NM-Similarity in KNN

NM-Similarity will be applied in KNN method in this paper.

The K-nearest neighbor's algorithm (KNN) is a method for classifying objects based on closest training examples in the feature space. KNN is a type of instance-based learning, or lazy learning where the function is only approximated locally and all computation is deferred until classification. The *k*-nearest neighbor algorithm is amongst the simplest of all machine learning algorithms: an object is classified by a majority vote of its neighbors, computing the Euclidean distance between an unknown object and its neighbors, with the object being assigned to the class most common amongst its *k* nearest neighbors (*k* is a positive integer, typically small).

IV. EMPIRICAL RESULTS

The classification accuracy of NM-Similarity will be compared with the method in [11] and [13].

A. Evaluation Methods

In this paper, classification system indicator which is used to measure the accuracy of the classification solution contains precision and recall. Precision can be seen as a measure of exactness or fidelity, whereas Recall is a measure of completeness.

Precision is defined as the number of true positives (i.e. the number of items correctly labeled as belonging to the positive class) divided by the total number of elements labeled as belonging to the positive class (i.e. the sum of true positives and false positives, which are items incorrectly labeled as belonging to the class).

Recall in this context is defined as the number of true positives divided by the total number of elements that actually belong to the positive class (i.e. the sum of true positives and false negatives, which are items which were not labeled as belonging to the positive class but should have been).

Given a particular category, setting the required terms: true positives, true negatives, false positives and false

negatives to tp, tn, fp, fn respectively, precision and recall are defined as follows(equation 4 and equation 5):

$$Precision = \frac{tp}{tp + fp} \quad (4)$$

$$Recall = \frac{tp}{tp + fn} \quad (5)$$

B. Results and Analysis

In order to validate the approach in this paper, we evaluate our approach on real classification data sets, which are quoted from Wikipedia datasets [14, 15]. The experiment environment is as follows: P8600 CPU at 2.4GHz, the RAM memory size is 2G-Byte, development tools is Microsoft Visual Studio 2005, development language is C#. We uses three types in Wikipedia data sets, which are created from the same DTD. The average size of the XML document is 20k, every type has 25 XML documents and the number of unlabeled documents is 80. The value of λ is set based on experience in observation and experiment in fact. In all experiments in this paper, set the value of λ to 0.8.

To check the effectiveness of our approach in this paper called NM-Similarity, we will use methods in [11] and [13]. These are in the following briefly described, as shown in table 3 and table 4.

TABLE III.
ACCURACY RESULTS

Method	Precision		
	K=5	*K=7*	*K=9*
NM-Similarity	0.9125	0.925	0.9125
Method in [13]	0.375	0.425	0.4625
Method in [11]	0.4125	0.4	0.3875

In [11], it improves the calculation of the edit distance between two XML documents, simplifies the representation of XML document tree by using repeated pruning and nested pruning, then computes the edit distance between two simplified XML documents. Although it can enhance the speed of the classification after simplifying the XML document tree, it ignores the importance of content of the XML document and it is less effective on classifying XML documents whose structure is more similar than content.

TABLE IV.
RECALL RESULTS

Method	Recall		
	K=5	*K=7*	*K=9*
NM-Similarity	0.9067	0.9289	0.9222
Method in [13]	0.4178	0.4756	0.5267
Method in [11]	0.5	0.5244	0.5178

In [13], the method is based on the KNN algorithm that relies on an edit distance measure. The approach explores both the content and the structure of XML documents for determining similarity among them. Compared to the method in [11], the method in [13] improves classification accuracy rate to a certain extent, but pruning make some important content nodes miss out, so it is less effective on classifying XML documents whose structure is more similar than content.

Using NM-Similarity in KNN, and setting the parameter λ to 0.2, 0.5, and 0.8 respectively, when let k=5, k=7 and k=9. We obtain the results displayed in table 5.

The data in table 5 shows that the content information in these datasets plays an important role on the performance of the clustering solution. As λ increases, the result becomes better. But when λ increases to a particular value, the result turns worse.

While the methods in [11] and [13] simplify the representation of XML document tree using pruning because XML documents are always large, the method in this paper does not use pruning. Consequently the methods in [11] and [13] perform better than method in this paper in the consideration of time complexity. However, many important information would be lost and hence the accuracy rate of classification would decrease because of pruning. So the method in this paper shows a significant improvement in accuracy rate of classification.

TABLE V.
ACCURACY AND RECALL RESULTS

K	NM-Similarity							
	λ =0. 2		*λ =0. 5*		*λ =0. 8*		*λ =0. 9*	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
K=5	0.6875	0.5844	0.7125	0.6977	0.9125	0.9067	0.8358	0.8535
K=7	0.7259	0.7257	0.8069	0.8275	0.925	0.9289	0.8525	0.8794
K=9	0.7439	0.7324	0.7846	0.7921	0.9125	0.9222	0.9105	0.9222

V. CONCLUSIONS

In this paper, we discussed an effective method for classifying XML data called NM-Similarity. This approach bases on k-nearest neighborhood algorithm which relies on Euclidean distance. It takes the structure

and content similarities of XML documents into account, improving the accuracy of classification, especially to the data whose structure is more similar than content. Differing from other methods, this method uses the advantage of Cosine measure to calculate content similarity. But NM-Similarity does not take the articles linked in the XML document into account. In further work, the articles linked in the XML document will be attached with great importance and k-nearest neighborhood algorithm will also be improved.

ACKNOWLEDGMENT

The authors would like to thank the members in the Database lab at Xiamen University for their helpful discussions and suggestions.

REFERENCES

[1] Joe Teklia,Richard Chbeir,and Kokou Yetongnon, "An overview on XML similarity: Background, current trends and future directions," Computer Science Review, vol 3, pp. 151-173, 2009.

[2] Ludovic Denoyer and Patrick Gallinari, "Overview of the INEX 2008 XML Mining Track: Categorization and Clustering of XML Documents in a Graph of Documents," Lecture Notes in Computer Science, vol 5631, pp. 401-411, September 2009.

[3] Inderjit S. Dhillon, Yuqiang Guan, Brian Kulis, "Kernel K-means,spectral clustering and normalized cuts," Proc of the 10 th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York: ACM Press, pp. 551-556, 2004.

[4] Sugato Basu, Mikhail Bilenko,Raymond J. Mooney. "A probabilistic framework for semi-supervised clustering," Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York:ACM Press, pp. 59-68.

[5] Nello Cristianini, Jaz Kandola, Andre Elisseeff and John Shawe-Taylor, "On kernel-target alignment," Studies in Fuzziness and Soft Computing, vol 194, pp. 205-256, 2006.

[6] Olivier Chapelle, Vladimir Vapnik, Olivier Bousquet and Sayan Mukherjee, "Choosing Multiple Parameters for Support Vector Machines," Machine Learning, vol 46, pp. 131-159, 2002.

[7] Wang Wenjian, Zongben Xua, Weizhen Luc and Xiaoyun Zhang, "Determination of the spread parameter in the gaussian kernel for classification and regression," Neuro Computing, vol 55, pp. 643-663, 2002.

[8] M. Zaki and C. Aggarwal, "Xrules: An effective structural classifier for xml data," Proc of the 10 th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York: ACM Press, pp. 316 – 325, 2003.

[9] TANG Kai, "Method of classification based on content and hierarchical structure for XML file," Computer Engineering and Applications, vol 43, pp. 168- 172, 2007.

[10] Theodore Dalamagas, Tao Cheng, Klaas Jan Winkel and Timos Sellis, "Clustering XML Documents by Structure," Lecture Notes in Computer Science, pp. 112-121, 2004.

[11] Gong An and Liu Huashan, "Improved Algorithm of XML Document Structural Clustering Based on Edit Distance," Microcomputer Applications, vol 29, pp. 88-91, 2008.

[12] A. Nierman and H. Jagadish, "Evaluating structural similarity in XML documents," In Proc. 5th Int. Workshop on the Web and Databases(WebDB), 2002.

[13] Abdelhamid Bouchachia and Marcus Hassler, "Classification of XML Documents," Proceedings of the 2007 IEEE Symposium on Computational Intelligence and Data Mining (CIDM 2007), pp. 390-396, 2007.

[14] Ludovic Denoyer and Patrick Gallinari, "The Wikipedia XML Corpus," ACM SIGIR Forum, New York: ACM Press, vol 40, pp. 64-69, 2006.

[15] Ludovic Denoyer and Patrick Gallinari, "Report on the XML Mining Track at INEX 2005 and INEX 2006 Categorization and Clustering of XML Documents," ACM SIGIR Forum, New York: ACM Press, vol 41, pp. 79 – 80, 2007.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15, August 2010, pp. 431-433

Software Package of Computer Network Course in Education

Qiao Yingxu [1],. Yang Hongguo [2]
[1] College of Computer Science & Technology, Henan Polytechnic University, Jiaozuo, 454000, China
qiaoyingxu@hpu.edu.cn
[2] Department of Computer Science & Technology JiaoZuo Teachers College Jiaozuo, 454000, China
yanghg@hpu.edu.cn

***Abstract*—The computer network course is the required course that college computer-related professional sets up , in regard to the current teaching condition, the teaching of this course has not formed a complete system , some new knowledge points can't be added in promptly while the outdated technology is still there in teaching .The article describes the current situation and maladies which appears in the university computer network related professional teaching ,the teaching systems schemes about the computer network course are presented.**

***Index Terms*—computer network, software,protocols**

I. INTRODUCTION

Computer network is a professional basic course, which must be learned by guaduated students in computer major. The Ministry of Education clearly states some documents about colleges teaching courses, which clearly poses that the computer network course teaching system should be divided into two directions,the professional and non-professional[1]. How to set up computer-related professional categories ofcomputer network course system becomes an important task of reforming computer network.

II. CURRENT TEACHING SITUATION

The computer network course mainly relates four modules including data communications, theory of network, application technique of computer network,

computer network safety and management. The computer course mainly takes the OSI model of the system frame of the computer network as theories ,take the practical TCP/IP framework as the main line.It narrates computer network-related layer model and protocol. It covers actual use of the relevant LAN and WAN technology.

At present, according to the computer network course teaching statistics, which offered by majority of domestic computer-related professional colleges,

There are no more than two ways of setting up courses.

(1)Data communication +Computer network

(pure theory)+ network technology (experiment);

(2)Data communication and Computer network (Theory+Experiment);

According to the statistical data analysis, college which adopt the way (1)is 45%,others which adopts the way (2) is 52%[2].The left 3% put forward the computer network course system structure systematically.

Although the way (1) expresses that the knowledge points of the computer network course is coherent,and it increases the proportion in the class hour,but the separation of computer network theory and experimental courses make it difficult to achieve the combination of theory and practice .The class hour of way (2) is short. The content of course covers a knowledge point that is numerous and jumbled,the experiment teaching

III ANALYSIS OF TEACHING CONTENT

The total period of the computer network course is 52, include 40 periods in prelection and 12 periods in experimentation[3]. The credit hour is 3. The teaching contents are shown as follow:

Chapter 1 introduces the set of core ideas that are used throughout the restof the text. Motivated bywidespread applications, it discusses what goes into network architecture, and it defines the quantitative performance metrics that often drive network design.

Chapter 2 surveys a wide range of low-level network technologies, ranging from Ethernet to token ring to wireless. It also describes many of the issues that all data link protocols must address, including encoding, framing, and error detection.

Chapter 3 introduces the basic models of switched networks (datagrams versus virtual circuits) and describes one prevalent switching technology in details. It also discusses the design of hardware-based switches.

Chapter 4 introduces internetworking and describes the key elements of the Internet Protocol (IP). A central question addressed in this chapter is how networks that scale to the size of the Internet are able to route packets.

Chapter 5 moves up to the transport level, describing both the Internet's Transmission Control Protocol (TCP) and Remote Procedure Call (RPC) used tobuild client/server applications in detail.

Chapter 6 discusses congestion control and resource allocation. The issues in this chapter cut across both the network level and the transport level. Of particular note, this chapter describes how congestion control works in TCP, and it introduces the mechanisms used by both the Internet and ATM to provide quality of service.

Chapter 7 considers the data sent through a network. This includes the problems of both presentation formatting and data compression. The discussion

of compression includes explanations of how MPEG video compression and MP3 audio compression work.

Chapter 8 discusses network security, ranging from an overview of cryptography protocols (DES, RSA, MD5), to protocols for security services authentication, digital signature, message integrity), to complete security systems(privacy enhanced email, IPSEC). The chapter also discusses pragmatic issues like firewalls.

Chapter 9 describes a representative sample of network applications and the protocols they use, including traditional applications like email and the Web, multimedia applications such as IP telephony and video streaming, and overlay networks like peer-to-peer file sharing and content distribution networks.

The experimentation is the important means for complementing the teaching contents in prelection.It could help the students consolidate the basic knowledge and the capability of program design. The current sets of exercises are of several different styles: Analytical exercises that ask the student to do simple algebraic calculations that demonstrate their understanding of fundamental relationships Design questions that ask the student to propose and evaluate protocols for various circumstances Hands-on questions that ask the student to write a few lines of code to test an idea or to experiment with an existing network utility library research questions that ask the student to learn more about a particular topic

The idiographic distribution of the periods in prelection and in experimentation was shown in Table 1.

Table 1: Distribution of the periods

contents	prelection	experimentation
foundation	2	
data link networks	4	2
packet switching	8	2
Internetworking	8	2
end-to-end protocols	4	2
resource allocation	6	2
End-to-end data	4	
Network security	4	2

IV MULTIMEDIA EDUATIONAL SOFTWARE

The computer software multimedia software has been done according to the need of didactical outline. The contents that are preciseness and abundance, and possess reasonable structure system and clearly arrangement combines closely with teaching book. The difficulty of this course is understanding and application for a great deal of complex theory.1200 cartoons has been made to help students understand the the theory and principle. Cartoon software can demonstrate procedure of the principle step by step, Teacher can control tenor of executed program ,as a result to help students understand these contents easily.

Take the “Go-Back-N Protocol “ for example, In a Go-Back-N (GBN) protocol, the sender is allowed to transmit multiple packets without waiting for an acknowledgment, but is constrained to have no more than some maximum allowable number, N, of unacknowledged packets in the pipeline. Figure 1 shows the sender's view of the range of sequence numbers in a GBN protocol. If we define base to be the sequence number of the oldest unacknowledged packet and next seqnum to be the smallest unused sequence number, then four intervals in the range of sequence numbers can be identified. Sequence numbers in the interval [0,base-1] correspond to packets that have already been transmitted and acknowledged. The interval [base, next seqnum-1] corresponds to packets that have been sent but not yet acknowledged. Sequence numbers in the interval [nextseqnum,base+N-1] can be used for packets that can be sent immediately, should data arrive from the upper layer. Finally, sequence numbers greater than or equal to base+N can not be used until an unacknowledged packet currently in the pipeline has been acknowledged.

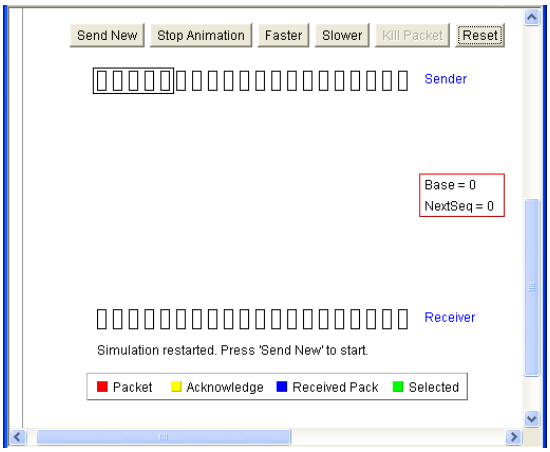


Fig1: the software of go-back-n protocol

Figure 2 and Figure 3 show the operation of the GBN protocol for the case of a window size of five packets. Because of this window size limitation, the sender sends packets 0 through 4 but then must wait for one or more of these packets to be acknowledged before proceeding. As each successive ACK (e.g., ACK0 and ACK1) is received, the window slides forwards and the sender can transmit one new packet (pkt4 and pkt5, respectively).

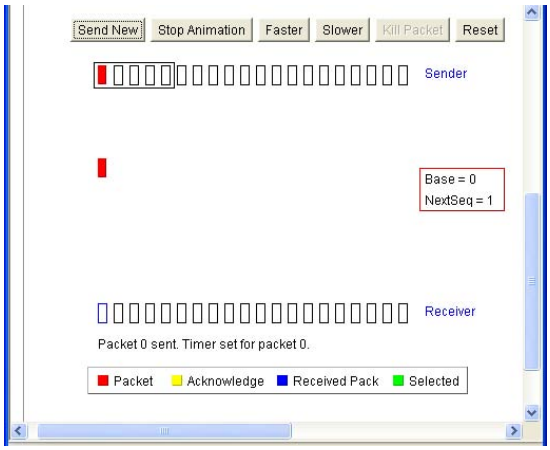


Fig2:Sending the first packet

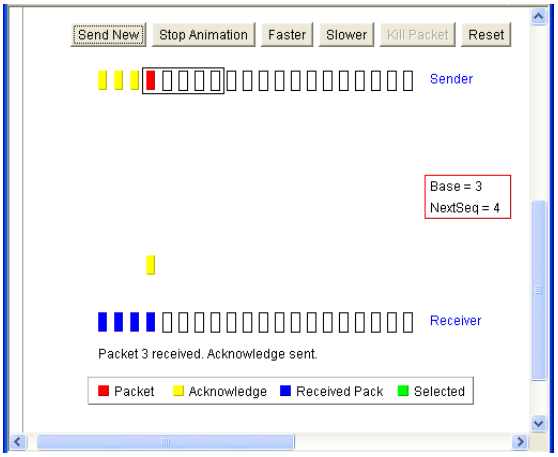


Fig 3: widows slider forward

Self-tests after class for each chapter are included in this software. After doing these self-tests, students master knowledge better. It makes 50 munites in class to prolong into after school to train student's ability of self-study. Operating this software possess good ability of tolerance and debug. It can be used after coping into computer without setup and tenor of cartoon demo can be controlled by "button" and be finished to exit at any time.

After adopting advanced multimedia technology , teaching work in class obtains good effect..

V CONCLUSIONS

After developing data structure multimedia software, abstract knowledge and procedure of theory and protocols are exhibited with the cartoon mode, accepted easily by students in the multimedia teaching. Thereby , students' enthusiasm in learning this course is exploded. This software is used in computer major 06,07,08 and information management 07,08 for computer network course, which obtains well didactical effect..

REFERENCES

[1] Sun Xiujuan,Deng Yun.iscussion on the Reform of the Computer Network[J] .Journal of Beijing Polytechnic College,2007,(04)

[2] WU Yi-zhi,XU Hong-an.Network Simulation Technology Aided Computer Networks Education[J].Journal of Donghua University (Natural Science,2007,(02)

[3] CHEN Ming-ming,U Xiao-yan,AN Ying.Research on Computer Network Course Teaching in University [J].Computer Knowledge and Technology,2007,(01)

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15, August 2010, pp. 434-437

Design of Core Modules on Three-dimensional Roadway Engine

Zhou Hong-bin[1], Liu De-jian[2], Wang Yu-kun[3]
[1] Wuyi Macro link Gas Limited Company, Jinhua, China
Email: 691334921@qq.com
[2] Institute of Computer Science & Technology, Henan Polytechnic University, Jiaozuo, China
Email: liudejianw@163.com, wyk@hpu.edu.cn

***Abstract*—In order to raise the development efficiency and quality of three dimensional tunnel roaming system, the paper utilizes object-oriented technology and underlying graphics interface to design three dimensional graphics engine which can be applied in the three dimensional laneway roaming system and other virtual reality systems. The paper proposes common framework model combination of object-oriented design patterns, focusing on the design of core module of the engine. The simulation results show that the three dimensional roadway based on engine pattern has higher degree of simulation and faster speed of roaming.**

***Index Terms*—three-dimensional roadway; engine; core modules; design patterns.**

I. INTRODUCTION

With the rapid development of virtual reality technology, high-realism, realistic three-dimensional nature of roadway roaming systems have become an urgent requirement for the industry and academia. Prior to research three-dimensional roadway roaming systems, mostly based on OpenGL, DirectX and other low-level graphical interface modeling and rendering mode of implementation. This three-dimensional graphics library, in the three-dimensional graphics rendering has outstanding performance advantages. But they were only calculated on the underlying graphics API interface library, object-oriented operational relatively poor. If you use these API interface directly to system development, not only require a long development cycle and the efficiency is relatively low, but also the reliability is low and the simulation is poor, the rate of roaming is also significantly low, does not meet the three-dimensional roadway practical requirements for the use of roaming systems and modern software development trends [1]. Therefore, using 3D engine technology can greatly improve development efficiency, is a new means of development and design.

Although the three-dimensional 3D engine technology developed fast at home and abroad, there are a lot of 3D products, but the cost is high, and the target is poor, it is difficult to find suitable systems in the development of three-dimensional roadway roaming products. Therefore, building a Three-dimensional graphics engine basic on graphics functions, application-oriented, modular is the key to develop the virtual reality Navigation System [2]. This research was designed to achieve the establishment of a subjective and interactive, high-realism and the ability to develop business in the common three-dimensional roadway structure of the engine, to discuss the structural model of core modules of the engine.

II. THE FRAMEWORK OF ROADWAY ENGINE

Three dimensional roadway roaming system is a large software engineering, which is made of by many modules. From developers' perspective, the roadway engine provides API needed for the roadway developers, which also offers some core libraries auxiliary means for developing. The roadway engine is divided into two parts: engine core and peripheral interface of the engine. The specific simulators interact with the engine by the peripheral interface, whereas the information conveyed is delivered over up to the core to deal with by engine.

The peripheral interface is subdivided into the following parts: the input system module is in charge of detecting the input to input device in the beginning of each frame, the results of which are given to the logical decision part of the engine to deal with; the audio system module takes charge of loading, playing the sound and sound effect; the message processing module is responsible for sending the events received to the main control module in the way of messages; GUI interface module is with responsibility for the interface display.

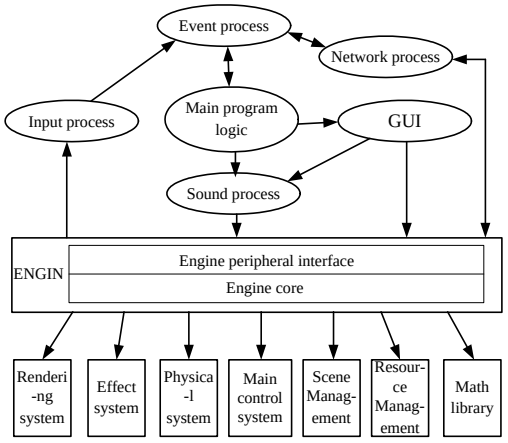


Figure 1. Conventional diagram of the framework model of laneway engine

The engine core module is subdivided into the following parts: rendering system takes charge of the parts of coordinate transformation, hidden-surface

removal, illumination and pigmentation, texture mapping and particle system; the physical system is in charge of detecting and responding to simulated collision of physical laws; the resource management system are given to manage the texture resources, the model resources and so on; the mathematics common module provides the points, the vector, the matrix and other three-dimensional mathematical operation. To enhance reusability, expandability and maintainability of the design, design pattern must be used synthetically to design each module. Use Factory pattern to load all kinds of resources; Use Composite pattern to manage the scene and organize resources; make use of Singleton pattern to guarantee the uniqueness of virtual tour characters; make use of Flyweight pattern to realize the resource sharing [3]; make use of observer pattern to realize the event processing and message routing. The general framework of roadway engine as illustrated in Fig.1.

III. CORE MODULES

By constructing the laneway engine, using object-oriented approach to rendering laneway, you can handle from the geometry level detached work out and instead deal with specific scenes and objects in the scene. In which object includes: movable objects (laneway car, roaming characters), composed of static objects in the scene itself (laneway scene itself), lights, cameras and others. Just putting the object into the scene, the engine will complete mess of the geometry rendering treatment to out of dependence on the API. Scene management module, rendering module, file system module and the main control module complete most of the functionality, are the core modules of the engine [4].

A. The design of the roadway engine's scene management module

Scene graph in the graphics engine, the position is without doubt, it not only provides space for users to find and search for objects provide high-speed optimization, but the need for rendering library, it provides the search, sort and remove function. Sometimes also used for collision detection. In some specific design inside, scene graph can even be used for all subsystems, such as voice and physical systems may rely on a scene graph to achieve the corresponding functions. OGRE reference in the laneway engine solution, discard the traditional design method, using a scene graph and scene design from content, the scene graph structure and its use of data nodes as equal inheritance system.

1) Scene graph structure

In the laneway engine we use the scene node to organize the laneway scene content; the scene node is the actual transforming elemental area in the scene. The scene node has own connection level (to have a father node and certain subnodes), the node operation supports three kind of different coordinate system spaces: World coordinate space, father node space and own space. Therefore, in the move, zoom, rotate, they can choose to use the coordinate space. Scene node can exist independent of the scene graph, where the scene shows the contents of independence and the benefits of scene graph: the content will not be affected by a specific scene graph; it can be re-attached to the other scene graph. Scene graph of the Scene management module is in Fig.2.

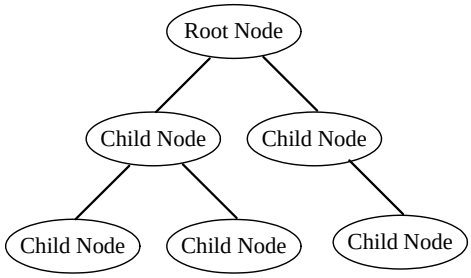


Figure 2. Scene graph of the Scene management module

In the laneway the concrete scene content needs to hang in the scene node to be able to display. The scene content has interior equipment, including: guide rail, car, and transformer substation and so on. Abstracting scene models to specific classes for the Mesh, Mesh class inherits entity (Entity Class); the entity inherits the active object (MovableObject), and created through the scene manager. After the scene content founded, binded to the scene node which already existed. When the scene contents attached to the scene node, you can use the scene node to manage the entity, and by changing the content of the scene node to change the scene. Scene management module in the integrated hierarchical graph is in Fig.3.

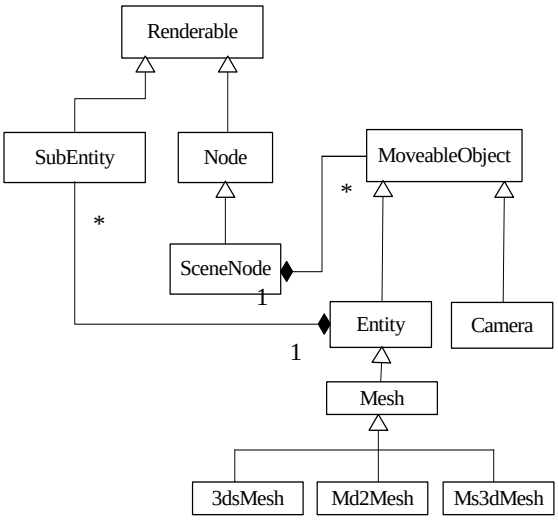


Figure 3. Class diagram of the Scene management module

And static models (laneway model itself, guide rail) and dynamic models (roaming character, car, electric lamp and so on) in the laneway may produce by the three dimensional modeling tools. After loaded into the engine and attached to the scene node, spatial relations and 3D transform (move, rotate, zoom) operation objectives are scene nodes rather than the scene content (content to follow the node with the transformation), then the scene nodes are Unified Management by SceneManagement class.

2) Scene management classes

Through constructing the SceneManager classes in the laneway engine, SceneManager class is the core class in the Scene management module, which is responsible for the creation and placement of the moving objects, lights and cameras in the scene, and maintains their travel in the scene graph and transform; Loading the laneway map of the whole model; on the scene to support queries, and will remove invisible objects and push visible objects into the rendering queue; According to current and exaggeration object perspective drawing, Organizing and sorting (by increasing distance) unidirectional lights from the perspective of the current renderable; Setup and rendering of any shadows in the scene, passing this organized content to the render system for rendering. According to the processing mode in OGRE, SceneManager class and interactive rendering system can complete the updating and rendering scenes. As the laneway level of detail of different scenes in different locations, different levels of focus rendering, SceneManager class abstract the interface, the specific details of the level of scene management to achieve by the sub-class, SceneManager realizes applies the Template pattern. SceneManager class of the Scene management module is in Fig.4.

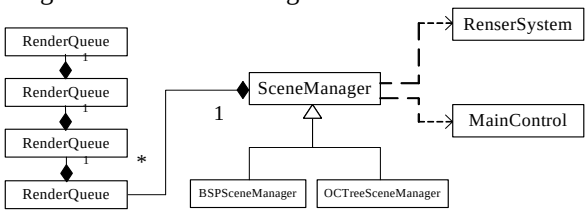


Figure 4. SceneManager class of the Scene management module

B. The design of the roadway engine's rendering module

1) the overall structure of the Rendering system

Rendering module is one of the most important functions to complete the model of the engine; it can finish model, animation, lighting, special effects and other comprehensive and display effect. In the laneway engin,renderTarget class is responsible for receiving the results of rendering operations, which can be a window on the screen can also be packaged FPS statistical information and also responsible for creating and maintaining Viewport. Viewport is a rectangular area; it is to get the contents of the mapping from the camera to render the above objectives. In the same render target can contain one or more viewports. Camera connects rendering and scene management system, inherits from MoveableObject and Renderable in SceneManager. RenderSystem is responsible to exaggerate system's supervisory work, is defined as the abstract class, and the realization is completed by subclass through the graph API, realized with the Template pattern in Fig.5.

2) Rendering flow

Rendering system and scene management module with the main control module complete the updating and rendering scenes. After the MainControl initialization completes,it calls StartRendering function to enter the message circulation, and then through the message system realized by the observer pattern and the responsibility chain pattern, calling UpdateAllTarget method in the RenderSystem to update all the RenderTarget. RenderTarget through the Update method to update all of the Viewport, Viewport and then calls the associated rendering method of Camera, then program is into the SceneManager's RenderScene method, through the calculation of the rendered scene to RenderSystem do real rendering.

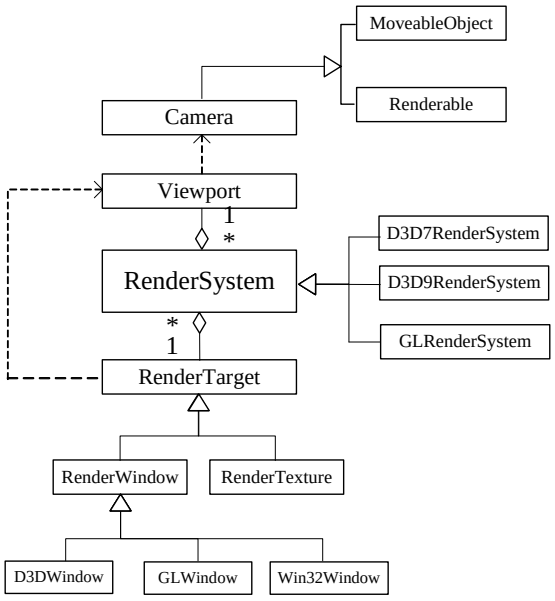


Figure 5. Class diagram of the Rendering system module

C. The design of the roadway engine's main control module

The main control module is in the core position in the system, driven by the message system ongoing implementation of the event logic processing and rendering, completes 3D scene rendering and moving objects. The main control module is the entire laneway engine's outward appearance class (realizes with Façade pattern) , may call each sub-system's interface in the engine through it; engine can be opened and closed by MainControl class, when the construction started the engine, the destructor method shuts down the engine. Processing message loop and dispatching events, the master control class realize with the SingleTon pattern, to ensure there is only one object and provide global access points [5]. Sequence diagram of the main control module is in Fig.6.

IV. CONCLUSION

The engine has played an important role in software development system; this paper builds an overall framework of the three-dimensional roadway system based on engine pattern to detach the reuse of code and modules in the systems so as to provide common solutions for the development of three-dimensional roadway system. Systems integration uses various design patterns to ensure system scalability. The paper focuses on the most important modules-the rendering modules

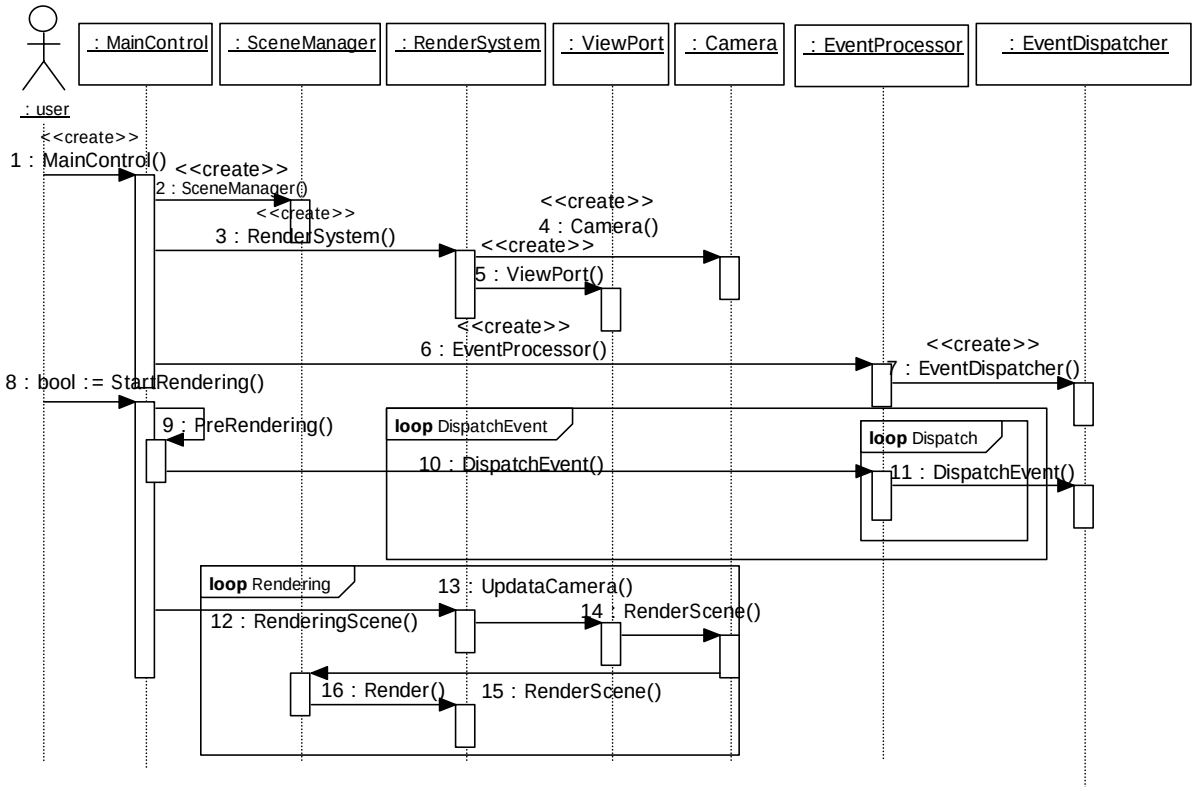


Figure 6. Sequence diagram of the Main control module

and the scene management modules and main control module.

ACKNOWLEDGMENT

The authors would like to thank School of Computer Science and Technology in Henan Polytechnic University for their sponsoring to the subject and all the numbers helpful for my paper.

REFERENCES

[1] ZHANG Mingmin, YANG Haoran, PAN Zhigeng. Virtual Walkthrough System Base on Game Engine [J]. Computer Science, 2007, 34(12):78-81.

[2] Li Zhaoming, MENG Xianfu. Analysis and Design of 3D Engine in Shoot Simulation System [J]. Computer Engineering, 2006, 32 (20):227-229.

[3] HE Zhiying, YU Jiangying. The Application and Research of Design Pattern in 3D Graphics Engine [J]. Journal of System Simulation, 2008, 20:158-161.

[4] LUO Guan, HAO Chong-Yang, HUAI Yong-Jian. Design and Implementation of a Virtual Reality Engine [J]. Computer Engineering, 2001, 24 (11):355-362.

[5] NIE Zhe, WEN Xiao-jun. Design and Implementation of Virtual Reality Engine Based on Visualization Technology [J]. Computer Engineering and Design, 2008, 29 (9) :2423-2425.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15, August 2010, pp. 438-442

The Research of Modulation Recognition Algorithm Based on Neural Network

Yanfang Hou[1], Hongmei Feng[2]
[1] School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, China
Email: yfhou@hpu.edu.cn
[2] School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, China
Email: fenghongmei@hpu.edu.cn

***Abstract*—Aimed at the non-stationary modulated signal of which signal to noise ratio (SNR) changes large, a modulation recognition algorithm based on neural network is presented. In this algorithm, instantaneous feature parameters of received signals are extracted using wavelet transform. Then, make singular value decomposition to the matrix which is composed of instantaneous parameters to get singular values. Error back propagation neural network (BPNN) with supervised training is to be made the classifier. The singular values obtained are used as feature vector and inputted to the classifier. So the automatic modulation recognition of signal is realized. The identification in category for FSK and PSK are simulated respectively, and the simulation results prove the approach proposed in this paper is efficient.**

***Index Terms*—recognition, feature vector extraction and matching, wavelet transform, singular value decomposition (SVD), error back propagation neural network (BPNN)**

I. INTRODUCTION

Modulation identification for communication signals is a very important research topic both in military and civilian areas. In the military field, mastery the exact modulation facilitates the identification of enemy threat and the decision on an appropriate interfering waveform. So a more targeted communication confrontation strategy can be made [1] [2]. In civilian areas, modulation identification for communication signals is one of key technologies for the non-cooperative communication.

Modulation recognition can be divided into two categories: identification between categories and identification in category. Identification between categories means the identification between different types of modulation mode, such as the recognition of ASK, FSK and PSK. The technology about it can be found in reference 3. And the identification in category means a further distinction of the same modulation mode, such as the recognition of BPSK, QPSK and 8PSK. For the identification in category, the traditional approach is using differences in amplitude to effectively distinguish MASK and using Fast Fourier Transform to distinguish MFSK effectively. But it still has not been found a very effective and simple classification method for the distinction of MPSK.

Communication signals are polluted by the channel noise in the process of transfer, so the signal received is time-varying and non-stable [4]. Wavelet transform has the characteristics of time-frequency localized quality and zoom-changed. And its results reflect a number of coefficients got by the wavelet decomposition, which contain plentiful and various character information of the signal. So it is particularly applicable to the analysis of non-stable signal, and has been widely used as a tool for feature extraction [5]. Neural network has characteristics of distributed information storage, large-scale self-adaptive parallel processing and a high degree of fault tolerance. Its learning ability and fault tolerance quality are unique to the uncertainty pattern recognition. The combination of wavelet transform and neural network can obtain a good quality of automatic identification for the signal, thus forming various approaches [6].

Based on the research of wavelet theory and neural network, a novel method for recognition of modulation signals is proposed. In this method, feature extraction is based on wavelet analysis and singular value decomposition while classification and recognition is based on neural network. The identification in category for FSK and PSK are simulated respectively, and the simulation results prove the approach proposed in this paper is efficient.

II. WAVELET TRANSFORM

Wavelet analysis is a time-frequency localized analysis method that the window size (viz. window area) is fixed but its shape can be changed, and both the time window and frequency window can be changed. That is, it has a higher frequency differentiate rate and a lower time differentiate rate in the low-frequency part while a higher time differentiate rate and a lower frequency differentiate rate in the high-frequency part [7]. The meaning of wavelet transform is: the function $\psi(t)$ known as the basic wavelet is made a displacement τ, and then inner-multiplied with the signal $x(t)$ in different scales α:

$$WT_x(\alpha,\tau)=\left\langle x(t),\psi_{\alpha,\tau}(t)\right\rangle=\frac{1}{\sqrt{\alpha}}\int_{-\infty}^{+\infty}x(t)\psi^*(\frac{t-\tau}{\alpha})dt \quad (1)$$

From the definition we can see that the basic wavelet has two parameters: scale α and displacement τ. So when a function was made wavelet transform, it means that a time function will be cast to the two-dimensional space: time-scale space. It is advantage to pick-up some essential characteristics of the signal. The wavelet

function used in wavelet analysis is not exclusive, that means wavelet function $\psi(t)$ has diversity. Therefore in engineering applications, a very important issue for wavelet analysis is the choice of optimal basic wavelet. At present the error among the results obtained by wavelet analysis and the theoretic results is used to judge the merits or demerits of the basic wavelet. So we can determine which one to choose. In this paper, Mexican Hat wavelet is used to realize the recognition of modulation signals.

III. ERROR BACK PROPAGATION NEURAL NETWORK (BPNN)

BP network namely the error back propagation neural network is the most widely used type of neural network models. Its topological structure adopts the three-tier feed-forward network structure mostly, including input layer, hidden layer and output layer. There are whole links between the layers, and no connection between neurons of each layer. The basic idea of BP learning algorithm is: learning process consists of two processes such as the positive transmission of signals and the back-propagation of errors. Given an input mode of network, it will produce an output mode by the disposal from input layer units to hidden layer units, and then from hidden layer units to output layer units. If the output doesn't agree with the expected output, it will turn to the error back propagation stage. Error back propagation transmits error signals from the hidden layer to the input layer in some form, and obtains the error signals of all units in every layer. The error signals will be used as the basis to modify the weight value of unit. This process will be carrying out up to the error of network decreases to a pre-determined value, or to a pre-determined number of training.

The training process of network is the process to adjust the weight value vector W in accordance with some rule, making error function E below the threshold that is required. Learning algorithm is actually based on the idea of linear approximation, so the convergent speed of the algorithm is slow. In order to accelerate the speed of convergence, learning algorithm-SCG (scaled conjugate gratitude) is proposed [8]. The algorithm based on the idea of second-order approximation to determine the search direction $d(k)$ and change the step $\alpha(k)$, thereby adjusting the weight value vector W. The idea of second-order approximation can be shown as follows:

$$E(W+\Delta W)\approx E(W)+E'(W)^T\Delta W+\frac{1}{2}\bullet\Delta W^T E''(W)\Delta W \quad (2)$$

The specific process of algorithm can be found in reference 8.

Trainscg uses scaling conjugate gradient back-propagation algorithm to train the network. It is a deformation of conjugate gradient algorithm. The algorithm combines the model confidence interval method in the Levenberg-Marquardt BP (LMBP) algorithm and conjugate gradient method. So it avoids line search process that time-consuming is huge, and improves the training speed of network. The improved BP algorithm is used in this paper for the learning and training of network.

IV. ALGORITHMIC IMPLEMENTATION

The modulation recognition algorithm based on neural network proposed in this paper consists of three parts. First is the extraction of instantaneous feature parameters. Adopt wavelet transform to get wavelet coefficients of signal in different scales α, and make the wavelet coefficients gained in different scales respectively compose coefficient matrix. The quantity of data for instantaneous parameters is large. In order to reduce them, and saving the signal information to the largest extent, so the extraction of feature vector should be realized. The specific method is: make SVD to the wavelet coefficient matrix obtained in the first stage, and identify all the nonzero singular values to be as the feature vector of signals. The last part is the feature matching namely the design of classifier. This part is implemented by BPNN. Using the feature vector of communication signal that modulation mode has been known and the expected output, adopt BP learning algorithm to adjust the weight values of BPNN and obtain the neural network classifier that has been trained. Extract the characters of unknown signal, and then put them into the neural network classifier obtained. Finally judge the modulation types of communication signal according to the output of classifier.

On the assumption that there is a communication signal sequence $x=\{x(n)\,|\,n=0,1,\cdots,L-1\}$ with a limited length L. To realize the classification of signals, algorithm can be designed as follows:

- Select appropriate basic wavelet function to make one-dimensional continuous wavelet transform to the sample signal sequence in positive, real scale respectively (α = 1, 2, 4, 8, 16, 32, 64, 128), and get wavelet coefficients.
- Make the wavelet coefficients gained in eight scales respectively compose coefficient matrix and make SVD to it to obtain the nonzero singular

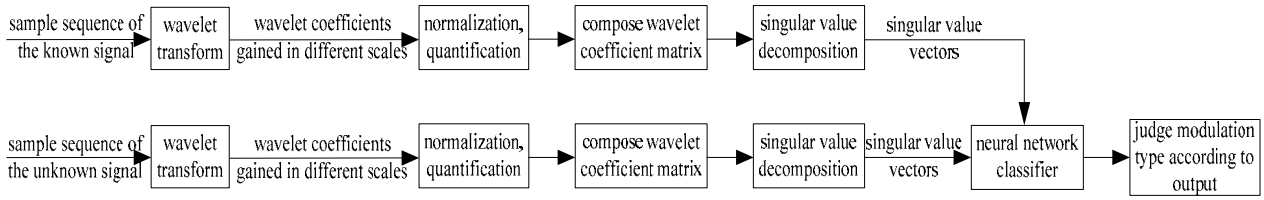


Fig.1 Flow chart of classification

values.

- Get the nonzero singular values as feature vector of signals, and input them to BPNN.
- Use the signal characters of which modulation mode has been known to train the network to obtain neural network classifier.
- Extract the characters of received unknown signal, and then put them into the neural network classifier that has been trained. Finally judge the modulation type of signal according to the output of classifier.

Flow chart of classification is shown as Fig.1.

In the training of neural network, the expected output vector can be set to the corresponding 0, 1 sequence. For example: for the recognition of modulation signal MFSK, when the input signal is BFSK, QFSK and 8FSK respectively, the expected output vectors should be set [1 0 0], [0 1 0] and [0 0 1]. That is the three nodes of output correspond to BFSK, QFSK and 8FSK signals respectively. The non-linear transformation function in BPNN used in this paper adopts continuous S-type function. So the actual output is a continuous vector rather than the corresponding 0, 1 sequence. Therefore it's necessary to make a judgment of 0 or 1 to each output

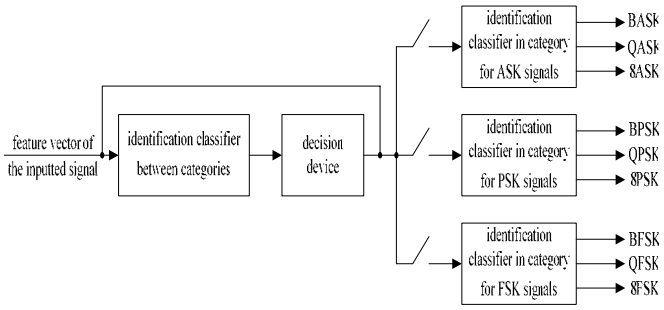


Fig.2 Specific composition of neural network classifier

value after the output vector has been achieved. Judgment rule is shown as follows:

$$y_i = \begin{cases} 1, (y_i \geq 0.5) \\ 0, (y_i < 0.5) \end{cases} \quad (3)$$

Based on the rule, modulation type of signal can be judged in accordance with the 0, 1 sequence of output. If there is more than one 1 or full 0 in the output nodes, we can't judge to which kind the signal belongs at the moment.

It's worth noting that the specific composition of neural network classifier in the flow chart of classification (shown in fig.1) can be expressed as fig.2.

From Fig.2 we can see that neural network classifier consists of three parts: identification classifier between categories, decision device and identification classifier in category. Use wavelet transform and SVD to get the characters of received unknown signal. The feature vector will be put into neural network classifier that already trained. First, the vector obtained is composed of 0, 1 sequence after feature vector goes through the identification classifier between categories. Receiving the vector, decision device judges to which modulation type the vector belongs, and closes the switch to the corresponding modulation identification classifier in category. Then, the link to the corresponding modulation identification classifier in category is on. So put feature vector of signal to the classifier can make a further distinction that in the same type of modulation.

V. EXPERIMENT ANALYSIS

Based on the classification method mentioned above, the identification in category for FSK and PSK will be simulated respectively. The results are obtained by MATLAB6.5.

A. Identification in category for FSK signals

Generate 1500 signals randomly as training signals, of which BFSK, QFSK and 8FSK is respectively 500. Carrier frequency f_c of the three signals is random variable that is uniformly distributed in [200MHz, 300MHz]. Sampling frequency $f_s = 800MHz$, code rate $f_d = 100MHz$, and the sampling points are 1024.

First process these signals in accordance with the algorithm proposed above. The number of non-zero singular values obtained after handling was all 8, so use 8 input nodes. As there are three types of signals to be classified, the number of output nodes is 3. 25 nodes are used as the intermediate nodes (the selection of intermediate nodes is relative to the performance of network and complexity of system. The more are intermediate nodes, the better classification results of network will be got. But at the same time the computing and complexity the system required will increase correspondingly. Specific selection should take both two factors in consideration). Training of classifier adopts SCG algorithm of BPNN to adjust the weight values of network. And the transfer function of neuron adopts logarithm S-type (Sigmoid) transfer function. Set the error threshold 0.001.

Table 1 RECOGNITION RATE IN NOISE-FREE CIRCUMSTANCES

	number of testing samples	number of recognized samples	recognition rate	average recognition rate
BFSK	1000	998	99.8%	99.1%
QFSK	1000	984	98.4%	
8FSK	1000	992	99.2%	

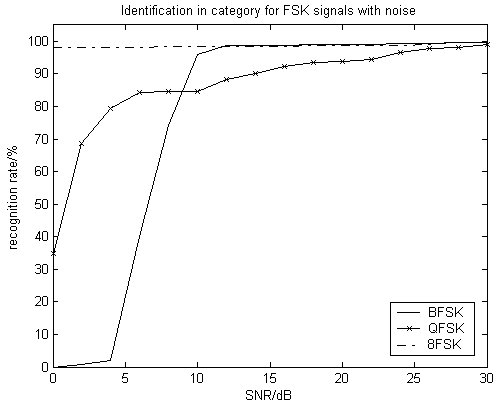


Fig.3 Identification in category for FSK signals with noise

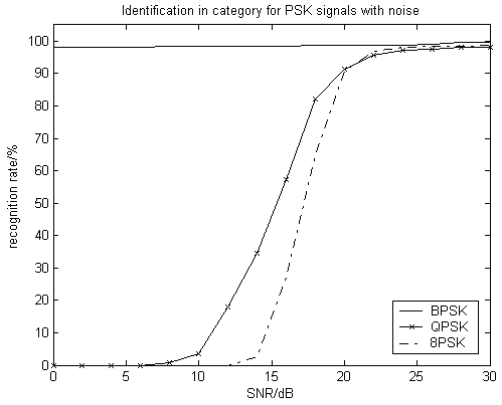


Fig.4 Identification in category for PSK signals with noise

Table 2 RECOGNITION RATE IN NOISE-FREE CIRCUMSTANCES

	number of testing samples	number of recognized samples	recognition rate	average recognition rate
BPSK	1000	993	99.3%	99.3%
QPSK	1000	987	98.7%	
8PSK	1000	998	99.8%	

Generate 3000 signals randomly as testing signals, of which BFSK, QFSK and 8FSK is respectively 1000. Modulation parameters of testing signals are consistent with training signals, and the information they carried is also a random sequence. After processed, they are put into the neural network classifier obtained previously. Then the judgment results can be got. Recognition rate in noise-free circumstances is shown in Table 1:

In order to test the recognition rate of this method under the circumstances of noise, signals with noise are simulated. 1000 samples for each type of signal are tested, and fig.3 shows the simulation results.

As can be seen from Fig.3, when the SNR is 10dB, recognition rates to the three types of signals are all more than 88%. The classification results are good. But when the SNR is lower than 10dB, recognition rate of BFSK declines fastest, and that of QFSK declines second. However 8FSK is basically out of the influence of SNR,

The simulation results got in the circumstances of noise are shown in Fig.4.

As can be seen from Fig.4, when the SNR is 20dB, recognition rates to the three types of signals are all more than 90%. The classification results are good. But when the SNR is lower than 20dB, recognition rate of 8PSK declines fastest, and that of QPSK declines second. However BPSK is basically out of the influence of SNR, and its recognition rate remains high. From feature vector to the three types of signals we can see that it is small for 8PSK, bigger for QPSK, and the biggest for BPSK. In the same SNR, the absolute values of impact caused by noise on the three types of signals are similar. Then the relative alues of impact decrease in the order: 8PSK, QPSK and BPSK. Therefore, the anti-noise performance of 8PSK is the worst, and that of QPSK is better. BPSK is almost out of the influence of noise, so its anti-noise performance is the best.

and its recognition rate remains high. The results above are decided by the classification algorithm proposed in the paper. From feature vector to the three types of signals we can see that it is small for BFSK, bigger for QFSK, and the biggest for 8FSK. In the same SNR, the absolute values of impact caused by noise on the three types of signals are similar. Then the relative values of impact decrease in the order: BFSK, QFSK and 8FSK. Therefore, the anti-noise performance of BFSK is the worst, and that of QFSK is better. 8FSK is almost out of the influence of noise, so its anti-noise performance is the best.

B. Identification in category for PSK signals

For the identification in category for PSK signals, the parameters and processing method are identical with those of FSK signals. Recognition rate in noise-free circumstances is shown in Table 2:

VI. CONCLUSION

Using wavelet to process the non-stationary modulated signal of which the SNR changes large. That means wavelet space is used as the feature space of pattern recognition. Instantaneous feature parameters of received signals are extracted using one-dimensional continuous wavelet transform. Then, make SVD to the matrix which is composed of instantaneous parameters to get singular values. BPNN with supervised training is to be made the classifier. The singular values obtained are used as

feature vector and inputted to the classifier. So the automatic modulation recognition of signal is realized. The identification in category for FSK and PSK are simulated respectively, and the simulation results prove the approach proposed in this paper is efficient. This method doesn't need to know any prior information, and doesn't need to make carrier frequency synchronized or signal parameter estimated before identification. The only thing it needs to do is obtain the sampling points of signal. The method is simple, requires less memory and its speed is faster. So in military, it can meet the requirements of real-time processing in communication counterwork and opens up a new route for communication reconnaissance. In civilian areas, it is also easy and quick to use.

ACKNOWLEDGMENT

This work was supported by the Youth Foundation of Henan Polytechnic University (No.646191).

REFERENCES

[1] Jiang Yuan, Zhang Zhao-Yang, and Qiu Pei-Liang, "Modulation classification of communication signals," Military Communications Conference Proceedings, IEEE, 2004, vol. 3, pp. 1470–1476.

[2] E. E. Azzouz and A. K. Nandi, "Automatic identification of communication signal modulation," Transl. Yu Rentao and Li Wugao, 57394 People's Liberation Army troops, pp. 3–6, September 1998.

[3] Fu Wei-hong, Yang Xiao-niu, and Zeng Xing-wen, "Novel method for blind recognition of communication signal based on time-frequency analysis and neural network," Signal Proccessing, Nov.2007, No.5.

[4] Wang Ming-san, "Principle of Communication counterwork," Beijing: Publishing House of PLA,1999, pp. 144～172.

[5] HONG Liang and HO K C, "Identification of digital modulation types using the wavelet transform," Military Communications Conference Proceedings, IEEE, 1999, Vol.1, pp. 427～431.

[6] Gao Meng, Zhao Pei-qing, and Zhang Fu-sheng, "A method of modulation identification for communication signals based on wavelet analysis and neural network," Journal of Shijiazhuang Railway Institute, Dec.2002, Vol.15, No.4, pp.32～36.

[7] R & D center of Feisi-technology Products, "Wavelet analysis theory and Matlab 7 achieved," Beijing : Publishing House of Electronics Industry,2005, pp.29～48.

[8] Moller, M. F, "A scaled conjugate gradient algorithm for fast supervised learning," Neural Networks, 1993. Vol. 6, pp. 525～533.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15, August 2010, pp. 443-446

A Comparative Study of Several Face Recognition Algorithms Based on PCA

Dong Xiaoqian, Huang Huan, and Wen Hongyan
Faculty of Information Engineeringand Automation
Kunming University of Science and Technology, Kunming, China
Email:dongxiaoqian1986@126.com, xhuan@21cn.com

***Abstract*—Principal component analysis (PCA) is one of the main methods for face recognition, and it does have some benefits and achieved some results. Although, PCA method also has some disadvantages, such as high images dimensionality, computational load and so on. In this paper, with consideration of 2DPCA and LDA with PCA, a study of PCA +2 DPCA and PCA + LDA based on PCA method is discussed. The experimental results show that PCA +2 DPCA and PCA + LDA are much more accurate and effective than traditional method that uses PCA only.**

***Index Terms*—PCA, 2DPCA, LDA, PCA+2DPCA, PCA+LDA, face recognition**

I. INTRODUCTION

Principal component analysis (PCA), two-dimensional principal component analysis (2DPCA), and linear distinguish analysis (LDA) are relatively common methods for face recognition which have achieved very good recognition rate, but they also have some defects. For example, when processing with people's face image with PCA method, we have to turn two-dimensional image matrix into one-dimensional column vector, then the dimension of the image is changed into multidimensional image matrix which complicated the calculations. 2DPCA feature in extract data to form a column number vector, but the speed of classification is lower than PCA. The computation of LDA is very complex, complicate, and easy to cause errors. Therefore, we introduced the PCA +2DPCA method and PCA+LDA method. With the comparison of experimental study, both of the methods are more accurate than method that uses PCA only.

II. PCA,2DPCA AND LDA

A. PCA algorithm

PCA (principal Components Analysis) [1-3] is derived from K-L transformation essentially. Its purpose is to find an optimal orthogonal transformation, and get the optimal orthogonal unit vectors which are called principal components, so we call this algorithm principal component analysis. It is the most widely used method of feature extraction in face recognition. The steps of this algorithm are as follows:

For a human face image, we connect to each of its columns and build up a column vector with the size of $D=M\times N$ dimensions. We suppose n is the number of training samples, x_i is the image vector of the i^{th} training sample , then the covariance matrix of the sample is:

$$S = \sum_{i=1}^{n}(x_j - \bar{x})(x_j - \bar{x})^T \quad (1)$$

In this formula, $\bar{x}$ is the average image vector of training samples, $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $Z = [x_1 - \bar{x},\ x_2 - \bar{x},\ \ldots,\ x_n - \bar{x}]$,and $S = ZZ^T$, the dimension is $D \times D$ .According to the theory of K-L transformation, the new coordinate system we need to obtain is composed of the non-zero vectors corresponding to non-zero eigenvalue of matrix S. By SVD theorem, we calculate the eigenvalue of ZZ^T through calculating the eigenvalue of Z^TZ ,then arrange the eigenvalue in descending order $\lambda_1 > \lambda_2 > \ldots > \lambda_n$, We define the eigenvectors corresponding m-largest eigenvalues as main elements, compose of these vectors into a subspace, we call it eigenface space. Then project training samples into the space of feature face, so we get a row of vector projections, then constitute these vector projections into recognition database. We select a image with concentrated testing samples and project it into the eigenface space, then compare their positions with the database of human faces by using the nearest neighbor classifier, so face can be recognited.Position figures and tables at the tops and bottoms of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Leave sufficient room between the figures/tables and the main text. Figure captions should be centered below the figures; table captions should be centered above. Avoid placing figures and tables before their first mention in the text. Use the abbreviation "Fig. 1," even at the beginning of a sentence.

To figure axis labels, use words rather than symbols. Do not label axes only with units. Do not label axes with a ratio of quantities and units. Figure labels should be legible, about 9-point type.

Color figures will be appearing only in online publication. All figures will be black and white graphs in print publication.

B. 2DPCA algorithm

2DPCA[4] is different from traditional methods based on PCA, it does not require to transform image into one-dimensional vector, and it calculates the eigenvectors of image covariance matrix based on two-dimensional matrix of the image directly. We calculate the eigenvalues and eigenvectors, and select several larger eigenvalue vector space structure of face, then project the training samples and testing samples of the image into the eigenface space matrix, get the characteristic matrix of the image, use the nearest neighbor classifier to determine the type of the testing samples.

C. LDA algorithm

Linear discriminant analysis (LDA) [5] is a algorithm which we select the Fisher criterion function in classical linear discriminant analysis, so sometimes linear discriminant analysis is also called Fisher linear discriminant analysis (FLDA). Its purpose is to extract the low-dimensional with the most discriminant ability high-dimensional feature space. These features can help make all the samples with the same type together and separate the samples into different types as many as possible, that is, it selects the features that make the ratio of between-class scatter matrix S_B and within-class scatter matrix S_W largest. Between-class scatter matrix and within-class scatter matrix are showed as follows:

$$\mathrm{S}_B = \sum_{i=1}^{M} N_i(\mu_i - \mu)(\mu_i - \mu)^T \tag{2}$$

$$\mathrm{S}_W = \sum_{i=1}^{M}\sum_{j=1}^{N_i} (x_j^i - \mu_i)(x_j^i - \mu_i)^T \tag{3}$$

Fisher criterion function is:

$$J(W) = \arg\max \left| \frac{W^T S_B W}{W^T S_W W} \right| \tag{4}$$

When S_W is a non-singular matrix, the feature vectors of $S_W^{-1}S_B$ are the column vectors of the best linear transformation that make $J(W)$ the largest, this group of vectors will be the feature vectors of feature subspace in FLDA.

III. PCA+2DPCA

Based on the above analysis, we can see that, we must transform the two-dimensional image matrix into one-dimensional column vectors when using PCA method to process face image. As the dimension of the image increasing, the computation gets more complex. 2DPCA method uses the original image to construct the image covariance matrix directly where computation is less than PCA method, but image feature extracted from 2DPCA is a column number vector, which causes the classification speed slower than PCA. To obtain a better method, we combine 2DPCA with PCA and call it PCA +2 DPCA [6]. We obtain the projection matrix by using 2DPCA first, compose the projection matrix of the sample training group, then extract the second feature by PCA to identify the human face, so both speed of feature extraction and classification speed are improved[7].

The group of training samples is: $\{S_j^i \in R^{m\times n}, i = 1,2,\ldots,\ \mathrm{M,j=1,2,\ldots,\ N}\}$, where i is the number of categories, that is the i^{th} individual; j is the j^{th} image; M shows the number of persons; N shows each person contains N images; K is the total number of samples, and K = MN.

Calculate the average image matrix of samples:

$$\mathrm{S}=\frac{1}{\mathrm{K}}\sum_{i=1}^{M}\ \sum_{j=1}^{N} S_j^i \tag{5}$$

Calculate the covariance matrix of samples:

$$\mathrm{Z}=\frac{1}{\mathrm{K}}\sum_{i=1}^{M}\ \sum_{j=1}^{N} (S_j^i - S)^{\mathrm{T}}(S_j^i - S) \tag{6}$$

We make eigenvalue decomposition of the sample's covariance matrix Z : $ZX_i = \lambda X_i$, select the larger eigenvalues $\lambda_1 \ldots \lambda_p$,and use the orthogonal eigenvectors $X_1 \ldots X_P$ for the projection space. We project training sample $\{S_j^i \in R^{m\times n}, i = 1,2,\ldots,\ \mathrm{M,j=1,2,\ldots,\ N}\}$ onto the space $x_1 \ldots x_P$, and get

$$\mathrm{Y}_j^i = [S_j^i X_1, \ldots,\ S_j^i X_p] = [\mathrm{Y}_j^i(1), \ldots, \mathrm{Y}_j^i(p) \in R^{m\times p}$$,

make up Y_j^i into a new training sample $x_j (j = 1,2,\ldots,\ \mathrm{K})$ where

$$x_j = \{\{\mathrm{Y}_1^1, \mathrm{Y}_2^1, \ldots,\ \mathrm{Y}_N^1\}, \{\mathrm{Y}_1^2, \mathrm{Y}_2^2, \ldots,\ \mathrm{Y}_N^2\}, \ldots, \{\mathrm{Y}_1^M, \mathrm{Y}_2^M, \ldots,\ \mathrm{Y}_N^M\}\}$$

Then we use PCA method to extract the second feature of this new sample, and get a group of projection feature vectors which we make them as recognition feature vectors so that the face recognition is completed.

IV. PCA+LDA

PCA method for face recognition are impacted by angle, illumination, size and expressions, which will lead to decreasing of the recognition rate .The process of LDA calculation calculates matrix repeatedly, and the computation is so complex that can easily lead to cumulative errors and affecting the accuracy. In normal circumstances, face recognition is always a problem with small sample, the numbers of training samples are much smaller than the dimensions of the image vector, and make the within-class scatter matrix always a singular matrix, which increase the difficultly of this method.

Therefore, we introduce a method which combines PCA with LDA algorithm, which is named PCA + LDA algorithm for face recognition[8]. We obtain the eigenspace of training samples through the PCA algorithm, and then calculate the eigenspace of LDA algorithm. Integrate the eigenspace of PCA and LDA into

one eigenspace, and get the integration eigenspace PCA+LDA algorithm. Training samples and testing samples are projected onto integration eigenspace separately and negotiation features are obtained, last we use the nearest neighbor criteria to complete the face recognition. The steps of particular algorithm are showed as follows:

a) Obtaine a subspace of eigenface by PCA method.

b) Calculate eigen subspace of LDA algorithm on the basis of step one.

c) Calculate the generalized eigenvalues and eigenvectors, which combine the best classification space.

d) Integrate the eigen subspace of PCA and LDA algorithm into a subspace of PCA+LDA algorithm.

e) Project training samples and testing samples onto integration subspace ,get the recognition features, use the nearest neighbor criteria to complete the face recognition.

V. RESULTS AND ANALYSIS OF THE EXPERIMENTS

This paper studied and discussed the similarity and difference of the recognition rates for face recognition by using PCA, PCA+2DPCA, PCA+LDA algorithm. The experiments are simulated on ORL face database, and it comprises 400 gray-level frontal view face images from 40 persons, each person with 10 images, each with the size of 112 × 92, and the gray level is 256. These images are obtained at the conditions of different times, lightings, facial expressions and changing facial materials. Simulations are conducted on MATLAB. For each algorithm, we take each person's first 5 images from ORL face database as training samples, and take the rest five images as testing samples. We compare the recognition rate of the various methods with different sizes of samples. All the experiments are carried out on a PC with 2.0GHz CPU and 512M memory. The following diagrams show the comparison of recognition rate of the three methods that discussed in this paper.

As showed in Figure 1, the usage of PCA+2DPCA and PCA + LDA method are more accurate than PCA. As the dimension increasing,the recognition rates of the three methods are increased gradually,but when dimensions

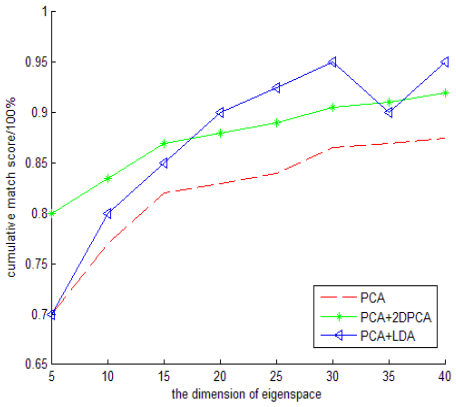


Fig.1 The relationship between cumulative match scores and the dimension of eigenspace

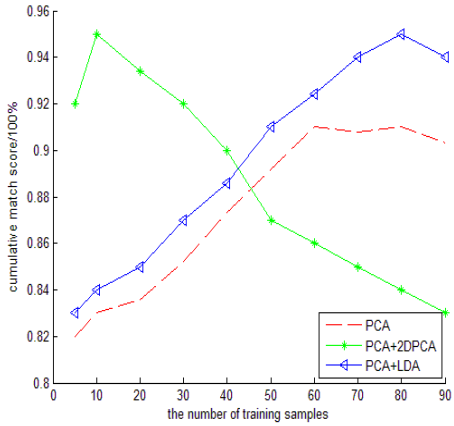


Fig.2 The relationship between cumulative match scores and the number of training samples

change to about 35, the recognition rate of PCA+LDA starts declining, which implies that the increasing of dimension does not necessarily lead to higher recognition rate. Figure 2 shows the relationship between cumulative match scores and the number of training samples when the dimension of eigenspace is 35. We find when processing small sample, PCA +2 DPCA is more accurate than PCA and PCA + LDA. As the number of the testing samples increasing, the recognition rate of PCA+2DPCA is decreasing. PCA+LDA is better than PCA and PCA+2DPCA significantly. Thus, PCA+2DPCA method is suitable for the problem of small sample, and PCA+LDA method is not influenced by the number of the testing samples.

VI. PCA+LDA

In this paper, we study the features of PCA, 2DPCA and LDA, and compare the three face recognition methods of PCA, PCA+2DPCA and PCA+LDA, then compare their correct rates of face recognition through simulation experiments. From the experimental results, we can see that PCA+2DPCA and PCA+LDA are much more accurate and effective than traditional method that uses PCA only. But the method of PCA +2 DPCA is more suitable for the problem of small sample, and PCA + LDA is much more accurate than PCA regardless of the number of samples. After all, PCA+LDA method has more theoretical significances in application.

REFERENCES

[1] Turk M A,Pentland A P.Face Recognition Using Eigenfaces[C].Proceedings of IEEE Computer Society Conference oil Computer Vision and Pattern Recognition,1991:586–591.

[2] Belhumeur P N,Hespanha J P,Kriegman D J.Eigenfaces VS.Fisherfaces :Recognition Using Class Specific I.inear Projection[J].IEEE Transaction on Pattern Analysis and Machine Intelligence,1997,19(7):711–720.

[3] Aleix M.Martinez,Avinash C Kak.PCA and LDA [J].IEEE Transaction on Pattern Analysis and Machine Intelligence, 2001,23(2)：228–233.

[4] Yang Jian,Avid Zhang.Two-dimensional PCA:A new approach to appearance-based face representation and

recognition[J]. IEEE Transaction on Pattem Analysis and M achine Intelligence, 2004,26(1):131–137.

[5] Belhumeour P N,Hespanha J P,Krlegman D J.Eigenfaces VS FLDA faces recognition using class specific linear projection[J]. IEEE Trans on PAMI,1997,19(7):711–720.

[6] Daoqiang Zhang, Zhi-Hua Zhou and Songcan Chen.Diagonal Principal Component Analysis for Face Recognition[J].Pattern Recognition,2006,39(1):140–142.

[7] Xingmin Qi,Guanmei Liu.A Comparative Study on Face Regonition Baesed On PCA[J].Modern Electronic Technique,2007,269（6）.77–79.

[8] Wei Wu,Jinhui Li.A Study of Face Regonition Baesed On PCA and LDA[J]. Science & Technology Information (Academic Research),2008(36).465–466.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15, August 2010, pp. 447-450

A Hybrid Structure of Spatial Index Based on Multi-Grid and QR-Tree

Guobin Li[1] ,Lin Li[2]
[1] School of Computer Science and Technology ,Henan Polytechnic University ,Jiaozuo ,China
Email:lgb99_99@163.com
[2] School of Computer Science and Technology ,Henan Polytechnic University , Jiaozuo ,China
Email:lilin0098@163.com

***Abstract*—Multi-level spatial index techniques are always used in the management of large spatial databases. In this paper, according to the characteristics of spatial index of grids and QR- tree, a hybrid structure of spatial index is presented based on multi-grid and QR-tree. Inspired by the theory of Spatial Information Multi-grid, rectangular geo-spatial is divided into a Multi-grid Index firstly, and then it is established spatial index for each small grid based on QR-tree. There shows the advantages of the structure and the algorithm. By algorithm analysis, it is showed that the indexing mechanism expends slightly larger space overhead for higher query performance, which has a good practical value.**

***Index Terms*—spatial index, grid index, QR-tree, hybrid index**

I. INTRODUCTION

Spatial index is a data structure according to a certain order, which based on the position and shape of spatial objects or a certain spatial relationships between spatial objects. Spatial index is a supplementary measure between the space objects and space operation algorithm, whose main objective is to screen and filter the spatial data. Through this measure, when we make spatial operations, a large number of spatial data is pre-excluded which is unrelated with the spatial object, so as to achieve the purpose of improving the efficiency of spatial operations [1].

Spatial index is the key technology to improve the performance of spatial database, which directly affects the efficiency of spatial data storage and the performance of spatial retrieval. Many scholars have done much research and have proposed a series of index structures on how to build a more effective spatial indexing. These structures have their own advantages. They solve some of the problems of spatial data index from different angles at different levels. Their actual effectiveness is often dependent on the specific structure of spatial data and methods of spatial data organization and storage. In terms of the current research of the spatial index, more representative include KD tree [2]、Quadtree [3-5]、R-tree and its improved [6] , the grid [7] and so on.

These methods of spatial index have their own advantages and disadvantages. By analyzing the advantages of grid index and QR-tree, this paper proposes a hybrid structure of spatial index based on multilevel grid [8] and QR-tree [9-10]. In this program, first, coarse mesh the rectangular geo-spatial repeatedly to establish a multi-level index, and then establish a QR-tree index for each sub-grid can not be further divided. Results show, in addition to storage space should be slightly larger, it's performance of insertion, deletion, query algorithm is superior to Grid Index and QR-tree.

II. REPRESENTATION OF HYBRID INDEX AND DATA STRUCTURES

A. Representation of Multi-level spatial Index

In actual applications, there are both exact queries and also non-exact queries which are raised by user. We need to do a division of non-exact queries, and then the results will be obtained. But for exact queries, if we only divide into M × N small pieces, it often fails to check requirements. For example, we query a point entity whether falls within a polygon or not. Even if they are stored in the same block, but that does not mean to meet the requirements, only show that they are nothing more than closer.

In order to achieve the requirements of precise queries, only make M and N is large enough, and even small pieces can not be divided. But it is obvious that, when the M and N is too large, the space and time efficiency would have become lower. Therefore, in order to improve the efficiency, we can divide the small pieces of the one level grid by using multilevel grid strategy. Blocks can be divided into a number of levels, but if the rank is too much, it will bring too much overhead of storage space and the reduction of time efficiency. For this reason, in the specific applications, we need formulate different index series according to the actual situation, which can be specified by the user, or optimized by certain conditions.

A specific partition method is that the entire space is divided into blocks of m_1 row and n_1 column by first-level grid division. Each block can also carry out a second-level division which is divided into lower blocks of any number of rows and columns. The number of lower blocks can be different. It depends on the actual need that whether each block is divided into the next level. Dynamic lists which point the next level division through the pointer can be used to store the structure. If the current block is no longer divided, the establishment of multi-grid index for this block is finished, and then a

QR-tree index is established for each block to build the hybrid index.

C	3	01	03	04

Figure 1. the relational table about cross-block object

B. Indexing Mechanism and Data Structures

This index structure we proposed is a multi-level hybrid index structure combined by multi-level Grid Index and the QR-tree index. First, the entire space is divided into N sub-spaces by multi-grid, and then we use QR-tree to search the final sub-index space. Its core idea is decompose a "Large" QR-tree into many "small" QR-trees. This approach has not only reduced the average depth of QR-trees, but also reduced the overlap of the index space in order to improve the overall search performance.

Rectangular geo-spatial will be divided into certain number of blocks by multi-grid index structure. Every block can be regarded a bucket. We can establish a QR-tree index for certain number of objects which MBR completely falling into a bucket. The bucket stores the pointer which point to the root node of QR-tree. And the index information of cross-block objects is also stored in the corresponding bucket directly.

We can set Block[i] for the objects in a bucket, S_Block[i] for the set which stored the index information of cross-block objects in the bucket, Block[i].QR-Tree for the corresponding QR-tree of Block[i].

(1) We set P_OBJ_x for an arbitrary point object, Oid_x for its unique identifier and $Point_x$ for its coordinate. If P_OBJ_x Block[i], $<Oid_x,Point_x>$ Block[i].QR-Tree.

(2) We set L_OBJ_x for an arbitrary line object, Oid_x for its unique identifier, L_MBR_x for its minimum bounding rectangle and $\{l_1, l_2, \ldots, l_m\}$ for the set of L_MBR_x across bucket number. If m=1, $<Oid_x, L\_MBR_x>$ Block[i].QR-Tree. If m>1, $<Oid_x,L\_MBR_x>$ S_Block[i] for $\forall$ $\{l_1,l_2,\ldots,l_m\}$.

(3) We set S_OBJ_x for an arbitrary range object, Oidx for its unique identifier, S_MBR_x for its minimum bounding rectangle and$\{s_1,s_2, \ldots,s_m\}$ for the set of S_MBR_x across bucket number. If m=1, $<Oid_x, R\_MBR_x>$ Block[i].QRTree. If m>1, $<Oid_x, S\_MBR_x>$ S_Block[i] for $\forall$ $\{s_1, s_2, \ldots, s_m\}$.

The data structure of this index consists of a number of bucket arrays, a single linked list contains the index information of cross-block objects, and a pointer contains the index information of QR-tree root node. If a block will be divided further, we can obtain a number of second-level blocks which stored in the region of the first-level block pointer point to after the second-level division. If the block doesn't contain a cross-block object, the pointer points to the single linked list is null. If the block doesn't completely contain any object, the pointer field of the QR-tree corresponding to the block is null. If the block completely contains several objects, the pointer field of the QR-tree corresponding to the block point to the root node of the QR-tree.

In order to manage the cross-block objects conveniently, we establish the relational table about cross block objects and its corresponding bucket. The table records Oid for the unique identifier of cross-block object, Num for objects across the numbers of barrels and $Block\_No_i$ for objects across the serial numbers of buckets, as shown in the following figure. In the operation of cross-block objects, we can find the object and its corresponding bucket directly according to the table to improve operational efficiency.

III. MAJOR OPERATIONS ALGORITHMS

A. Insertion Algorithm

Insertion process involves the point object's coordinate insertion process and the line and range's bounding rectangle insertion process.

As follows:

(1)The insertion algorithm of the point object

Input: point(x, y) for the coordinate of the point object and Oid for the unique identifier of the object

Output: the index with punctuate logo and its feature information

Search the corresponding block number i of the point(x, y) in the dynamic linked list.

We can find the corresponding bucket according to Block[i]. If the pointer which point to the root node of QR-tree is null, establish the QR-tree which root node is the new node. Then link the new QR-tree to the bucket. Go to .

If the pointer which point to the root node of QR-tree is not null, calculate the R-tree corresponding to the point object. If the leaf node can accommodate the index entry, the index entry of the new node could be inserted into leaf nodes directly. If the leaf node overflow resulting node split, use the split algorithm in the reference [11].After the node split, use the R-tree adjustment algorithm in the reference [11].If the node split communicate upward, it leads to root node split generating a new root node. Then modify the pointer which point to the root node of R-tree, making it point to the new root node. The insertion algorithm is used the method in the reference [11].

End algorithm.

(2)The insertion algorithm of the line object or range object

Input: mbr (min_x ,min_y ,Max_x ,max_y) for the object's minimum bounding rectangle and Oid for its unique identifier

Output: the index with logo of the line or range object and its feature information

Calculate i and j for the corresponding block of lower left and upper right corner of the object's bounding rectangle, also k and l for its upper left and lower right corner.

If i=j or k=l, the rectangle completely falls into the same block. We can use the algorithm similar to the insertion algorithm of the point object. Go to .

If $i \neq j$ and $k \neq l$, the rectangle falls into multiple blocks which set is { i , i+1, …l,l+1, … ,k,k+1 , … , j}. The set is marked as Set_B. Each block corresponds to a bucket, then the bucket can be expressed as Block[x]

(x Set_B).The new node index entry is inserted into Block[x] separately, at the same time the Oid of the object and its relational bucket number write to the relational table about cross-block objects and its corresponding bucket.

End algorithm.

B. Deletion Algorithm

Deletion algorithm includes the deletion of point object's index information and deletion of line or range object's index information.

As follows:

(1)The deletion algorithm of the point object

Input: point(x, y) for the coordinate of the point object to be deleted and Oid for the unique identifier of the object

Output: none

Search the corresponding block number i of the point(x, y) in the dynamic linked list.

We can find the corresponding bucket according to Block[i]. Then take out the pointer pointing to the QR-tree. If the pointer is null, go to , or else calculate the R-tree corresponding to the point object. Using the R-tree deletion algorithm in the reference [11], the index entry of the object can be deleted from R-tree.

End algorithm.

(2) The deletion algorithm of the line or range object

Input: mbr (min_x ,min_y ,Max_x ,max_y) for the minimum bounding rectangle of the object to be deleted and Oid for its unique identifier

Output: none

Calculate i and j for the corresponding block of lower left and upper right corner of the object's bounding rectangle.

If i=j or k=l, the rectangle completely falls into the same block. We can use the algorithm similar to the deletion algorithm of the point object. Go to .

If $i \neq j$ and $k \neq l$, the rectangle falls into multiple blocks. According to Oid of the object, we can search the block number which the object crosses from the relational table about cross-block objects and its corresponding bucket. Then delete the information of the object index entry in each bucket.

End algorithm.

C. Query Algorithm

Query algorithm is more complicated, we only discuss the queries based on window query here. Given a query window S, we set Set_B for the set of the block number which covered by query window, and Set_Block(i)(i Set_B) for the set of the bucket numbers corresponding to the query window.

Query algorithm is described as follows:

Input: S(x_1, y_1, x_2, y_2) for query window, (x_1, y_1) for the lower left corner coordinates of the query window,(x_2,y_2) for the upper right corner coordinates

Output: Set_S for the set of the index entries of the object which fall into the window

Put Set_S = Φ, and find out the set of bucket numbers corresponding to the query window S Set_Block (i) (i Set_B).

If Set_Block (i) = Φ, go to .

Take any Block[x] . Set_Block(i) , if S $\cap$ Block[x] .QR-Tree. MBR $\neq \Phi$, it will have to traverse all of the quadtree nodes, to find out the R-tree corresponding to the window targets. Using the region query algorithm in the reference [11], we can find out the index entries of the objects which fall into the query window from the leaf nodes of the R-tree. Then bring them into Set_B.

Find out S_Block [x].

If S_Block[x] = Φ, Set_Block(i) $\Leftarrow$ Set_Block(i)-S_Block[x]. Go to .

Take an index entry of arbitrary cross-block object<Oid_j ,MBR_j> S _Block[x].If S $\cap$ $MBR_j \neq \Phi$, Set_S $\Leftarrow$ Set_S {< Oid_j ,MBR_j >} .

Set_Block[x] $\Leftarrow$ Set_Block[x]-(<Oid_j ,MBR_j>), go to .

Output Set_S, end algorithm.

IV. ANALYSIS OF THE CAPABILITY

The technology this index used is mesh the rectangular geo-spatial repeatedly first, thus the number of cross-block objects in every grid reduced greatly, in order to reduce the duplication of storage. For the objects which completely contained in each grid, we use QR-tree to store them. This based on object segmentation way further reduces the storage overhead.

Multi-level grid index can help you to locate the object you want to query, just need simple address calculation and a small amount of disk access. Its query speed is fastest. But because of duplication of storage, the insertion and deletion of the object is time-consuming. So we combine multi-level grid index with QR-tree which efficiency of insertion, deletion and query is outstanding, even superior to R-tree. The performance of QR-tree is proportional to the depth of quadtree. Deeper the quadtree is, better the performance is. However, the storage overhead of QR-tree and the depth of quadtree is proportional too. So in practical applications, we should select the appropriate depth for the quadtree to exchange for higher performance.

Overall, this index structure has better time complexity and space complexity. It expends slightly larger space overhead for higher query performance. Especially for large-scale spatial databases, it has an excellent practical value.

REFERENCES

[1] Dean Luo，Liqiong Liao . Spatial Object Boundary Index Based on Wide Grids [J]. Journal of Southwest Jiaotong University，2003；38(6)：271-275.

[2] Yu Liu, Zhongying Zhu ,Songjiao Shi . Novel Strategy for Spatial k-NN Query[J]. Journal of Shanghai Jiaotong University, 2001；35(9):1298-1302.

[3] Samet H. The quadtree and related hierarchical data structures[J] .Computing Surveys,1984；16(2):187-260.

[4] Shaffer C, Samet H. Optimal quadtree construction algorithm [J] .Computer Vision, Graphics and Image Processing ,1987；37:402-419.

[5] Jianya Gong. A Natural Digits Based Linear Quadtree

Encoding[J]. Cartographica Sinica, 1992；21(2): 90-98.
[6] Xiutao Cui,Jianping Wu .Development of Approach in Storage and Index for Spatial Vector Data[J]. Remote Sensing Technology and Application,2002； 17(4):215-219.
[7] Nievergelt J , Hinterberger H , Sevcik KC.The grid file:an adaptable symmetric multi2key file structures[J].ACM Transactions on Database Systems,1984;9 (1):38-71.
[8] Jing Guo,Wei Guo,Zhiyong Hu. QR-tree:An Efficient Spatial Indexing Structure for GIS with Very Large Spatial Database[J]. Geomatics and Information Science ofWuhan University, 2003 , 28(3) :306 - 310.
[9] Jianhua Qiu,Xuebing Tang,Huaguo Huang. An Index Structure Based Quad-tree and R * -tree—QR * -tree[J] .Computer Applications,2003,23(8):124 – 126,152
[10] Deren Li, Xinyan Zhu, Jianya Gong. From Digital Map to Spatial Information Multi-grid——A Thought of Spatial Information Multi-grid Theory[J]. Geomatics and Information Science of Wuhan University, 2003,28 (6) : 643～650.
[11] Guttman A. R - Trees :A Dynamic Index Structure for Spatial Searching[C]. Proc of ACM SIGMOD. Boston :ACM Press ,1984 :47 - 57.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15, August 2010, pp. 451-454

Multi-scale Representation of Global Vector Data on Sphere Based on Map Accuracy

Fang Lin[1], Hu Bailin[2]

[1] Anhui Xuancheng hydropower construction survey and design institute, Xuancheng, China
Email: fanglinlonglong@126.com
[2] School of Surveying and mapping, Henan Polytechnic University, Jiaozuo, China
Email: hpuhbl@qq.com

***Abstract*—With the application of large scale even global scale geospatial data, the multi-scale representation of global vector data on sphere becomes very important. In order to represent the multi-scale global vector data efficiently and accurately, a real-time global vector data simplification method based on map accuracy is presented in this paper. Firstly, this method generates simplification thresholds according to the principle of map accuracy and the distance between viewpoint and the 3D global surface. Then, an improved-Li-Openshaw algorithm and Daglaus-Peauker algorithm are used to simplify line objects in two different directions in the data multi-scale representation process. Finally, an experiment with ESRI shp file data of the roads in China is given. The result illustrates that the method can simplify line objects efficiently, and the simplification ratio can reach about 80% or higher. The result is good and receivable.**

***Index Terms*—Global GIS, multi-scale representation, global vector data, spherical space**

I. INTRODUCTION

In recently years, the large scale even global scale geospatial data for analyzing and decision-making are required in many applications, such as global environmental monitoring, meteorological forecasting, sustainable development and utilization of resources, national security, and digital earth and so on. In order to avoid some significant drawbacks caused by projecting mode of traditional GIS, such as geometric distortions, data discontinuity and inconsistency of spatial relation, etc [Lukatela 2000; Kolar 2004], the geo-spatial data must be expressed on the spherical surface directly. It becomes necessary to construct a global oriented geography information system, namely Global GIS.

One of the key technologies for constructing a Global GIS is representing multi-scale geo-spatial data on the global surface. Related researches have been done in recently decade, and some software systems have been developed, such as Google Earth, WorldWind, and ArcGlobe, etc. However, these systems provide much raster data multi-scale representation and visualization function, while vector data multi-scale representation and visualization function are weaker. As a result, spatial entities related functions are few[Goodchild 2008; SUN Min, etc. 2008; Wang Zhipeng 2008b], which are very important for most applications in GIS, such as interactive retrieval, query, operating, analyzing and so on. If vector data could be efficiently represented on the sphere, spatial entities related functions will be achieved, the capabilities of those virtual geo-space systems would be significantly extended. At the same time, problems of global multi-source, multi-dimensional, multi-type, multi-resolution data seamless integration, representation, analyzing, unified treatment and sharing will be solved, and the application of digital earth will be more extensive and deeper. It is well known that the quantity of the global spatial data is huge. In order to improve the efficiency of operations, it should be simplified during the process of multi-scale representation and visualization. So the main aim of our work is to study simplification methods of global vector data when they will be multi-scale represented on the sphere.

II. RELATED WORKS

A lot of works have been done about the vector data simplification in traditional 2D and general small scale 3D Euclidean space environment, but related works are very few in sphere environment. Yet a spherical surface space is not topologically equivalent to Euclidean space in geometry, methods used in Euclidean space can not be used directly in a spherical surface. Some approaches to simplify vector data on the sphere have been suggested. These approaches can be classified two kinds of simplification method: based on spherical grid and based on viewpoint.

Simplification based on spherical grid. The principle is: combining vector data with spherical grid, and take grid unit as the filter, choose one point to stand for all points in the same grid unit, and linked these selected points as the new lines or polygons; the vector lines or polygons would be simplified when the spherical grid levels changed from high to low. The typical works are Dutton [1997] presented a simplification strategy by combined vector curves with QTM and JAO Jian [2005] proposed a new algorithm based on Dutton's. Dutton took local curvature as the parameter to control points selected, overcame the defect of selecting points mechanically, but related computation, non-line conversion, classification and storage made the simplification become complicated. JAO Jian selected a lower grid level as the initiative level to choose points. This initiative level grid unit was larger than the scheduled target QTM level. She judges the line is zigzag or smooth according point numbers in the grid unit. If there are many points in the gird unit, the line is zigzag, or the line is smooth. If the line is smooth,

selecting the midpoint of the point range to stand for all points in the same grid unit. If the line is zigzag, subdividing the grid, up to the scheduled target level. This method is easier and the simplification efficiency is higher. But it is little reliance that judging the line is zigzag or smooth according point numbers in the grid unit. Summarizing the two methods above, they both can simplify the data to a certain degree. They both have weakness. Firstly, there would be obvious difference between the simplified curves with the origin one because some feature points would be ignored during simplification process. Secondly, self-intersection phenomenon appears sometimes.

Simplification based on viewpoint. View-dependent simplification methods are based the principle of eye observation. “The farer it is small, the closer it is big. The closer can be seen the details, the farer only can be seen the outline.” Sun Min, et al [2007] proposed a QIP method for online vector data 3D visualization which based on the contrived priority and automatic priority of points to control real time global data generalization process. Wang Zhipeng, et al [2008a; 2008b] presented a view-dependent simplification algorithm. This algorithm used screen space projection error as the threshold realized the multi-resolution representation of the vector curve on sphere. But screen space projection error is intimately related with the screen pixels, simplification result maybe different if the screen pixel’s size is different.

Comparing these two kinds of approaches, it is easy to realize simplification by combining vector data with spherical grid. Simplification methods based on viewpoint corresponding with the principle of human observation, it can give us a fine visual effect, and can be understood and used easily. But these works seldom care for the representation accuracy. Representation accuracy determines the accuracy of retrieval, query, operating, analyzing and so on, we should pay much attention to it during simplification. In this paper, a real-time global vector data simplification algorithm based on map accuracy will be presented. In order to simplify the data simplification problem, in this research work, curve objects are mainly discussed in this paper .

III.The Principle And Algonithm Implementation

A.The principle of algorithm.

When making a topographic map, those targets whose size is larger than 0.1mm will be drawn on the map, whose size is smaller than 0.1mm won’t be drawn on the map. Choosing the 0.1mm as the threshold is based on vision principle. The reason is that our eyes’ Minimal Angle of Resolution is 1′, the smallest size is 0.1mm which our eyes can make out when the distance between our eyes and the target is about 0.5m, as shown as Fig. 1.

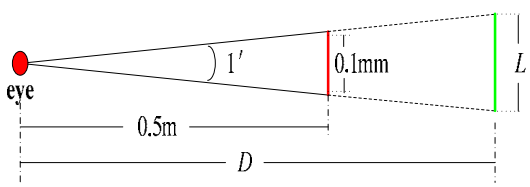


Figure 1. The principle of map accuracy.

The viewpoint is our eyes in a 3D visualization system. According to the vision principle and Fig 1 we will have this equation:

$$\frac{0.1mm}{Lm}=\frac{0.5m}{Dm}\Rightarrow L=0.0002\times Dm \qquad (1)$$

where D is the distance between the viewpoint and the target, and L is the smallest target size which viewpoint can make out.

That is to say the target will not be made out if its size is less than L when the distance between the viewpoint and the target is more than D, it is unnecessary to be expressed during the representation process. If these targets are eliminated during representation, data will be simplified. This is the principle of the global vector data simplification algorithm based on map accuracy.

B.The method implementation procedure

Li-OpenShaw algorithm is famous data simplification algorithm which based on a natural principle of objective generalization, so this research chooses it as the basic algorithm. Li Z. L. and Openshow[1992,1993] have described the simplification procedure in detail. But during the simplification procedure using Li-OpenShaw algorithm, sometimes the simplification-circle will intersect with multi parts of the curve, especially at the bottleneck, which will cause several different simplification results, even significant difference with the original. As shown as Fig.2. [Ying Shen, 2002]

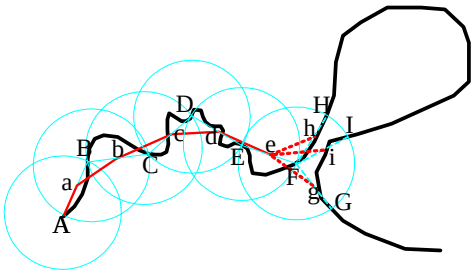


Figure 2 Li-Openshaw algorithm and the bottleneck problem

To overcome this deficiency, an improvement is done for Li-OpenShaw algorithm in this paper. During simplification procedure, comparing the arc length between points with the threshold directly instead of the simplification circle. Several different simplification result phenomenon will never appear.

Using the improved Li-OpenShaw algorithm and map accuracy, for a curve on the sphere surface,, the simplification procedure as follows:

(1) Firstly, saving the first point and the last point of the curve, and defining two index pointers namely *Location_pointer* and *Move_pointer* and letting them points to the beginning two points. Then measuring the distance D viewpoint to global surface, and generating the simplification threshold L according the principle of map accuracy and D.

(2) Measuring the minor arc length *l* of the great circle which through the two points *Location_pointer* and *Move_pointer* point to.

(3) Comparing *l* with *L*. If *l* is shorter than *L*, eliminating the point that *Move_pointer* points to, moving *Move_pointer* a point toward the end. Or else, moving both *Location_pointer* and *Move_pointer* a point toward the end, and saving the point which *Location _pointer* points to. Go to the (2) until *Move_pointer* points to the last point. Fig 3 shows the simplification procedure.

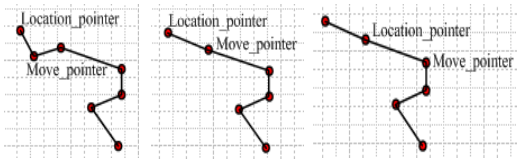


Figure 3 The simplification procedure based on improved Li-OpenShaw algorithm and map accuracy

From the simplification procedure above, it is obvious that that the improved Li-OpenShaw algorithm simplified the curve in curve direction only; in curve vertical direction did no thing. As Fig 4 shown, if the curve is a relatively flat curve, and the length *l*s of the minor arc between points are all longer than *L*, all points will be saved. In fact, if the minor arc length *l*s between points P_2, P_3, P_4, etc. and the minor arc of the great circle which through P_1 and P_n is shorter than *L*, the minor arc through P_1 and P_n can substitute for the curve. So another simplification algorithm which can simplify in curve vertical direction is needed. Douglas-Peucker algorithm is a classical data simplification algorithm. It can keep general characteristics and fractal dimension of the curve, and simplify curves in curve vertical direction effectively. So Douglas-Peucker algorithm is imported for further simplification.

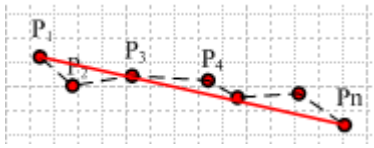


Figure 4 Using an arc instead of a curve

For a curve and a distance *D*, the total simplification procedure can be described as this: Firstly, generating the simplification threshold *L* according to map accuracy principle and the distance *D* viewpoint to sphere surface. Then using the improved Li-Openshaw algorithm to simplify the curve in curve direction for the first time, and using the Douglas-Peucker algorithm with the same threshold *L* to simplify the curve which has been simplified by the improved Li-Openshaw algorithm for the second time in curve vertical direction. Thus the curve has been simplified twice in two directions, those unnecessary or beyond expression points will be eliminated mostly. Repeating the above process when the distance *D* changes, the data are real-time simplified.

It is important to note that all middle points except the first and the last of a curve may be eliminated if *D* increase without limit. There will be a quite different between simplification result with the original data, which will bring significant errors for analyzing and measurement. In addition, if *D* is very long, the virtual 3D earth will be very small; this is no real meaning. Thus we should set the maximum value for *D* according to the purpose and scope of the work.

IV. EXPENIMENT

In order to test the method, one vector data set about road network data in China is used, the data set has 310057 points. These data are rendered on 3D global surface in real time. 11 viewpoint distances to 3D global surface are selected, and point numbers after simplification and simplification ratios are respectively recorded, the result shows in Table.1. Obviously, the data are simplified at different degree with different viewpoint distances. When the viewpoint distance to global surface increases, point numbers after simplification decrease rapidly and the simplification ratio increase quickly. The simplification ratio reaches about 80% when the distance increases near to the predefined maximum value 20000km.It is to say, four out of five of points are eliminated. 6 pictures are copied respectively from 6 viewpoint distances to one area during viewpoint running away from global surface (shown as Fig.5).

V. CONCLUSION

With the application of large scale even global scale geospatial data, the multi-scale representation of global vector data on sphere becomes very important. Data simplification is one of important aspects for realizing multi-scale representation and fast visualization. This real-time global vector data simplification algorithm based on map accuracy, it does double simplification for a curve: The improved Li-Openshaw algorithm simplifies the data in curve direction and Daglaus-Peauker algorithm in curve vertical direction,which improves the simplification ratio rapidly which can reaches about 80% or higher. In addition, it uses the principle of map accuracy generating the simplification threshold which assures the expression precision, and then assures the precision of interactive retrieval, query and analysis. In a word, it can simplify the data efficiently while maintaining high accuracy.

Although the algorithm is effective, there are also some defects: the simplification result is not very good if the line is very long or it is far away from the vertical line from viewpoint to global surface. In addition, Daglaus-Peauker algorithm will cause self-intersection sometimes and bring about wrong topological relation. All these problems need being further studied.

ACKNOWLEDGMENT

The authors hereby deliver their thanks to National Science Foundation of China for its financial supports to this research work of project No: 40771169 and 40701152.

Table. 1. Point numbers after simplification and simplification ratiosratios

Viewpoint distance to global surface(km)	190.6	1009.7	1855.0	3078.9	4124.1	6094.7	8004.7	10556.7	13184.8	17062.7	19422.1
Point numbers after simplification	272646	260500	246108	221811	199743	164224	138921	114204	95928	76347	68102
simplification ratio	12.07%	15.98%	20.63%	28.46%	35.58%	47.03%	55.19%	63.17%	69.06%	75.38%	78.04%

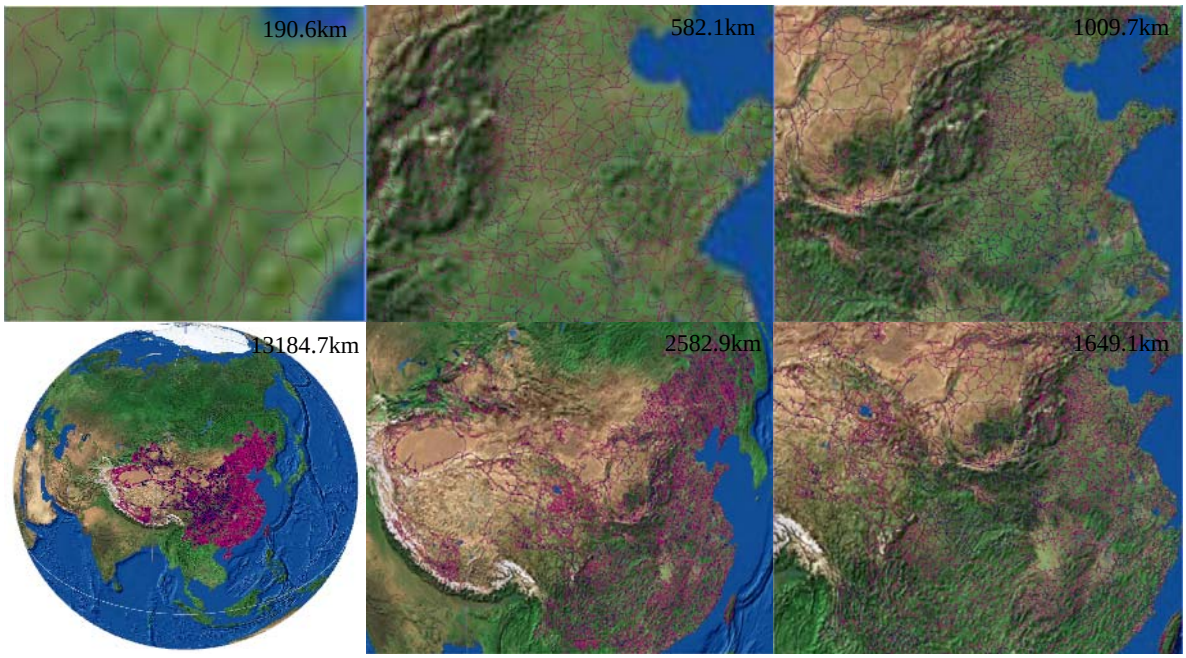


Figure 5. The screen shot sequence from different viewpoint distance to global surface

REFERENCES

[1] Dutton G. "Digital Map Generalization Using a Hierarchical Coordinate System". Proc.Auto Carto 13 ACSM-ASPRS '97 Technical Papers. Washington, 267-376(1997).

[2] J IAO Jian, WEI Lili and ZENG Qiming. "Algorithm for automated line simplif ication based on QTM ". Science of Surveying and Mapping, Vol.30 (5), 0089-0091(2005).

[3] Kolar, J. "Representation of The Geographic Terrain Surface Using Global Indexing", *Proceeding of 12th International Conference on Geoinformatics*, Gavle, Sweden: 321–328, (2004)

[4] Li Z.-L., Openshow S. "Algorithms for Automated Line Generalization Based on a Natural Principle of Objective Generalization". INT. J. Geographical Information Systems, Vol. 6(5): 373 – 389(1992),

[5] Li, Z.L., Openshaw, S.A. "Natural Principle for the Objective Generalization of Digital Maps", Cartography and Geographic Information Systems, Vol.20 (1):19-29(1993).

[6] Lukatela H, http://www.ncgia.ucsb.edu/globalgrids/papers

[7] Sun min, Wang Zhipeng, Yin Dan, and Wu Huan. "A Spatial Data Generalization Method for Online Vector Data 3D Visualization, " The 15th International Conference on GeoInformatics, Nanjing, Vol.2, 293-301 (2007).

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15, August 2010, pp. 455-457

Wireless CO sensor in mine hardware design based on ZigBee

Wang Wei[1],Shang Hua[2], Li Changqing[2]
1 Gas Institute of Geology Safety Institute Henan Polytechnic University, Jiaozuo, China
Email:shangh-04@163.com
[2] School of Computer Science & Technology Henan Polytechnic University, Jiaozuo, China
Email: cnwangwei@hpu.edu.cn , shangh04@yahoo.cn

***Abstract*—ZigBee is an emerging short-range, low-rate wireless network technology, which is a proposal cross between wireless tag technology and Bluetooth. The ZigBee agreement based on Wireless Sensor Networks overcome the past shortcomings in sensor network routing, for example, Wiring problems, poor effects of real-time detection, poor performance on proofing explosion. Based on the analysis of ZigBee technology in an underground mine wireless transmission, an electrochemical co wireless sensor has been designed, AT89C52 and wireless transceiver chip CC2420 constitutes a wireless transmission module，the sensor can display concentration values real-time, transfinite sound and light alarm, etc, it has strong anti-interference ability,it is suitable to be installed underground where needs wireless co sensor.**

***Index Terms*—ZigBee, wireless sensor network, AT89C52, cc2420, electrochemistry CO sensor**

I. INTRODUCTION

Carbon monoxide is a colorless, tasteless, odorless gas, lighter than air, and can be uniformly mixed with air, when the air CO concentration range from 13 to 75 percent risk of explosion. Explosion, spontaneous combustion of coal, mine fires and coal dust explosion caused by gas produced a major source of carbon monoxide underground [1]. as a key technical indicators of forecasting and detection of spontaneous combustion of coal,by the addition of carbon monoxide sensors in the mine and monitoring carbon monoxide concentration real-time is necessary. Existing mine carbon monoxide sensor multi-use wired technology to monitor carbon monoxide levels, Such programs scalability, cumbersome wiring, affecting appearance. Due to hard-wire connection, the line was easy to aging or corrosion, rat bite, abrasion, high incidence of failure. Wireless carbon monoxide sensor constructioned by wireless transmission can just avoid these problems. And, for more flexible, it avoid the trouble of re-wiring ,the network infrastructure is no longer needed buried in the ground or hidden in the wall, you can adapt to the needs of changing or moving.

II. ZIGBEE TECHNOLOGY

ZigBee uses 2.4GHz frequency band, ZigBee's PHY layer uses direct sequence spread spectrum (DSSS) technology, which has a Advantage of covering, confidentiality, strong anti-interference, resistance to multipath interference, increase system capacity, etc, particularly suitable for underground communication, can resist the interference of underground equipment and the environment;ZigBee has a advantage of low power consumption, low-cost , which happens to apply to more tortuous, strict restrictions on power supply, the shortage of funds, etc in mine[2]. In view of these advantages, in order to reduce coal mine hazards, contain gas accidents, full and unattended real-time monitoring to carbon monoxide concentration, this paper designed electrochemical wireless carbon monoxide sensor based on ZigBee.

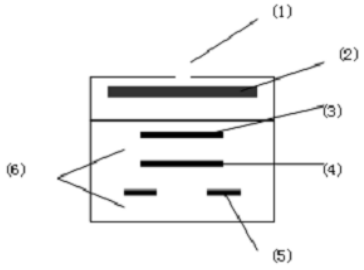
A. ZigBee-based wireless sensor networks

Node realization of the mechanism of ZigBee-based wireless sensor networks[3]is use ZigBee transfer module instead of the traditional serial communication module to sent collected information date wirelessly. The node includes ZigBee wireless communication module, micro-controller module, sensor module, DC power supply module.

.SENSITIVE ELEMENT COMPOSITION AND WORKING PRINCIPLE OF SENSOR

A. The composition of sensing element

Composition sensing element shown in Figure 1, Here electrochemical wireless carbon monoxide sensor developed by constant potential electrolysis type working principle, composed by the ventilation holes sensors, filters, electrolyte, working electrode, counter electrode and reference electrode. Holes is the gas channel, that CO, O2 and other gases can go through, but it is not only to prevent the leaking electrolytic tanks, but also to prevent the infiltration of water vapor outside the electrolytic cell, so it needs a layer of porous hydrophobic strong plastic film. Filter absorbed organic molecules in the air mainly by physical adsorption to prevent these



（1） Ventilation holes;（2）Filters;（3）Working electrode;
（4）Counter electrode;（5）Reference electrode;（6）Electrolyte

Figure 1. Compositions of CO sensor

substances from get into the sensor to contaminate electrode and impact the electrolyte performance, so the extension of the sensor filter plays an important role in extending the life of the sensor, stableing the state of the sensor. Working electrode, counter electrode contains catalytic activity of metal particles to carbon monoxide, its coated in a breathable buthydrophobic membrane, reference electrode contains gold particles that has a relatively low catalytic activity to carbon monoxide but chemical stability.

B. Work

Reference electrode does not participate in redox reactions, but can the working electrode and reference electrode potential make the stable between the working electrode and reference electrode potential. Gas containing carbon monoxide diffusion through the permeable membrane to contain the catalyst film on the working electrode, under the constant potential, in the gas, liquid and solid phase redox reaction between the interface and the electrolyte, generate current at the external circuit While reaction, Measurement is conducted between the working electrode and reference electrode, This allows the counter electrode potential change, which does not affect the measurement of the working electrode. When the carbon monoxide gas go through the semi-permeable membrane and get into the sensor, after the occurrence of redox reactions, redox reaction occurs:
Anode:

$$CO + H_2O \rightarrow CO_2 + 2H^- + 2e^- \quad (1)$$

Cathode:

$$1/2O_2 + 2H^- + 2e^- \rightarrow H_2O \quad (2)$$

According to Fick diffusion law, electrolysis current produced between the WE-CE is:

I-electrolytic current ,A;n- Transfer of electrons,2;F-Faraday constant, 96500c / mol; A-surface area of proliferation, m2; D-Diffusion coefficient of CO, m2/s; C-CO concentration, mol / m3; 6- Diffusion layer thickness,m.

For n, F, A, D and 6 are fixed values, therefore, the electrolytic current I proportional to CO concentration in gas. Therefore, as long as the electrolytic current I measured, the CO concentration ,C can be known.

. HARDWARE CIRCUIT DESIGN

A. Wireless sensor networks

Shown as Figure 2 ,Sensor node is make of the sensor modules, processor modules, wireless communication module and power supply modules. Sensor module is responsible for information collection and data conversion; Processor module is responsible for control of the sensor nodes, processing the collected data and the data sent by other nodes; Wireless communication module is responsible for wireless communication with other sensor nodes, the exchange of controling

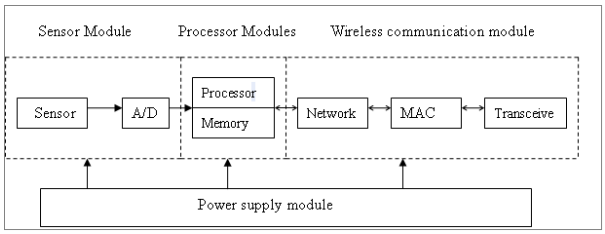


Figure 2. Sensor node hardware architecture

information, send and receive collected data ; Power supply module provide the energy required for the sensor nodes, in a miniature battery.

Wireless communication module select CC2420 produced by TI, it is a high performance, low power RF chip which the baseband processing and CSMA function integrated in inter, control it Through the SPI interface to achieve data transceiver.

B. Hardware design of mining wireless carbon monoxide sensor

The system of mining wireless carbon monoxide sensor make AT89C52[4]for the MCU, the SPI bus connects AT89C52 with wireless transceiver chip CC2420[5] , the two constitute a wireless transmission module. Other hardware circuit connected by the amplifier circuit, A / D converter circuit, LED display circuit[6], sound and light alarm circuit[7], infrared remote control circuit [8] and other components. Hardware circuit shown in Figure 3:

AT89C52 is low-voltage, high-performance CMOS 8 bit microcontroller produced by ATMEL in the U.S, with 8K bytes read-only Flash program memory that can be repeated erase and 256bytes random data memory (RAM), devices is produced by high-density nonvolatile memory technology in ATMEL, compatible with the standard MCS-51 instruction set and 8052 products pins, applicable to many more complex control applications.

1) Signal conditioning circuit

signal detection circuit designed using AD8572 with a zero-drift, single supply, rail to rail op amp,to ensure that the detection sensor within the signal sensitivity, stability and linearity.

2) A / D converter circuit

This paper adopts AD integrated circuit MAX197 that with multi-channel, multi-range input, it has an internal clock and the reference voltage, the sampling rate up to

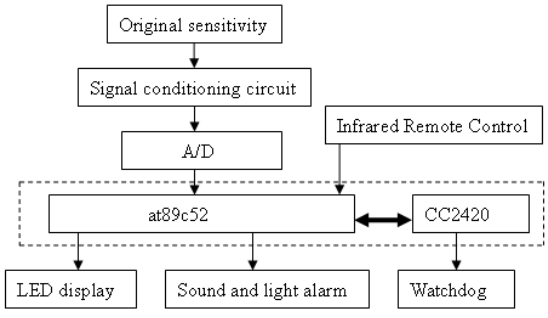


Figure 3. hardware components of CO sensor in mine

100 kHz, to meet the design needs.Need only four interface lines with MCU- Selection of films, read, write, and the high number of controls, and with a line more than 8-bit AD; analog input signal is 0 ~ 5 V, the actual control in the 0.5 ~ 4.5 V; with internal reference voltage, typical values +4.096 V.

Clock Circuit

The system uses internal self-excited oscillation, AT89C52 crystal up to 24MHz, using on-chip oscillator circuit, the XTAL1 and XTAL2 jumper at both ends of the crystal oscillator and two capacitors C1, C2 to form a stable self-excited oscillator, capacitor.

Sound and light alarm circuit

Components using piezoelectric buzzer sounds, Just needs add their two lead 3V ~ 24V DC voltage, P2.4 connects to Transistor base input through 74LS14, when the P2.4 output low "0", the transistor ,both ends of the piezoelectric buzzer was buzzing about +12 V voltage; When P2.4 output high level "1", the transistor out of conduction state, the buzzer sound stops.

Infrared Remote Control circuit

System uses infrared communication mode, TOSHIBA's TC9148P[9]infrared modulation transmitter chip with the speed and continuously firing function. Remote control only has four keys, namely: identification, function, plus (+), subtraction (-).The "OK", "function" two keys for the single button, add (+)"," minus (-) "two keys for the continuous bond. Infrared receiver circuit uses an integrated IR receiver HS0038, the data output can be directly connected to the external interrupt input of AT89C52 (in this system connected to P3.2), using the second function of the port, once the infrared signal coming, P3.2 is pulled low, MCU suspend the current work and transfer to the receiving, processing infrared signals, decoding the binary coded waveform sent by the infrared receiving, restore the date coming from the transmitter.

Ⅴ. CONCLUSIONS

ZigBee-based wireless sensor network system of carbon monoxide has the advantage of low cost, low power consumption, significant commonality, With the further development of wireless network technology, it will be mine development trend of carbon monoxide sensors. Although ,so far there is not applications that fully support ZigBee, there is reason to believe that ZigBee will lead the short-distance wireless communication network, and have very broad application prospects.

REFERENCES

[1] Bang-Chao Yang, Duan Jianhua. Monoxide sensor applications and progress [J]. Sensor Technology. 2001, 20(12):1～2.

[2] LiLi,ZigBee Technology[J].Technology and Market, 2009, 5(16).

[3] Li-Min Sun. Wireless sensor networks. Beijing: Tsinghua University Press, 2005 .

[4] YuTao, Shuang-Xi Liu. X25045 color change in stage lighting design in the application [J]. Journal of Computer Applications .2001 (7): 66 ~ 69.

[5] WangShuai, Safety monitoring system design based wireless sensor networks in mine[J]. Electrical Technology. 2008(10).

[6] Li Qi, Jin Zhi, Wu Chi power. Based on Single Chip LED display control system design [J]. Computer .2009,25 (2): 110.

[7] Xin-min Pan, Yan-Fang Wang. Micro-computer control technology. Higher Education Press, 2001.

[8] Qi-wen Liu, Zou Xu-chun, Wang Wei. Microcontroller on the universal infrared remote control transmitter signal decoding [J]. Practical Testing .2001 (6): 33 ~ 34.

[9] Hai-TaoDu.Smart Gas Sensor System Design and Research[D].JiaoZuo: Henan Polytechnic University,2004.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15, August 2010, pp. 458-462

Study on the Framework of College Student Honesty-credit Evaluation System

Xingxiang Qi[1, 2]
[1] Shandong Economic University, Jinan, China
[2] Fashion College of Donghua University, Shanghai, China
Email: qixingxiang@163.com

***Abstract*—Honesty-credit is vital to shaping the morality. Colleges and universities are the important base and effective body for the honesty-credit education. In this paper, a college student honesty-credit evaluation system is introduced, which not only directs, dominates and inspires the behaviors of student, but also is concise and easy to operate. The framework of management information system that contributes to implement the evaluation system is described in detail.**

***Index Terms*—college student, honesty-credit, evaluation system, management information system**

I. INTRODUCTION

Honesty-credit is vital to shaping the morality, which are the moral standards that conduct inter-relations between people, and which is the foundation and essence of all morality. College students, a significant part of human society, are precious human resources and hope of country's development. In order to preferably shoulder the important task of society in the future, colleges and universities should educate students to be good character citizens who both abide by the principle of honesty-credit and honor a pact.

Colleges and universities are the important base and effective body for the honesty-credit education. Honesty-credit education is the indispensable part of the moral education in colleges and universities, and honesty-credit education is the important component of social honesty-credit education. Shandong Economic University (SEU) is an institute of finance and economics, whose graduates in a great measure are on the post of duty related to accountant, public finance and monetary economics. These occupations and characteristics of the position require practitioner to have good personal cultivation and honesty-credit. Since the founding of our university, we always give high priority to facilitating honesty-credit education of college students. The whole system of honesty-credit education has three parts. Firstly, the content system of honesty-credit education is completely built up. Secondly, the implemental system of honesty-credit education has become mature, such as the class education that is the main channel, the campus culture sports education that is based on many different activities, the conventional education in the daily life, the special education in the educational process, and the open education with social interaction. Thirdly, supervision and evaluation system of honesty-credit education was established, of which there are four aspects: learning honesty-credit, economic honesty-credit, living honesty-credit, social honesty-credit [1].

II. HONESTY-CREDIT EVALUATION

The evaluation of college student honesty-credit is to synthetically estimate the honesty-credit condition of student according to student's behavior in school. The key to this evaluation lies in designing a set of science reasonable evaluation system. The evaluation system not only directs, dominates and inspires the behaviors of student, but also is concise and easy to operate. SEU have established a system on the status of college student honesty-credit [2].

A. Evaluation Methodology

The evaluation system of college student honesty-credit has two portions. One is records management. The other is democratic appraise. The main task of records management is to record the key behaviors listed in the table of college student honesty-credit evaluation. There are four parts in the table: learning honesty-credit score, economic honesty-credit score, living honesty-credit score, social honesty-credit score. The proportion of the four parts is as the following formula:

$$RMS= (LHS+EHS+LIS+SHS)*25\% \quad (1)$$

RMS: Records Management Score
LHS: Learning Honesty-credit Score
EHS: Economic Honesty-credit Score
LIS: Living Honesty-credit Score
SHS: Social Honesty-credit Score

Records Management Score can be calculated according to the above formula.

The implement of democratic appraise is more complex during operation. Every class has a standing body and a provisional organ. The standing body named class honesty-credit evaluation group is responsible for records management mentioned above. The provisional organ named class democratic appraise group. The former works for an entire semester, while the latter is hurriedly dissolved after it has finished its task.

The class provisional organ of five people democratic appraise group is built before the start of each semester. The group firstly discusses the performance of every student of the class. In the end, each member of the group rates every student of the class, the average of which is the democratic appraise score. The final score of each

student of the class is obtained by summing the records management score and the democratic appraise score according to certain proportion as the following formula:

$$FS= RMS*80\%+LHS*20\% \quad (2)$$

FS: Final Score of Honesty-credit
RMS: Records Management Score
LHS: Democratic Appraise Score

The final score can be obtained according to the above formula. Under that system, student win points not just for democratic appraise score such as casting a vote, but also for including records management score.

B. Evaluation Content

Evaluation content is a set of student's behavioral expression concerning honesty-credit, which is the criterion of college student honesty-credit evaluation that has a complex architecture and many items. Through scientifically setting evaluation content, the items can help to guide and encourage student honesty-credit behavior. In our study, after years of practice, we have established a scientific and reasonable evaluation system.

TABLE I. TABLE OF FORBIDDEN ITEMS

Type	Item	Score
Learning Honesty-credit	Cheating on exams or illegally altering score	-40
	Arriving late or leaving early in class	-2
	Absence from class	-5
	Faking papers	-20
Economic Honesty-credit	Maliciously outstanding fees(tuition, course material fee, accommodation fee)	-20
	not return on time loan or interest	-10
	Defraud of financial aid	-30
	Extravagance and waste	-20
Living Honesty-credit	Concealing health condition	-20
	Breach of the public safety	-10
	Criticalnon-conformity of public health	-10
	Breach obligations	-10
	Violation of school rules: Criticized	-10
	Violation of school rules: Warning	-20
	Violation of school rules: Serious warning	-30
	Violation of school rules: Demerits	-40
	Violation of school rules: Probation	-60
	Disorder engraved graffito in the furniture and walls	-10
	Scuffles	-30
	Using of improper means in the democratic appraisal or election	-20
	Smoking in non-smoking area	-10
Social Honesty-credit	Intentionally sabotaging public property	-20
	Forgery of a certificate	-20
	Employment resume false	-20
	Non-performing employment contract	-20
	View, spread reactionary, pornographic books, videos, etc	-20

The raw score of every type credit is 80 points. Items listed in TABLE I are forbidden to do. The last column is the score that should be subtracted from the raw score of corresponding type.

TABLE II. TABLE OF ENCOURAGED ITEMS

Type	Item	Score
Learning Honesty-credit	Sharing their learning experience, enthusiasm to help students to learn	+5
	Taking the initiative to help teachers do a good job of teaching and research work	+5
	Won the school model of studying	+10
	Departmental awards won	+2
	Won the school award	+5
	Won the provincial award	+10
	Won the national award	+20
Economic Honesty-credit	High integrity, on time payments, but also interest, but also material	+10
	Simple life, not extravagance	+5
	Subsidized student with financial difficulties	+5
Living Honesty-credit	Score of dormitory cleanliness 95 points or more (Once a week)	+1
	School civilized dormitory	+5
	Returning lost money, recognition by the school or municipal	+10
	Returning lost money, recognition by the provincial	+15
	Returning lost money, national recognition	+20
	High morality, concerned about the collective, helping others, and outstanding deeds	+5
	Student leaders who have a good work attitude and dutifully complete their work	+5
Social Honesty-credit	Courageous and dare to fight bad guys, recognition by the school or municipal	+10
	Courageous and dare to fight bad guys, recognition by the provincial	+15
	Courageous and dare to fight bad guys, national recognition	+20
	Often participate in public good, Outstanding achievements, recognition by the school or municipal	+10
	Often participate in public good, Outstanding achievements, recognition by the provincial	+15
	Often participate in public good, Outstanding achievements, national recognition	+20
	Protection of state property, recognition by the school or municipal	+10
	Protection of state property, recognition by the provincial	+15
	Protection of state property, national recognition	+20

Items listed in TABLE II are encouraged to do. The last column is the score that should be added on the raw score of corresponding type.

III. FRAMEWORK OF EVALUATION SYSTEM

With the college expansion in successive years, the number of students in school has become increasingly. Student affairs management becomes more complicated. Information management system of student affairs is the assistant for the team of student affairs management, of which college students honesty-credit evaluation system is one part. Its main function modules of framework are listed in the following showed as Fig. 1:

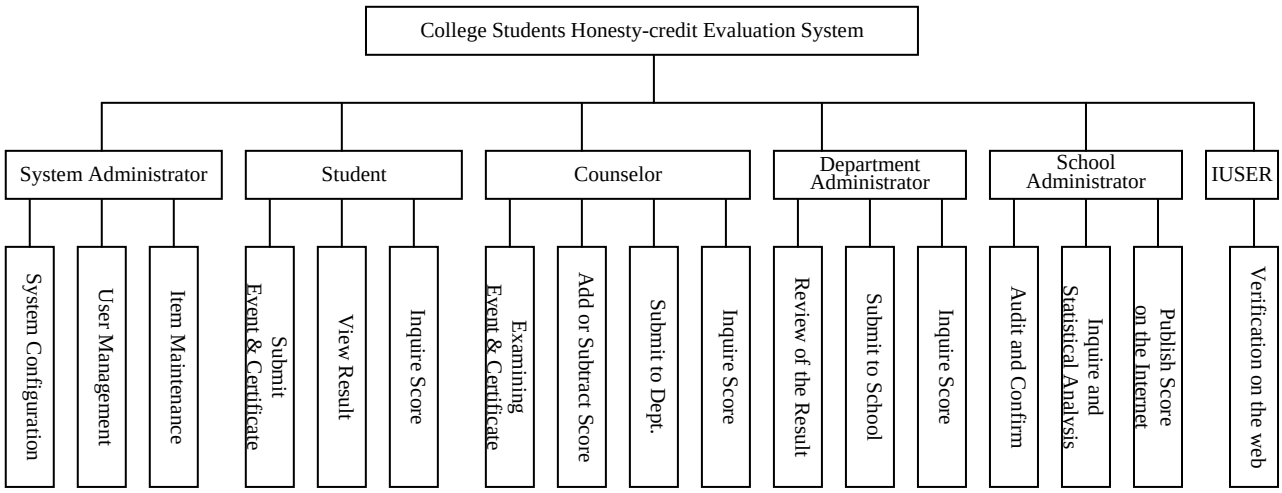


Figure 1. Framework of Evaluation System.

A. *Main Features of Framework*

The main features of college student honesty-credit evaluation system are listed in the following:

- User permissions can be set according to the actual staff level, such as student, counselor, department administrator, school administrator, etc. According to different permissions, the function menus of each user are also different.
- Scientific and efficient documentary function can help users to complete records management. Powerful data processing automatically calculate score.
- Workflow is a group of officers to complete a certain task with all the work carried out by the process of automatic transfer.
- Powerful and accurate search function is necessary. The results of honesty-credit evaluation play an important role in honor selection process. In the end, the results must be queried, sorted, and printing out.
- Score has been published on the Internet. Government, Enterprise, Financial institution and so on attaches great importance to student honesty-credit evaluation in recruit and business. Score can be verified on the webpage according to the personal information that student supplied.

B. *Main Function Modules of Framework*

In Fig. 1, there are six function modules, which are described one by one in the following.

- System Administrator. This function module has three menus. Above all, the menu of System Configuration is to configure parameters of system. For example, inputting every department and adding each class in corresponding department. Secondly, User Management menu is to add username, build password and set authorization. Thirdly, the menu named Item Maintenance is to set the forbidden items and the encouraged items of evaluation system mentioned in TABLE I and TABLE II. Fig. 2 is illustrated the function module of system administrator in details.

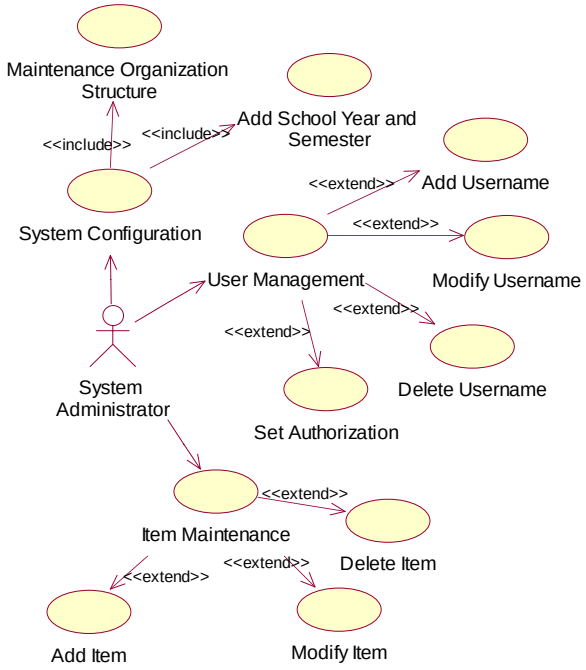


Figure 2. Use Case Diagram of System Administrator.

- Student. This module introduces the function of student interface. Student can submit events and certificates to score points. When counselor examined what student submitted is true, he/she will approve it. Otherwise, he/she will reject student's request. The result can be looked up by using the View Result menu. The Inquire Score menu can offer the total sore of this semester and all past semester. This function module is illustrated in Fig. 3.

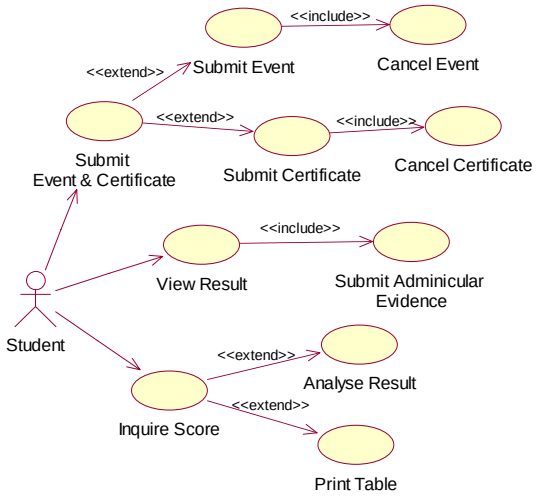


Figure 3. Use Case Diagram of Student.

- Counselor. In this framework of evaluation system, counselor is the key role. He/she is in charge of not only verifying the evidence that student submitted, but also submitting certificate and adding or subtracting corresponding value. At the beginning of semester, he/she need to arrange and supervise committee of class that he/she has jurisdiction over to democratically appraise every student honesty-credit in previous semester. He/she must provisionally set up appraising accounts of each member of democratic appraise group (AADA). After democratic appraise scores create, system will automatically generate the final scores of honesty-credit. Without objection by the public or non-objection to the final scores, counselor submits to department. The all tasks of counselor are shown in Fig. 4.

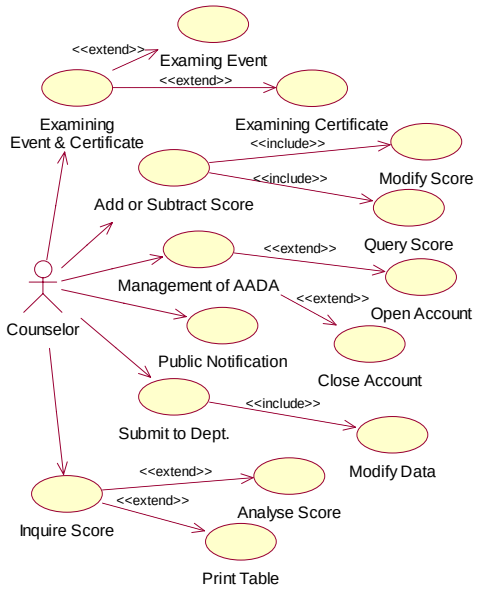


Figure 4. Use Case Diagram of Counselor.

- Department Administrator. The onerous task for department administrator is to review of the results submitted by counselors. The staff of this module play dual role. One is to appraise the counselors of the same department. Other is appraised by school. When he/she is convinced that there are no potential errors, the data of whole department are submitted to school. The function of Inquire Score menu is as same as module of counselor's except that the range of query is the whole department. The all functions of this module are shown in Fig. 5.

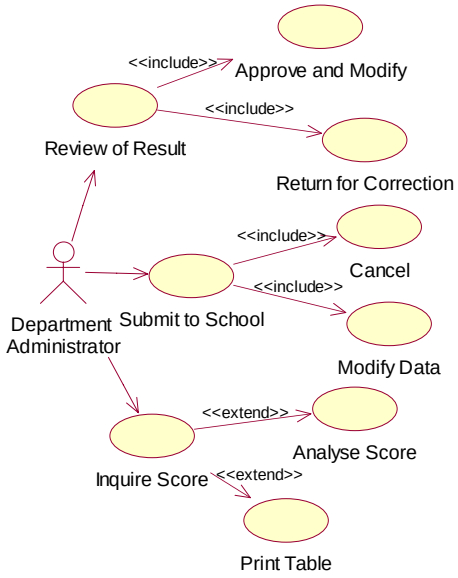


Figure 5. Use Case Diagram of Department Administrator.

- School Administrator. In this module, school administrator has three main functions. Firstly, the data submitted by all departments must be audited and confirmed. Secondly, the function of inquiring and statistical analysis is of the essence. Thirdly, publishing score on the internet is a useful function. The all functions of this module are shown in Fig. 6.

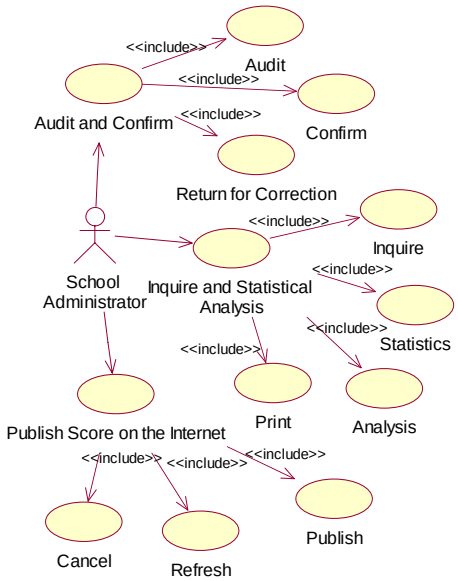


Figure 6. Use Case Diagram of School Administrator.

- IUSER. This is an open module that can allow anonymous access. Its function and features have been mentioned in advance.

IV. Conclusion and Forecast

It is very meaningful to set up college student honesty-credit evaluation system, which is conducive to enhance the pertinence and effectiveness of honesty-credit education. In terms of colleges and universities, honesty-credit education has become an important part of moral education. But the manner and content of honesty-credit education is relatively simple, the evaluation system is reference for honesty-credit education. As for college student, evaluation system is the behavioral constraints. Management information system of honesty-credit evaluation is an efficient and exact tool for students and educational institutions. In our study, the framework of the evaluation system can bring great changes in honesty-credit education and evaluation.

For simplifying application, the final score of honesty-credit is scaled into four grades which are described in TABLE III.

TABLE III. Grades of Honesty-credit

Grade	Score	Introduction
A	90—100	Honest and trustworthy, have good moral character, with models.
B	80—89	Basically honest and trustworthy, with good moral quality, timely correction of errors.
C	60—79	Often acts without integrity.
D	Below 59	Can not be honest and trustworthy, more serious.

The application of honesty-credit grade:

- College student whose honesty-credit rating scale is the grade A, or grade B, has the selection qualification of scholarship and advanced personal. Under the same conditions, the high grade has priority.
- The grade C and grade D can not participate in selection of scholarship and advanced personal, and the grade D are not entitled to various types of funding. The class, the number of grade D students more than 10%, will not participate in the selection of advanced collective in current year.
- College student honesty-credit evaluation system is opening to bank. When student apply for student loan, or graduate apply for loan, the likelihood of approval is greatly increased.
- In the long term, when the file of student honesty-credit is provided for employers, employers have a better understanding of graduates, and the employment rate of graduates can be greatly improved.

Morality is a person's stable, long-lasting, the overall state of mind that is shown by a long series of behaviors. The situation of college student honesty-credit can be more reasonable and quantitatively evaluated.

College student credit file that is a part of the honesty-credit education system must be established. Through the establishment of college student credit file, ethical values and integrity of the college students consciousness, behavioral norms of their faith are educated and guided, and the purpose of system control is achieved.

Acknowledgment

This work was supported in part by the Research Project of Shandong Economic University and the Student Affairs Office of Shandong Economic University.

References

[1] Tiqin Zhang, Silun Mu, *Honesty-credit Cultivation Introduction*, 2nd ed., Jinan: Shandong People's Publishing House, 2009, pp.225-235.

[2] Student Affairs Office of Shandong Economic University, "Student Handbook," unpublished, 2007, pp.204-205.

[3] Xingxiang Qi, "Design and Development of Student Affairs Management System," *Science & Technology Information*, vol.36, pp. 170, December 2009.

[4] Xingxiang Qi, "Application of Bar-Code in Career Management of University Graduate," *Modern Enterprise Education*, vol.2, pp.103-104, January 2010.

[5] Xiangxin Liu, Silun Mu, Shuchen Hao, et al. *Seek New Mechanisms of the Educational Work in Colleges and Universities*, Beijing: People Press, 2008.

[6] Min Cai, Huihui Xu, Bingqiang Huang. *UML basic and Modeling with the Rose*, Beijing: People's Posts and Telecommunications Press, 2006.

[7] Peihuan Gen, "The lack of Honesty and Credit of College Students and the Correction," *Journal of Pingyuan University*, vol.22, pp.115-116, December 2005.

[8] Hao Ren, Xiangming Fang, "Commenton Structuring the evaluation system of college students' honesty and credit," *Zhejiang Gongshang University of Commerce*, vol.70, pp. 88-91, January 2005.

[9] Chunming Wang, "On Instructing Estimate Mechanism of College Students' Honesty and Faith Education," *Journal of Xinyang Normal University (Philos. &Soc. Sci. Edit.)*, vol.27, pp. 76-77, October 2007.

[10] Website[online]: http://www.online.sdie.edu.cn.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15, August 2010, pp. 463-465

Study on Urban Spatial Structure Changes of Jiaozuo City Based on SLEUTH Model

Guan Zhongmei[1], Wang Yucun[2]
[1] College of Surveying and Land Information Engineering of Henan Polytechnic University, Jiaozuo, China
Email: gzm@hpu.edu.cn
[2] College of Architecture and Urban Planning of Suzhou University of Science and Technology, Suzhou, China
Email: suzhoujianzhuxi@163.com

***Abstract*—Taking the metropolitan area of Jiaozuo as a case study, this paper utilized the urban model of SLEUTH to forecast the spatial structure change of Jiaozuo in four alternative scenarios based on four Landsat TM/ETM+ images(eg.1988, 1992, 2001 and 2008). In the research, RS and GIS were adopted. The result indicated that SLEUTH model can effectively simulate the urban growth and sprawl; both back-past (1988-2008 year) map and forecasting simulation of Urban Geospatial growth and sprawl (2009-2020 year) were mapping. Thus, the SLEUTH model simulation on city dynamics geospatial change can provide useful information for the urban future planning and development.**

***Index Terms*—SLEUTH model; spatial structure; Jiaozuo city**

I. INTRODUCTION

As a carrier for human production and life, city has its characteristics like openness, dynamic and self-organization etc. The evolution of urban spatial structure is a very complex phenomenon. As a complicated phenomenon, the changing of urban spatial structure has been the focus of all aspects of academic research. Because of the high complexity of the process, the traditional method can no longer truly and exactly simulate of the evolution of urban spatial structure [1-3]. In this case urban model is an effective tool on studying the evolution of urban spatial structure [4]. In the existing urban model, cellular automata (CA) model as a strong spatial dynamic simulation capability, are widely used in simulating the evolution of urban spatial structure and it had achieved many significant research results [5-7].

Jiaozuo City is a typical resource-based city. Its spatial structure changes are typical in our country. Some scholars have analysis the reasons of spatial structure of Jiaozuo City, but comprehensive analysis of the urban spatial structure by using the remote sensing data and the CA model in Jiaozuo City has not been published. This paper discussed the law of evolution of spatial structure of Jiaozuo City, using SLEUTH model to simulate possible future scenarios, on the purpose of providing reference to the urban planning.

II. STUDY AREA AND DATA

A. Study Area

This study area includes Jiefang district, Shanyang district, Macun district and High-tech Industrial Development zone, E113° 06′ 31″ —E113° 26′ 14″ , N35° 20′ 51″ —N35° 09′ 14″ . This area lie to the east of Boai County, the west of Xiuwu County, the north of Wuzhi County, the south of Taihang moutain, topography from north to south, and the terrain changed greatly.

B. Data and Method

Remote sensing data is SLEUTH model's basic data, this research include four Landsat TM/ETM+ images of 1988, 1992, 2001 and 2008 year. There are two considerations to use remote sensing images of this period: firstly, this period from 1988 year to 2008 year is its rapid urbanization period; secondly, this period is its transition period from resource-based cities to the tourist city.

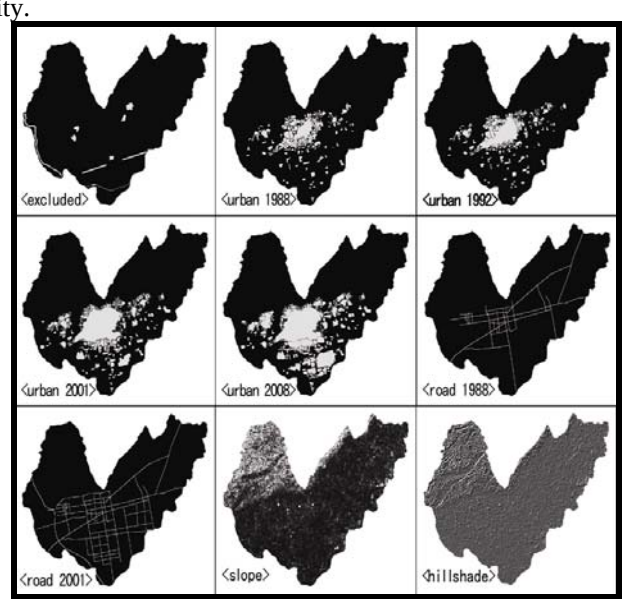


Figure.1 Input images of SLEUTH model in Jiaozuo City

According to the urban transportation, terrain conditions, the SLEUTH model set the appropriate coefficients. As the improvement product based on cellular automaton, it is loosely coupled with GIS. The name SLEUTH was derived from the simple image input requirements of the models: Slope, Land cover,

Excluded, Urbanization, Transportation, and Hillshade[8-10].

SLEUTH model includes three modules: test, calibration, prediction; five parameters: dispersion—coefficient、breed — coefficient 、 spread — coefficient 、 road — coefficient and slope—coefficient; four kinds of growth: Spontaneous Growth，New Spreading Center Growth，Edge Growth，Road Influenced Growth[11-13].

The study took the method of grading coefficient to reduce the range of coefficient on the purpose of filtering unnecessary coefficient. We began our study with setting the initial value of breed coefficient 、 spread coefficient、slope—coefficient、dispersion coefficient and road coefficient in order to find out Coarse of coefficient, then fine and final based on the Coarse through the way of narrowing the value range of coefficient, finally, we got the suitable coefficient combination[14-15].

Input the parameters and data to the model. First, input the data layer of the Urban of 1988 as a seed point to the model. After running it we can get the simulated result of the phase 1992 as a set of parameters; then, use the actual data layer of 1992 as the seed point to run the model. After that we can get the simulated result of the phase 2001 as another set of parameters; At last, use the data layer of 2001 to run it, we can get the result of the phase 2008 as the third set. From amending and checking each set of parameter at three levels, we are able to acquire three optimal sets of parameters. Use these three optimal sets as the simulated parameters of the three phases. Pick the best set as the seed point of 2008 to input and we can obtain the optimal parameters for subsequent phases. Finally, it can be possible to make a forecast on urban spatial structure evolution with the optimal parameters.

Table.1 Statistical results of model calibration

coefficient	Coarse		Fine		Final		final
	Lee-salee=0.56648		Lee-salee=0.58374		Lee-salee=0.59798		Parameter
	Scope	step	Scope	step	Scope	step	value
Dispersion coefficient	1-100	25	1-25	5	5-10	1	9
Breed coefficient	1-100	25	25-50	5	35-45	1	40
Spread coefficient	1-100	25	25-75	5	60-70	1	63
Slope coefficient	1-100	25	75-100	5	80-90	1	86
Road gravity coefficient	1-100	25	1-100	25	40-60	5	45

III. SIMULATION RESULT AND VALUATION

By calibrating different combinations of the coefficients through three stages, chose the most appropriate group, then input this group of the coefficient to SLEUTH model for simulating, obtain simulation map and prediction map of the urban spatial structure for Jiaozuo city. Analyzing based on actual data and prediction map; we could receive the following results:

(1) Model performance in study area was improved with increased spatial and parameter resolution. As we can see from initial values of 100 or one(in the case of diffusion) the five coefficients (diffusion, breed, spread, slope resistance, and road gravity) were narrowed down to more accurately reflect study area. From an initial breed coefficient of 100 in the study area it was possible to narrow down to a breed coefficient of 50 in the case of study area.

(2) A first improvement in model performance took place initially in the coarse calibration phase. Before coarse calibration, the maximum extreme values were given, from that maximum of 100, in all the coefficient values except diffusion that was given one. From that initial value, the resulting set of values output from coarse calibration was 25, 50, 75, 100, 100 in the case of the study area. As already explained before, these values fed the next calibration phase (fine calibration).

(3) The most substantial improvement in model performance was reached between the coarse and the fine calibration phases. For instance, during coarse calibration, and for study area, the maximum value of breed coefficient was 50, and in the study area it assumed the opposite extreme value of one (reflecting the previously mentioned erratic behavior of the model trying to adjust itself to an “unknown reality”). In the case of diffusion, because this is a coefficient that measures organic growth, we wanted to see how far it could increase, so we began assuming that it spread outward one cell per year in the coarse calibration in the study area.

(4) An adjustment of the values, less intense than in the previous calibration phases, happened between the fine and final calibration phases. In the study area, it was possible to see that from fine to final calibration a slight adjustment was made to the values of the study area. The value of Road gravity coefficient adjustment was higher.

(5) Initial values for the coefficients of diffusion, breed, spread, slope, and road gravity improved from coarse to fine and then to final calibration in the study area.

Consequently:

- The SLEUTH model reproduces the development of urban space during the 1988-2008 years. The simulation results match with the actual data (Figure 2). It proved that the SLEUTH model has the capability of spatial modeling and operating.

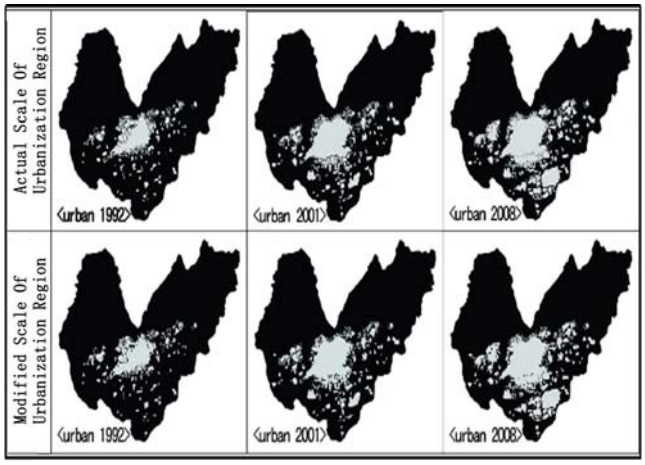


Figure.2 Comparative map about actual scale and modified scale

- Through the SLEUTH model, we can get the forecast map for the growth of the city from 2008 year to 2020 year. In the prediction map, gray and light gray represent new urbanization area, light gray represent for the high-growth areas, gray represent for the low-growth areas (Figure 3). With time, there is a growing trend to outer space, and urbanization area is expanding.

- The development of the spatial structure of Jiaozuo City is subject to geographical conditions. The northern part of the city is Taihang Mountains, and the eastern and western sides is a coal mining subsidence areas, urban space can develop to the east an west by skipping the east, west side of the coal mining subsidence area. Because the south is the alluvial plain, the city space can develop to the south as its main direction with great development conditions. Through the above analysis, we find the SLEUTH model is useful when we do urban planning.

Figure.3 Forecasting map by SLEUTH module

IV. CONCLUSION

As the dynamic simulation of a city, the SLEUTH model can be used both in simulating the generation of virtual cities, revealing the structural features and development laws of urban development, and stimulating and predicting city's development. Simulate the future urban patterns based on a cellular automata model and GIS technology, which integrate mechanism of urban spatial growth into the simulation model. Based on the characteristics of resource-based cities，using SLEUTH model，validating 5 coefficient of transformation rules repeatedly, this paper simulate and analysis the dynamic behavior in the development process of the main city. Of course, any model is limited, including SLEUTH model, such as it is not very well in the simulation results of newly built-up area, this model needs to be improved in the future research.

REFERENCES

[1] Lambin E F, Turner B L, "Geist H J, etal. The causes of land-use and land-cover change: Moving beyond the myths", Global Environmental Change, 2001, vol.11, pp. 261-269.

[2] Batty M, "Urban evolution on the desktop: Simulation with the use of extended cellular automata s", Environment and Planning A, 1998, vol.30, pp. 1943-1967.

[3] United Nations, "World Urbanization Prospects: The 2003 Revision", New York: United Nations Press, 2004.

[4] Wu F, "alibration of stochastic cellular automata: The application to rural-urban land conversions", International Journal of Geographical Information Science, 2002, vol.16, pp.795-818.

[5] Feng Jian, "patial-temporal evolution of urban morphology and land use structure in Hangzhou", Acta Geographica Sinica, 2003, vol.58, pp. 343-353..

[6] Wu Xiao-bo, Zhao Jian, Wei Chen-jie,etal, "The urban expanding simulation with the cellular automata model in Haikou", Urban Planning, 2002, vol.26 , pp.69-73.

[7] Dietzel C,Clarke K C, "The effect of disaggregating land use categories in cellular automata during model calibration and forecasting", Computers, Environment and Urban Systems, 2006, vol.30 , pp. 78~101.

[8] Ding Han, "LandUse/Land CoverChanges, Driving ForceAnalysis and Prediction in the Developed CoastalRegion in China: A Case Study in Zhejiang Province", Hangzhou: ZhejiangUniversity, 2006.

[9] Robert M, "Golledge:Understanding special concepts at geographic scale without the use of vision ", Progress in Human Geography, 1997, vol. 2 , pp.225-242.

[10] Wang Anyou, Yu Jicong , "Study of Fuxin economic transformation mode", China: USA Business Review .2005, vol. 4 , pp.29～32.

[11] Krugman, Paul, "Increasing Returns and Economic Geography", Journal of Political Economy, 1991, vol.99, pp.483-99.

[12] Chen X G, "Study on simulating urban growth based CA", Wulumuqi:College of Resources and Environmental Science, Xinjiang University,2005.

[13] Silva E A, ClarkeK C, "Calibration of the SLEUTH urban growthmodel forLisbon and Porto, Portugal", Computers, Environmentand Urban Systems, 2002, vol.26, pp.525-552.

[14] Zhang Yan, LI Jing, Chen Yun-hao, " Simulation of Beijing urbanization using SLEUTH ", Remote Sensing Information, 2007, vol. 2 , pp.50-54.

[15] Ye J A,Song X D, Niu X Y, etal, " Geographic information system and planning support system", Beijing: Science Press, 2006.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15, August 2010, pp. 466-470

Non-Malleable Non-Interactive Zero Knowledge Proof Using InstD-VRF

Guifang Huang[1], Lei Hu[1], and Dongdai Lin[2]
[1] State Key Laboratory of Information Security, the Graduate School of Chinese Academy of Sciences
Email: {gfhuang, hulei}gucas.ac.cn
[2] Institute of Software, Chinese Academy of Sciences
Email: ddlin@is.iscas.ac.cn

***Abstract*—In asynchronous network communication, non-malleability is a necessary security requirement to resist against man-in-the-middle attack. In [6], two non-malleable non-interactive zero knowledge proofs were presented, in which the first scheme was obtained by using a specific form of InstD-VRF. In this paper, we present how to construct non-malleable non-interactive zero knowledge proof by using the general InstD-VRF. Our construction is a framework and contains many non-malleable non-interactive zero knowledge proofs. With this framework, the security analysis of some complicated non-malleable non-interactive zero knowledge proofs can be simplified, as long as they are consistent with the framework.**

***Index Terms*—non-interactive zero knowledge (NIZK) proof, simulation-soundness, non-malleability, InstD-VFRF**

I. INTRODUCTION

Non-interactive zero knowledge (short for NIZK) proof was introduced by Blum, Feldman and Micali [2]. In an NIZK proof, prover and verifier share some randomness called the common reference string (short for CRS). Using the CRS, prover can send only one message to convince verifier the validity of a statement without revealing anything else. It was shown that any language in NP has a NIZK proof [11]. Since its introduction, many useful results and applications of NIZK have been worked out [1,3,8,10].

Instance-dependent verifiable random function (short for InstD-VRF), a variant of verifiable random function [15], is a new cryptographic primitive introduced in [7] to resolve the simultaneous resettability conjecture. The public key of InstD-VRF is of the form $(y,\cdot)$, where y is called a key-instance. The property of InstD-VRF is dependent on whether the public key-instance y is in some language L_1. For InstD-VRF, different definitions of language L_1 give rise to different constructions of InstD-VRF. For example, in [7], an InstD-VRF was constructed from zap which is a two-round public-coin witness indistinguishable proof [9,13]. In addition, InstD-VRF can be constructed from simulatable verifiable random function (short for S-VRF) [4]. Observed that for the construction of S-VRF in [4], take $y = n$ as the key-instance, we can get a specific InstD-VRF based on number theoretical assumption.

The notion of non-malleability was firstly brought out in [5]. After that, Sahai defined non-malleability for NIZK [17]. In addition, a weaker notion--simulation-soundness was defined. Intuitively, non-malleability means that even if having seen polynomial many proofs adaptively, the probabilistic polynomial-time adversary can only do what he could have done before. The non-malleable NIZK scheme in [17] was a bounded one. That is, the number of left proofs is bounded by a fixed polynomial in advance. Subsequently, De Santis et al strengthened the definition of non-malleability of NIZK and presented two unbounded non-malleable NIZK schemes [6]. Unbounded non-malleability means that non-malleability holds even if the number of left proofs is an arbitrary polynomial, instead of a fixed one. After the introduction of InstD-VRF, it is observed that the first scheme in [6] was obtained by using a specific InstD-VRF constructed from zap [7].

In this paper, we propose a generalized method of constructing non-malleable NIZK by using InstD-VRF. This method provides a framework of using different forms of InstD-VRF to construct non-malleable NIZK. With this framework, no matter how complicated the non-malleable NIZK proof is, its security analysis can be simplified, as long as it is consistent with the framework.

It is organized as follows. In section 2, some related notions and definitions are given. In section 3, we present the construction of non-malleable NIZK scheme.

II. PRELIMINARIES

We use the following standard notations and conventions for probabilistic algorithms and experiments.

If A is a probabilistic algorithm, then $A(x,r)$ denotes the output of running algorithm A on input x and coin r. Let $y \leftarrow A(x)$ denote the experiment of selecting r randomly and y is the output of $A(x,r)$. If S is a finite set, let $\alpha \leftarrow S$ denote the process of selecting an element α uniformly from S.

A function $f(n)$ is said to be negligible if for every polynomial $q(n)$ there exists a positive integer N such that for all $n \geq N$, we have $f(n) \leq 1/q(n)$.

Supported by NSFC No.60773134, the national 863 Program No.2006AA01Z416, the national 973 Program No.2007CB311201 and the 47th postdoctoral Fund of China No.20100470598
Corresponding Email: gfhuang@gucas.ac.cn

Definition 1 (NIZK Proof) $\prod = (\ell, P, V, S = (S_1, S_2))$ is called a non-interactive zero knowledge proof for the language $L \in NP$ with the relation R if ℓ is a polynomial, V and $S = (S_1, S_2)$ are probabilistic polynomial-time machines such that the following conditions hold:

- Completeness: For every $x \in L$ of length k, all w such that $R(x, w) = true$ and all strings σ of length $\ell(k)$, we have $V(x, P(x, w, \sigma), \sigma) = true$.
- Soundness: For any adversary A, if $\sigma \in \{0,1\}^{\ell(k)}$ is chosen uniformly, then the probability that $A(\sigma)$ will output (x, p) such that $x \notin L$ and $V(x, p, \sigma) = true$ is a negligible function in k.
- Zero Knowledge Property: For any non-uniform probabilistic polynomial-time adversary $A = (A_1, A_2)$, we have that

$$|\Pr[Expt_A(k) = 1] - \Pr[Expt_A^S(k) = 1]| \le \alpha(k)$$

where $\alpha(k)$ is a negligible function in k and two experiments $Expt_A(k)$ and $Expt_A^S(k)$ are defined as follows:

$Expt_A(k)$:	$Expt_A^S(k)$:
$\sigma \leftarrow \{0,1\}^{\ell(k)}$	$(\sigma, \tau) \leftarrow S_1(1^k)$
Return $A^{P(\cdot,\cdot,\sigma)}(\sigma)$	Return $A^{S_2(\cdot,\sigma,r)}(\sigma)$

Definition 2 (InstD-VRF) An InstD-VRF with respect to language $L_1 \in NP$ associates with the following algorithms:

- $KG\,\Pr ot$: the key generation protocol between a querier and the owner of the function. It takes security parameter k as input and outputs a pair of public/secret keys (pk, sk) where pk is of the form $(y, \cdot)$ and is public, $y \in L_1 \cap \{0,1\}^n$ is called a key-instance.
- $F = (f, prov)$: is the evaluator of the function. f is a deterministic algorithm and $prov$ is a probabilistic algorithm. Given (pk, sk) and $a \in \{0,1\}^{d(n)}$, F outputs a function value and a proof for the correctness of this function value. That is,

$$F_{(pk,sk)}(a) = (f_{sk}(a), prov(a, f_{sk}(a), pk, sk)) = (b, \pi)$$

- Ver: On input (a, pk, b, π), the algorithm verifies whether (b, π) is is the correct value of the InstD-VRF under pk on input a. If so, it outputs 1; otherwise, outputs 0.

$Fake$: Suppose $y \in L_1$ and w_y is its witness. For any $a \in \{0,1\}^{d(n)}$ and randomly chosen b of length $\ell(n)$ in the range, $Fake_{(pk,w_y)}$ can output $(b, prov(a, b, pk, w_y)) = (b, \pi)$ such that

$$Ver(a, b, pk, \pi) = 1.$$

$F_{(pk,sk)}$ is called an InstD-VRF if the following properties hold:

Provability: If $(b, \pi) = F_{(pk,sk)}(a)$, then $Ver(pk, a, b, \pi) = 1$.

2. Pseudo-randomness on yes key-instance: If $y \in L_1$ and w_y is its witness, then for any probabilistic polynomial time oracle machine M and sufficiently large n, for every polynomial $p(\cdot)$, we have

$$|\Pr[M^{Fake_{(pk,w_y,h)}}(1^n) = 1 : h \leftarrow H_n] - \Pr[M^{F_{(sk,pk)}}(1^n) = 1]| < 1/p(n)$$

Where H_n is the set of all the functions from length $d(n)$ to length $\ell(n)$, $d(n)$ and $\ell(n)$ are polynomials in n;

3. Uniqueness on no key-instance: If $y \notin L_1$, the probability that there exists $(a, b_1, b_2, pk, \pi_1, \pi_2)$ such that $b_1 \neq b_2$ and $Ver(a, b_i, pk, \pi_i) = 1$ for $i = 1, 2$ is negligible.

Definition 3 (Simulation-soundness) Suppose $\prod = (\ell, P, V, S = (S_1, S_2))$ is a NIZK for language $L \in NP$. It is simulation-sound if for any non-uniform probabilistic polynomial-time adversary A, it holds that

$$\Pr[Expt_{A,\prod}(k) = true] \le \alpha(k)$$

where $\alpha(k)$ is a negligible function and experiment $Expt_{A,\prod}(k)$ is defined as follows:

$Expt_{A,\prod}(k)$:
$(\sigma, \tau) \leftarrow S_1(1^k)$;
$(x, p) \leftarrow A^{S_2(\cdot,\sigma,\tau)}(\sigma)$;
Let Q be the set of proofs given by S_2 above
Return true iff $p \notin Q \wedge x \notin L \wedge V(x, p, \sigma) = 1$

Definition 4 (Non-Malleable NIZK) Suppose $\prod = (\ell, P, V, S = (S_1, S_2))$ is a NIZK for language $L \in NP$ with witness relation W. It is said to be a non-malleable NIZK proof for L if there exists a probabilistic polynomial-time oracle machine M such that for all non-uniform probabilistic

polynomial time adversary A and all non-uniform polynomial-time relation R, there exists a negligible function $\alpha(k)$ such that

$$|\Pr[Expt^{S}_{A,R}(k)=1]-\Pr[Expt'_{A,R}(k)=1]|\le \alpha(k),$$

where experiments $Expt^{S}_{A,R}(k)$ and $Expt'_{A,R}(k)$ are respectively defined as follows:

$Expt^{S}_{A,R}(k)$:
$(\sigma,\tau)\leftarrow S_1(1^k)$;
$(x,p,aux)\leftarrow A^{S_2(\cdot,\sigma,\tau)}(\sigma)$;
Let Q be the set of proofs produced by S_2
Return 1 iff
$p\notin Q\wedge V(x,p,\sigma)=1\wedge R(x,aux)=1$
$Expt'_{A,R}(k)$:
$(x,w,aux)\leftarrow M^{A}(1^k)$
Return 1 iff $(x,w)\in W\wedge R(x,aux)=1$

Security for Signature Scheme: A signature scheme is a tuple $(Gen,Sign,Ver)$ where Gen is a probabilistic key generator that outputs a pair of keys (sk,vk). $Sign$ is a randomized signature algorithm. Given the message m and the signature-key sk, $Sign$ outputs a signature s for m. Ver is a deterministic verification algorithm. Given (pk,m,s), Ver outputs 1 if s is a valid signature for m and otherwise 0.

A signature scheme is existentially unforgeable against adaptive chosen message attack, if for any polynomial time adversary, he can not produce a valid signature for a new message, even if he has seen many signatures for the messages adaptively chosen by himself.

A strong one-time signature scheme is required that even if a polynomial time adversary has seen a signature for a message, he can not generate a different valid signature-message pair. Strong one-time signature schemes can be constructed based on the existence of universal hash functions and one-way permutations [16].

III. Non-Malleable NIZK Construction

In this section, Using InstD-VRF, a simulation-sound NIZK proof $\prod$ is first presented. Then, as in [17,6,14], in protocol $\prod$, replacing the NIZK proof by the corresponding NIZK proof of knowledge for the same statement as the sub-protocol, a non-malleable NIZK proof is obtained.

In the simulation-sound scheme given below, the value of InstD-VRF taken at some point is binded with a pair of keys of strong one-time signature scheme. Simulation-soundness property comes from the following fact: on one hand, from security of the signature scheme, the public keys in accepting proofs produced by the adversary will not appear in the left simulated proofs; on the other hand, in the simulated proofs, the public key-instance of InstD-VRF is a no instance. For any polynomial time man-in-the-middle adversary, if he can prove a false statement successfully, he must generate two different function values and their correctness proofs at the same point, which contradicts with the uniqueness of InstD-VRF on no key-instance.

Suppose h is a one-way permutation, $Sig=(gen,sig,ver)$ is a strong one-time signature scheme, F is an InstD-VRF with respect to language L_1.

Here it is required that the key-instance contained in a randomly generated public key of InstD-VRF is a no instance. This is a natural assumption and easy to implement when language L_1 is a difficult language.

Construction of Protocol $\prod$:

- **Common Input:** $x\in\{0,1\}^n$
- **Common Reference String:** (σ,pk,a),where σ and a are random strings, (pk,sk) is a pair of randomly generated public/secret keys of InstD-VRF.
- **Prover Algorithm:** On input $x\in L$ and a witness w for $x\in L$, do:
  1. Randomly select $c\in\{0,1\}^{\ell(n)}$;
  2. randomly Generate a pair of keys (VK,SK) of Sig;
  3. Use σ as the CRS to generate a NIZK proof π to prove that $x\in L\vee(a,c,VK)\in L_2$, where L_2 is defined as follows:

$$L_2=\{(a,c,VK):\exists b,\pi_1,s.t.\ c=h(b)\wedge (VK,SK)=gen(1^k,b)\wedge(b,\pi_1)=F_{(sk,pk)}(a)\}$$

  4. Let $trans=(c,\pi)$;
  5. Generate a signature $s=sig(SK,trans)$;
  6. Output $(x,VK,s,trans)$.
- **Verifier Algorithm:**
  1. Verify that s is a valid signature for $trans$;
  2. Verify that π is a valid proof for $x\in L\vee(a,c,VK)\in L_2$;
  3. Output 1 if the above two checks are correct; otherwise output 0.

Theorem 1 If there exists an efficient InstD-VRF and a one-way permutation, then any language $L\in NP$ has an efficient simulation-sound NIZK proof.

Proof: For $x\in L$, prover can use a witness w of the Statement x to prove that $x\in L\vee(a,c,VK)\in L_2$. Therefore, the completeness property of protocol $\prod$ holds. For $x\notin L$, from the soundness property of the sub-protocol in step 3, protocol $\prod$ is sound.

The simulator of $\prod$ works as follows: First, S_1 generates randomly a pair of public/secret keys (sk, pk) of InstD-VRF and two strings a, σ. Then, S_2 computes $(b, \pi) = F_{(sk,pk)}(a)$, $(SK, VK) = gen(1^k, b)$ and $c = h(b)$. Using the witness (b, π), S_2 can simulate step 3 of $\prod$. At last, S_2 generates a signature for the message $trans$ under SK

From the simulator algorithm, by the standard hybrid argument technique, we conclude that the proofs produced by the simulator are indistinguishable from the proofs produced in the real conversation.

Next, we prove that protocol $\prod$ is simulation-sound. Define the following two indistinguishable experiments:

1. $Expt_0(1^n)$

$Expt_0(1^n)$:
Phase 1: Preprocessing
The algorithm works the same as S_1 and outputs (a, σ, sk, pk), where the public key-instance y contained in pk is a no instance.

Phase 2: When the adversary asks for proofs for x_i
Run the algorithm $S_2(x_i, a, \sigma, sk, pk)$.

Phase 3: When outputs $(x^*, VK^*, s^*, c^*, \pi^*)$, the adversary does:
Let Q be the set of proofs generated by the simulator.
Return 1 iff
$(VK^*, s^*, c^*, \pi^*) \notin Q \wedge x \notin L \wedge V(x^*, VK^*, s^*, c^*, \pi^*) = 1$

From the soundness property of statement $x \in L \vee (a, c, VK) \in L_2$, if the adversary can output an accepting proof $(x^*, VK^*, s^*, c^*, \pi^*)$ with a non-negligible probability, where $x^* \notin L$, then except for a negligible probability we have $(a, c, VK) \in L_2$. Therefore, $Expt_0(1^n)$ is computational indistinguishable from the following experiment.

2. $Expt_1(1^n)$

On one hand, from the security of the signature scheme, it is concluded that for all j, the public keys VK^* that the adversary generates in the right proof are different

$Expt_1(1^n)$:
Phase 1: Preprocessing
Same as Phase 1 of $Expt_0(1^n)$.

Phase 2: When the adversary asks for proofs for x_i
Same as.Phase 2 of $Expt_0(1^n)$.

Phase 3: When outputs $(x^*, VK^*, s^*, c^*, \pi^*)$, the the adversary does:
Let Q be the set of proofs generated by the simulator.
Return 1 iff $(VK^*, s^*, c^*, \pi^*) \notin Q \wedge x \notin L$ and
$c = h(b) \wedge (VK, SK) = Gen(1^k, b) \wedge (b, \pi_1) = F_{(sk,pk)}(a)$

from any of VK^j, where VK^j is the public keys of the signature scheme in the j-th simulated proof. On the other hand, $VK^* \neq VK^j$ means that $b^* \neq b^j$. Since the public key-instance of InstD-VRF is a no instance, by the uniqueness property of InstD-VRF on no key-instance, this case happens only with a negligible probability. Therefore, the probability that $Expt_1(1^n)$ outputs 1 is negligible. From the indistinguishability, we get the probability that $Expt_0(1^n)$ outputs 1 is negligible. That is, protocol $\prod$ is simulation-sound.

Corollary 1 In the above construction of protocol $\prod$, if using a NIZK proof of knowledge instead of a NIZK proof as the sub-protocol, then the resulting protocol is a non-malleable NIZK proof.

Remarks: Because there are many different ways of generating the key-instance y, protocol $\prod$ in fact is a framework to construct non-malleable NIZK scheme. We point out that in this scheme, if an InstD-VRF constructed from zap is used, the resulting non-malleable NIZK scheme is exactly the scheme given in [6]. Therefore, the new construction is a more general way of constructing non-malleable NIZK using InstD-VRF.

Ⅳ. CONCLUSION

InstD-VRF is a new cryptology primitive. In this paper, InstD-VRF is used to achieve non-malleability for NIZK. A generalized way of constructing non-malleable NIZK using InstD-VRF is presented. In the new construction, when using InstD-VRF constructed from zaps, the resulting non-malleable NIZK is the one proposed in [6]. Therefore, the scheme given in [6] is a special case of our construction.

REFERENCES

[1] M. Blum,., DE Santis, S. Micali. and G. Persiano: Non-interactive Zero Knowledge Proofs. SIAM Journal on Computing, 20(6): 1084-1118, 1991.

[2] M. Blum, P. Feldman and S. Micali: Non-Interactive Zero-Knowledge and Its Applications. Proc. of the 20th Annual Symposium on Theory of Computing-STOC'88, pp. 103-112, 1988.

[3] M. Bellare and S. Goldwasser: New Paradigms for Digital Signatures and Message Authentication based on Non-Interactive Zero Knowledge Proofs. In advances in Cryptology-Crypto'89, LNCS 435, pp. 194-21, 1989.

[4] M. Chase and A. Lysyanskaya: Simulatable VRFS with Applications to Multi-Theorem NIZK. In advances in Cryptology-Crypto'07, LNCS 4622, pp. 303-322, 2007.

[5] D. Dolev, C. Dwork and M. Naor: Non-Malleable Cryptography. SIAM Journal on Computing, 30(2): 391-437, 2000.

[6] A. De Santis, G. Di Crescenzo, R. Ostrovsky, G. Persiano and A. Sahai: Robust Non-Interactive Zero Knowledge. In advances in Cryptology-Crypto'01, LNCS 2139, pp. 566-598, 2001.

[7] Yi Deng and Dongdai Lin: Instance Dependent Verifiable Random Function and Their applications to Simulataneous Resettability. In advances in Cryptology-EUROCRYPT'07, LNCS 4515, pp. 148-168, 2007.

[8] A. De Santis, S. Micali and G. Persiano: Non-Interactive Zero Knowledge Proof Systems. In advances in Cryptology-Crypto'87, pp. 52-72, 1987.

[9] C. Dwork and M. Naor: Zaps and Their Applications. Proc. Of the 41st Symposium on Foundations of Computer Science-FOCS'00, pp. 283-293, 2000.

[10] A. De Santis and G. Persiano: Zero Knowledge Proof of Knowledge without Interaction. Proc. of the 33rd Symposium on Foundations of Computer Science-FOCS'92, pp. 427-436, 1992.

[11] U. Feige, D. Lapidot and A. Shamir: Multiple Non-Interactive Zero Knowledge Based on A Single Random String. Proc. of the 31st Annual Symposium on Foundations of Computer Science-FOCS'90, pp. 308-317, 1990.

[12] O. Goldreich: Secure Multi-Party Computation. http://www.wisdom.weizmann.ac.il.

[13] J. Groth, R. Ostrovsky and A. Sahai: Non-Interactive Zaps and New Techniques for NIZK. In Advances in Cryptology-Crypt'06, LNCS4117, pp. 97-111, 2006.

[14] Hongda Li and Bao Li: An Unbounded Simulation Sound Non-Interactive Zero-Knowledge Proof System for NP. Proc. Of INSCRYPT'05, LNCS 3822, pp. 210-220, Springer, 2005.

[15] S. Micali, M. Robin and S. Vadhan: Verifiable Random Function. Proc. of the 40th Annual Symposium on Foundations on Computer Sciences-FOCS'99, pp. 120-130,1999.

[16] M. Naor and M. Yung: Public-Key Cryptosystems Provably Secure Against Chosen Ciphertext Attacks. Proc. of the 22nd Annual Symposium on Theory of Computing-STOC'90, pp. 427-437, 1990.

[17] A. Sahai: Non-Malleable Non-Interactive Zero – Knowledge and Adaptive Chosen-Ciphertext Security. Proc. of the 40th Symposium on Foundations of Computer Science-FOCS'99, 1999.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15, August 2010, pp. 471-474

Research and Design of Mine Electromechanical Equipment Closed-Loop Inspection System Based on Wireless Sensor Network

Cui Lizhi[1], Xu Meng[2], Yu Fashan[1]
[1] He Nan Polytechnic University /School of Electrical Engineering and Automation, Jiaozuo, China
Email: clzh0308@hpu.edu.cn,Yufs@hpu.edu.cn
[2] Northwestern Polytechnical University/ College of Software and Microelectronics , Xi'an, China
Email: xumengpanda@gmail.com

***Abstract*—This paper has analyzed the Significance of applying point-Inspection-system in Coal mining enterprises and proposed a solution of mine electromechanical equipment closed-loop inspection system. Handheld network terminal embedded by WinCE has been adopt to collect inspection information and manage Maintenance Workers. Wireless-LAN technology has been used here to establish the connection between handheld network terminal and intranet. Zigbee technology has been applied to obtain inspection data. The test has proved that this system has solute Practical problems exist in point-Inspection-system effectively. This paper has mainly research and designed on wireless sensor network.**

***Index Terms*—point inspection system, wireless sensor network,embedded handheld terminal,wireless LAN**

I. INTRODUCTION

Coal industry is an important energy production industry in our country. The application of large number of electromechanical equipments have reduced the labor intensity, increased the efficiency of coal production. However, lots of equipments maintenance work has been brought into coal industry in the meanwhile. Currently, point-inspection system has been accepted in most coal industry as the main equipment maintenance system. But in practice, the point-inspection can not be finished only by human because of large work. So point-inspection instrument based on computer technology has been induced to assist worker to finish maintenance. According to the instrument used in point-inspection, there are several problems when applying to coal industry: (1) Lacking of network function. Most instruments need cable to connect with intranet. (2) Inconvenience of operation. The worker can not operate the instrument with so many buttons under the environment in mine. (3) Incompletion of signal. The signals detected by most instruments are limited to temperature, vibration and speed. (4) Inconvenience of data input. There is no standard database for equipment maintenance included in the instrument.

Against the limitation above, this paper has proposed the closed-loop point-inspection system. As the fig 1 showing, the system has been composed of four parts. (1)Intranet which realizes the network management of point-inspection; (2)Wireless LAN which realizes the connection between intranet and the intelligent point-inspection instrument. (3) Intelligent point-inspection instruments (abbreviated to handheld) which collects the point-inspection information. (4)Wireless sensor network which provides the point-inspection information. Among them, the handheld is the link between point-inspection information and intranet. This paper has mainly done research and design on wireless sensor network.

II. ZIGBEE TOPOLOGY ANALYSIS AND DESIGN

Zigbee is a kind of short-range, low-power wireless network technology based on IEEE.802.15.4. Complete zigbee protocol stack is composed of four parts: application layer (APL), network layer (NWK), middle access control layer (MAC), physical layer (PHY).

A. Zigbee network topology analysis

Zigbee network supports two different kinds of physical device which are full function device (FFD) and reduced function device (RFD). Generally speaking, FFD can be used as network coordinator which can communicate with any other devices. RFD can not be used as network coordinator which just can communicate with FFD. Both of the two different devices must exist in every zigbee network. FFD is responsible for build up network. And RFD connected with FFD to form a data-collecting network[1].

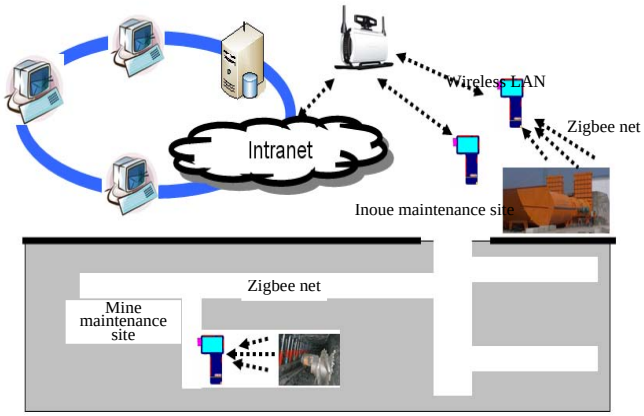


Fig1. the principle of mine electromechanical equipment closed-loop point inspection management system

With FFD and RFD, zigbee has two common network topologies which are star network topology and peer network topology. In star network topology, one FFD communicates with several RFDs. And in peer network topology, every device can communicate directly with each other within the communication. Peer network topology can be realized by very complex network form, in which every device can communicate with target device after following the multi-hop routing.

B. Closed-loop point inspection management system zigbee topology design

According to the analysis above, we have gotten a conclusion that the star network topology is simple to build up, and can be realized conveniently. So this paper has designed the multi-star network topology which is shown in fig 2. In the topology, the net has been divided into different regions according to maintenance sites. In every region, several RFDs have been installed according to the signals collected. And the FFD is movable which is designed in the handheld. There is an electrical tag installed in every maintenance site which includes the site information. There is a RFID reader circuit designed in the handheld. When the workers arrive at the site, the site information is read and recorded by the handheld. After the site is confirmed, the handheld drives the FFD which designed in the handheld to link with the specific RFDs. The handheld collects the point-inspection information after the star topology has been build up. When the data communication finished, the RFDs enter sleeping again in order to save energy. When the handheld comes to the wellhead with workers, it will detect the wireless LAN and upload the point-inspection data to intranet. The network topology designed in this paper is simple and flexible. It not only solves the problem of wireless signal transfer path, but also overcomes difficulty of wireless signal's transmission in mine.

III. HARDWARE DESIGN

According to the topology descript above, we here separately designed the RFD installed in maintenance site and the handheld. The chip named CC2430 produced by CHIPCON is adopted here to form the RFD. And embedded technology is used here to design the handheld.

A. Hardware design of RFD

The principle of the RFD's hardware is shown in fig 3. In the circuit, the CC2430 is used as the control center. The CC2430 is a kind of SoC CMOS component which embedded with an enhanced and low-power 8051 micro-controller. The 14 bit ADC integrated in the CC2430 can be used to convert the analog signal from sensors. The radiofrequency transceiver which is operating in the 2.4GHz has the advantage of high sensitivity and great immunity. The current is lower than 27mA at the receiving and transmitting mode. CC2430 also can run at the sleeping mode, which can hop into active mode at a short time. This feature is fit to our system which need long battery life and long time to run[2].

Besides, the DS2401 is used as the unique identifier memory in which a unique 64 bit license is stored. Except ground pins, DS2401 just has one function pin which finish the power supply and data input-output. Because the power is supported by two batteries, MAX1724 is used to stable the system operating voltage.

B. Hardware design of handheld

Handheld adopts S3C6410 to be the controller which is a kind of ARM chip based on ARM1176. This chip contains many hardware structure and peripheral circuits such as AXI bus, AHB bus, APB bus, sports video process circuit, audio process circuit, 2D accelerator, 3D accelerator, camera interface, TFT 24 bit true color LCD controller, 4 channels of UART, 32 channels of DMA, 4 channels timers, general I/O port, I2S bus, I2C bus, USB host, high speed USB OTG, SD host, high speed MMC card interface and inner PLL. The Falsh/ROM/DRAM port support several kind of external memory such as NOR-Flash, NAND-Flash, OneNAND, CF, ROM. According to the features refer to above and the site requirements, this paper designed the handheld whose principle is shown in fig 4.

In the fig 4, the main peripheral devices connected with S3C6410 and their function are descript below[3].

(1)WI-FI module. The handheld connects with intranet through WI-FI module, and finishes the download of maintenance task and the upload of maintenance information.

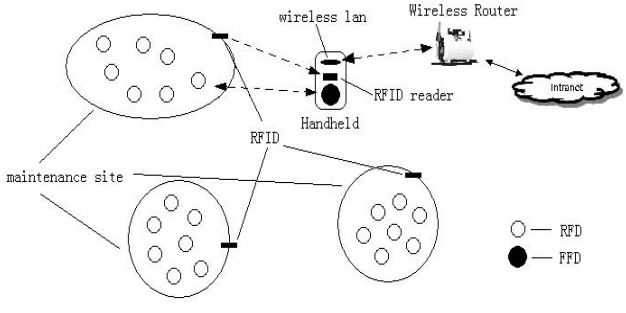


Fig2 zigbee network topology

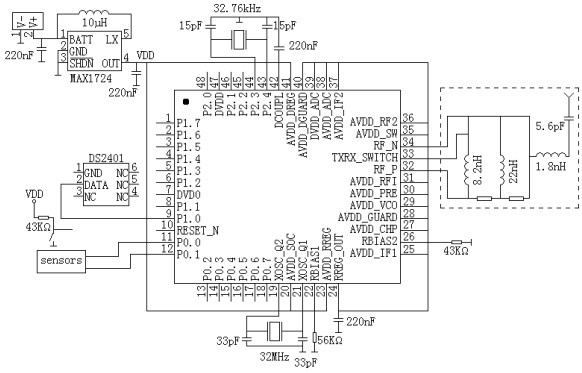


Fig3 circuit of RFD

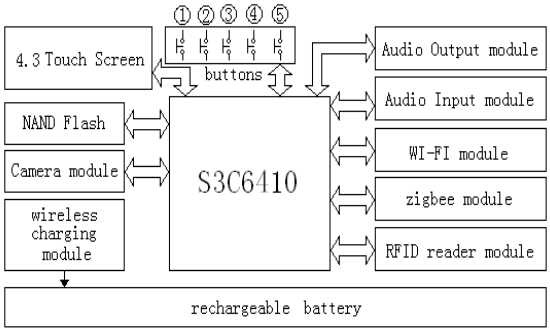


Fig4 hardware schematic of handheld

(2)RFID reader module. The handheld reads workers' public license and electrical tag installed in maintenance site, and records the workers' information, inspection line and site information.

(3)Zigbee module. After recording the maintenance site, the handheld wake up the RFDs in the site through zigbee module. And after RFDs enter the star network, the handheld receives the date from RFDs and processes. The zigbee module here is similar with the RFD in fig 3, which is not descript here. The difference is that the CC2430 connects with S3C6410 through UART.

Besides, a 4.3 inch touch screen is used to show the graphical interface. A NAND Flash is used here to store information. A wireless battery charging module is adopt in order to realize all-closed shell design. A audio I/O module and a camera module are used to collect audio-visual information on-site.

The point need noting specially is that the environment for mine electromechanical equipment is very bad. So it is impossible for the worker to use the point-inspection instrument as a mobile-phone. We design five buttons on the handheld for convenient operation, whose function are power on/off, forward, backward, enter and cancel. In inoue, the handheld can be operate through touch screen, and just the five buttons can be operate when the handheld used in mine.

IV. SYSTEM SOFTWARE DESIGN

According to the zigbee network topology, this paper has designed separately the data acquiring program in RFD, zigbee module program in handheld, and handheld main program.

A. Program in RFD design

Because the energy of the zigbee RFD is supplied by two batteries, all the RFDs enter sleeping mode if not work in order to save energy. In the sleeping mode, just the radiofrequency transceiver works in monitor state. When the handheld sends signal, the RFD will be interrupted into active mode. And then, the RFD convert the voltage information from sensors and transfer them to the handheld.

Fig 5 has shown the NS flow of interrupt program designed in RFD. When the RFD enters interrupt program, it will build up the link of the network, then finish the receive setting. If the data is not received, the RFD will enter sleeping mode again after series inquire.

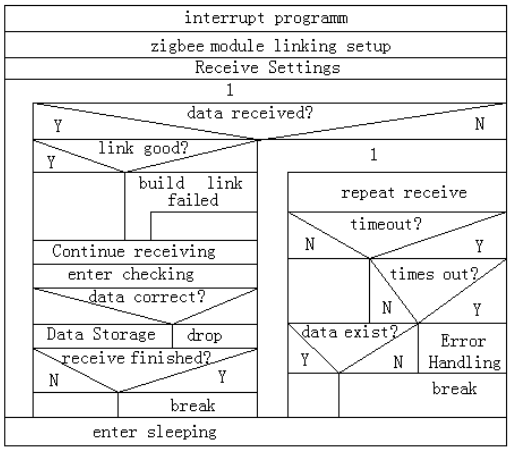


Fig5 NS flow of sensor net code

If the data has been received, RFD will determine whether the network link is normal according to the data frame head firstly. If the link is abnormal, RFD will build up the link again until the link is normal. When the receiving is finishing, RFD will store or drop the data after verification. After the communication, the RFD enter sleeping mode again.

B. The program design of zigbee module in the handheld

The zigbee module is connected with S3C6410 through UART. When workers drive zigbee module to operate through the handheld, the interrupt program will be triggered. The interrupt program wakes up the zigbee module. After waken up, the RFD judges the instruction from UART and operates accordingly. The kinds of instruction include read data and enter sleeping. This program is simple, and is not descript here.

C. Handheld main program design

Fig 6 has shown the architecture of program development based on WinCE. The bottom layer is hardware descript in fig 4. Board Support Package (abbreviated to BSP) located between hardware layer and operation system layer, which helps the system operating on the hardware designed. Operating system layer is the core of WinCE, which provides interface and service for BSP and application program. And the top layer is the user application layer[4].

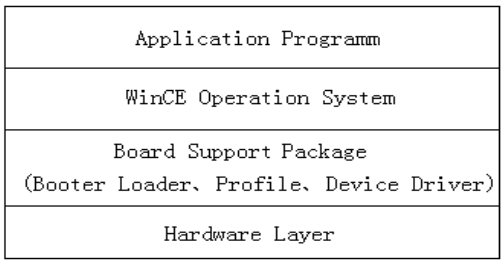


Fig 6 WinCE architecture

In this design, the handheld operates the zigbee module through UART. The handheld transfers instruction to zigbee module and receive the data acquired from RFDs to process, display and store.

V. ANALYSIS OF TEST RESULTS

We install ten RFDs in a laboratory with $40m^2$ area. And every RFDs is connected with a potentionmeters to simulate the signal from sensors. Two different electrical tags are install and besides the laboratory's door which simulate two different maintenance site. Testing person reads one the electrical tag by the handheld, four of the ten RFDs is waken up and transfer data to the handheld. Then the testing person reads another electrical tag, the other six electrical tag also work normally. The result shows that the design is feasible.

REFERENCES

[1] WANG Dong, ZHANG Jin-rong, WEI Yan, CAO Chang-xiu, TANG Zheng ,"Building Wireless Sensor Networks(WSNs) by Zigbee Technology", Journal of ChongQing University(Natural Science Edition). ChongQing, vol. 29, pp. 95–97,110, August 2006.

[2] Yan Lian-long,Luo Jun, "Design of mine monitoring network nodes based on zigbee technology", International Electronic Elements, Xi'An,vol.16, pp.56–58,August 2008.

[3] He Dingxin, Ye Gang, Xu Jinbang, "Research of CAN bus device driver by using WINCE", J.Huazhong Univ. of Sci. & Tech.(Nature Science Edition),WuHan,vol. 9,pp.104-106,2007.

[4] Deng Chengzhong, Huang Weigong, Wan Songfeng, "Design of Small System for Monitor and Control Based on Embedded ARM & WinCE", Control & Automation,Bei Jing, vol. 21,pp.47-49, 2005.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15, August 2010, pp. 475-479

Cellular Automata to Study Mode-I Crack Propagation

He Junlian[1], Li Mingtian[1, 2]
[1] Department of Civil Engineering, Shandong Jiaotong University, Jinan, China
hejunlian@hotmail.com
[2] Geo-technical and Structural Engineering Research Center, Shandong University, Jinan, China
lmt21st@163.com

***Abstract*—Mode-I crack propagation in quasi-brittle material such as rock and concrete is studied by a new numerical method, lattice cellular automata. Cellular automaton method is an efficient method that simulates the process of self-organization of the complex system by constructing some simple local rules. It is of the advantage of localization and parallelization. Lattice model can transform a complex triaxial problem into a simpler uniaxial problem as well as consider the heterogeneity of the materials. Lattice cellular automata integrate advantages of both cellular automata and lattice model. In this paper the importance of the study of the mode-I crack propagation, fundamentals and applications of cellular automata are briefly introduced firstly. Then the cellular automata model is presented, and in order to verify lattice cellular automata, the propagation of mode-I crack in homogeneous material is studied. Results of the numerical simulation are in good accordance with the experimental results and theoretical results of classical fracture mechanics. Furthermore, based on lattice cellular automata, the size effect of quasi-brittle materials is studied. The simulation results agree well with the size effect presented by Bažant. Finally the influences of length of crack on the propagation of mode-I crack are studied. The simulation results are also consistent with the results of classical fracture mechanics.**

***Index Terms*—cellular automata, numerical method, mode-I crack propagation, size effect, heterogeneity**

I. Introduction

There exist cracks with diverse scales in the natural rock mass. And all kinds of cracks arise during the construction of concrete. Failure of quasi-brittle materials such as rock and concrete is related to the initiation, propagation and coalescence of the cracks. Quasi-brittle material such as rock and concrete is sensitive to tensile strength. So it is very important to study the tensile fracture process, which is about the propagation of mode-I crack in fracture mechanics. The propagation of mode-I crack has been studied extensively by classical linear or non-linear fracture mechanics. But the classical fracture mechanics is based on the homogeneity, when it is applied to quasi-brittle materials, many difficulties arise. Kaplan [1] and Mindess et al. [2] detected the fracture process zone before the crack. Bažant et al [3] pointed out neither continuum material models including local, non-local continuum models and stochastic finite methods nor fracture mechanics including linear elastic fracture mechanics and non-linear fracture mechanics can realistically model the fracture process of rock, concrete and so on. Hudson [4] firstly observed the strength of rock samples changed with the size of rock samples, which was called as size effect. Rilem [5] experimented on the size effect of concrete in detail and drew a conclusion that strength of samples would increase with the decrease of size. Bažant et al. [3,6,7], Guinea et al. [8] and Tang et al. [9] studied the size effect of quasi-brittle material such as rock, concrete on theory and numerical simulation. Bažant et al. [7] pointed out that the size effect of quasi-brittle material was due to its heterogeneity. Previous studies have proved that heterogeneity was the underlying cause, which leaded to the complexity of fracture process of quasi-brittle materials.

Lattice model is a typical meso-model. The thought of lattice model appeared about fifty years ago. But because the computation was too slow at that times lattice model was only taken as a theoretical model. With the development of the computer lattice model was also used to simulate fracture process of heterogeneous materials. Schlangen et al. [10,11] and Van Mier et al. [10,12] firstly used lattice model to model fracture process of concrete. Lattice model is able to simplify a complex tri-axial problem into a uniaxial problem and work with simple fracture law. On the other hand, because local poles will be removed when simulating fracture process, local rule is hoped to be adopted.

Cellular automata were presented by Von Neumann in 1950's to simulate self-organization among biological cells. Cellular automata model is made up of cell, cell lattice, neighbors and rules. The state of one cell depends only on the states of itself and its neighbors at the previous step. Cellular automata have the advantages of time and space discretization, localization and parallelization. These years, with the rapid development of computer techniques, cellular automata model has been widely used in fluid mechanics [13], earthquake [14] and solid mechanics [15]. Based on the scalar cellular automata (physical cellular automata), Zhou et al. [16] simulate the rock fracture qualitatively. Li et al. [17-20] presented lattice cellular automata to quantitatively study the tensile fracture, influences of heterogeneity on rock failure and mechanisms of interaction between two cracks under uniaxial compression.

II. Lattice Cellular Automata

Based on the failure mechanics, lattice cellular automata (LCA) were presented to simulate the fracture process, crack initiation, propagation and coalescence of crack in quasi-brittle materials such as rock and concrete. This model has the advantage of both cellular automata and lattice model. The structure of the lattice cellular automata is shown in Fig. 1. The lattice grid point denotes the cell, which connects its neighbors with beams that have three degrees of freedom; the red region around the cell denotes the region influenced by this cell. The states set of cell is expressed as follows,

$$\phi_i = \left\{\{u, v, \varphi\}, \{f_x, f_y, m\}\right\}_i \tag{1}$$

Where u, v, φ is x-direction, y-direction displacement and nodal rotation, respectively; f_x, f_y, m is x-direction, y-direction force and bending moment, respectively.

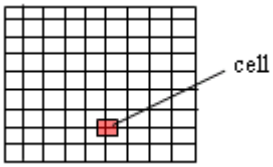


Figure 1. Schematic of cell structure

The local rules, which are used to update the states of the cell, are attained according to the equilibrium equations of each cell, the deformation equations and constitutive equations.

Actually, quasi-brittle materials are heterogeneous. So heterogeneity has to be considered while simulating the fracture process. Lattice cellular automata can be used to study the heterogeneity of quasi-brittle materials. Here heterogeneity is introduced by assigning the mechanical parameters such as elastic modulus E , compression strength f_c following Weibull distribution whose shape parameter m can present the heterogeneity of quasi-brittle materials.

In order to simulate the fracture process of quasi-brittle materials, we must introduce some fracture laws to judge whether the beams start to fail or not. Here two types of fracture laws, tension failure and shearing failure are adopted.

To judge tension failure, the maximum tensional strain criterion is adopted and is expressed as

$$\varepsilon \geq \varepsilon_{t0} \tag{2}$$

Where ε is the strain of the beams, ε_{t0} is the strain corresponding with the uni-axial tensional strength. Here the ultimate tensional strain ε_u is introduced to judge fully tensional fracture.

To judge shearing failure, Mohr-Coulomb criterion is adopted and it can be expressed as below,

$$F = \frac{1+\sin\phi}{1-\sin\phi}\sigma_1 - \sigma_3 \tag{3}$$

Where θ is frictional angle of the meso-element; f_c is uniaxial compression strength of meso-element; σ_1 and σ_3 are the major principal stress and minor principal stress respectively.

So the basic thought of lattice cellular automata to simulate failure and crack propagation of quasi-brittle materials can be expressed as:

1. To divide materials into equivalent lattice model.
2. To introduce the heterogeneity of the rock by assigning the mechanical parameters such as strength and elastic modulus.
3. To update the states of all the beams according to local rules.
4. To judge whether the beams fail or not according to the fracture laws.
5. To implement failure analysis for the failed beams with failure mechanics in order to be able to analyze the strain softening, simulate the discontinuity caused by fracture based on lattice cellular automata.

III. NUMERICAL SIMULATION AND DISCUSSION

A. Influence of Heterogeneity on Mode-I Crack Propagation

Heterogeneity of quasi-brittle materials can be considered conveniently using lattice cellular automata. In order to study the influence of heterogeneity on tensile fracture, a mechanical model with single-side pre-existing crack (shown in Fig. 2) is adopted and two samples will be studied. Here elastic modulus E=20GPa, compressive strength $f_c = 40$Mpa and tension strength $f_t = 4$Mpa. One is homogeneous; the other is heterogeneous and the Weibull distribution parameter $m = 4.0$ (shown in Fig. 3).

The fracture process of the homogeneous sample and its corresponding stress-strain curve are shown in Fig. 4. The fracture process of the heterogeneous sample and its corresponding stress-strain curve are shown in Fig. 5.

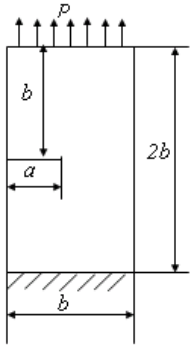


Figure 2. Mechanical model with single-side crack

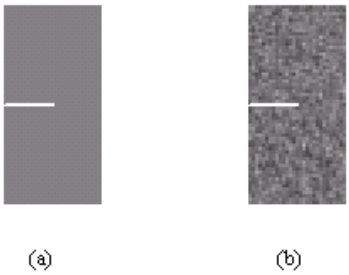


Figure 3. Test samples (a) homogeneous sample (b) heterogeneous sample, here a/b=0.5

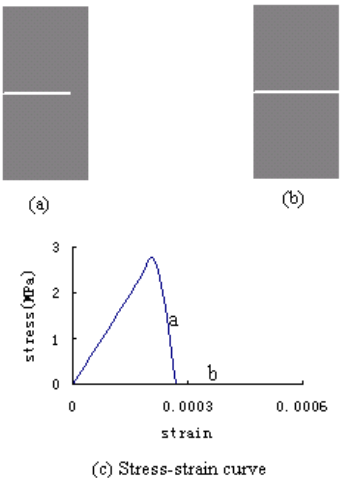


Figure 4. Fracture process of homogeneous sample shown in Fig.3(a)

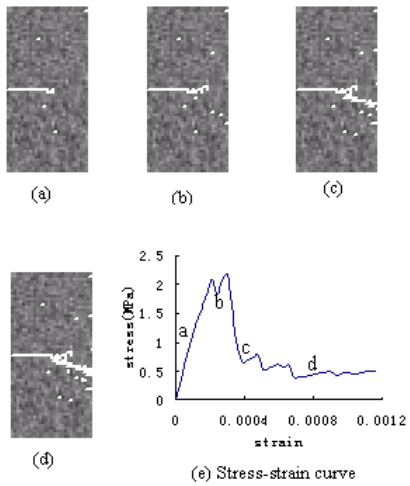


Figure 5. Fracture process of heterogeneous sample shown in Fig. 3(b)

According to Fig. 4, the lattice cellular automata is verified through the comparison between the numerical results and the theoretical results of classical fracture mechanics, where mode-I crack in homogeneous material propagates smoothly along the pre-existing crack . According to Fig. 5, the crack propagated zigzag because of the influence of heterogeneity, but the crack pattern is approximately perpendicular to the direction of load.

B. Size Effect

Experiment is an important method to characterize mechanical properties of material. However some researchers find that the strength of different samples with different sizes made of the same quasi-brittle material differs. It reveals that stress-strain curve is not the real material property but the presentation of material property in the samples. In order to fully understand mechanical properties of quasi-brittle materials, it is necessary to study the size effect of quasi-brittle materials. Here the study of the size effect is based on lattice cellular automata. Four samples with different sizes 100mm × 50mm, 80mm × 40mm, 60mm × 30mm, 40mm × 20mm, 20mm × 10mm (shown in Fig. 6) are studied. The samples are stretched by displacement in the vertical direction. The loaded displacement at each step is 0.0005mm. Here elastic modulus E=20Gpa, compression strength $f_c = 40$ MPa and tensional strength $f_t = 4$ MPa. Elastic modulus, compression strength and tension strength follow Weibull distribution whose parameter m=4.0. Stress-strain curves of these five samples are shown in Fig. 7, which shows that the tension strengths decrease with the growth of the size of the samples.

Bažant et al. [7] presented the size effect that fits quasi-brittle materials:

$$\sigma_N = \frac{Bf_u}{\sqrt{1+\beta}}, \beta = \frac{D}{D_0} \tag{4}$$

Where σ_N is strength of samples with different sizes, f_u is the given norm strength which can be taken as uniaxial tensional strength of the standard sample, β is brittle index, B is an empirical parameter, D_0 is the norm size of sample.

Here $f_u = 4MPa$, $D_0 = 1mm$, size effects of samples with single-side crack can be obtained. According to Fig. 8, numerical results are in good agreement with (4) presented by Bažant. Therefore lattice cellular automata can address the size effect of tensile strengths of quasi-brittle materials very well.

C. Influence of Crack Length on Tensile Fracture

According to fracture mechanics [21], not all cracks are very terrible. If the size of crack can be controlled to a certain limit failure from crack propagation will be avoided. Therefore the length of crack has great influence on the crack propagation and failure of samples. Based on lattice cellular automata several samples with different long single-side cracks are made to study the influence of the crack length on the propagation of the crack and the fracture process. Samples with different long cracks are shown in Fig. 9. The ratio of the length of crack to the width of samples (a/b) is 0.5, 0.25, 0.1 and 0, respectively. Mechanical parameters and distribution parameters of these four samples are the same with sample (c) shown in Fig. 6.

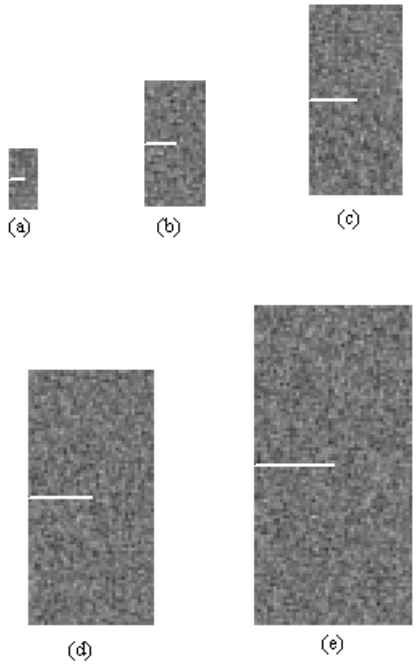


Figure 6. Samples with different sizes (a) 20mm×10mm (b) 40mm×20mm (c) 60mm×30mm (d)80mm×40mm (e) 100mm×50mm

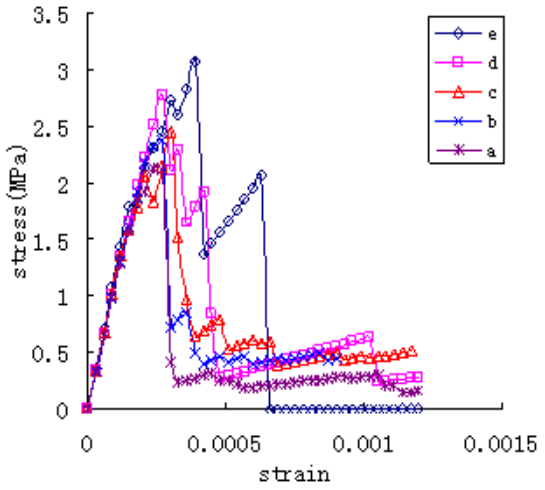


Figure 7. Stress-strain curves of samples with different sizes

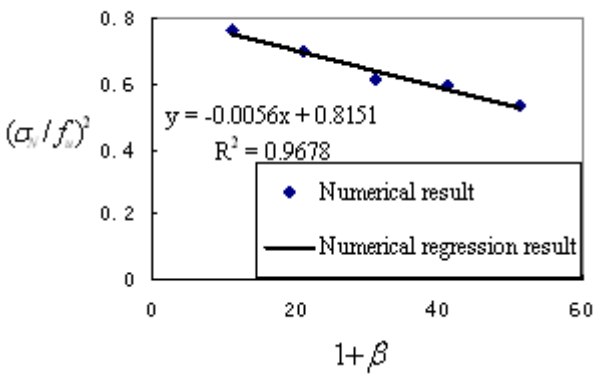


Figure 8. Size effect of samples with single-side crack

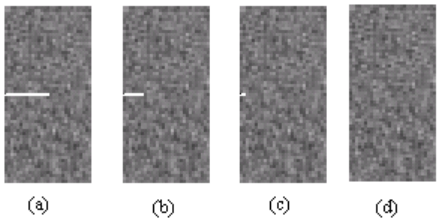


Figure 9. Samples with different cracks (a) a/b=0.5 (b) a/b=0.25 (c) a/b=0.1 (d) a/b=0

Fracture processes and stress-strain curves of these four samples are shown in Fig. 5, Fig. 10, Fig. 11 and Fig. 12, respectively. According to Fig. 5 and 10, crack propagates firstly at tips of pre-existing cracks when pre-existing cracks are longer. On the other hand, because of the influence of heterogeneity cracks propagate zigzag. With the decrease of the length of pre-existing cracks, the influence of heterogeneity on the propagation of cracks is less. When the ratio of the length of crack to the width of the sample is 0.1, the heterogeneity of quasi-brittle materials controls the propagation of the cracks and pre-existing cracks cannot grow and cause failure. So it can be concluded that the length of pre-existing cracks and the heterogeneity of quasi-brittle materials control the propagation of the cracks. And the numerical results also proved that pre-existing crack will not cause failure if the length of pre-existing crack can be controlled in a certain range.

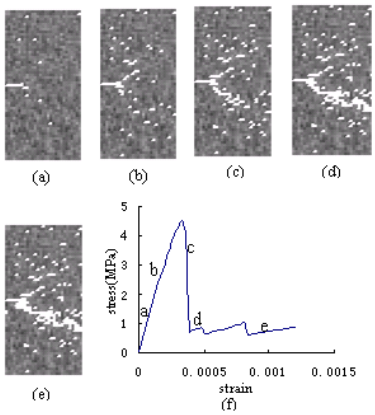


Figure 10. Fracture process of the sample shown in Fig. 9(b)

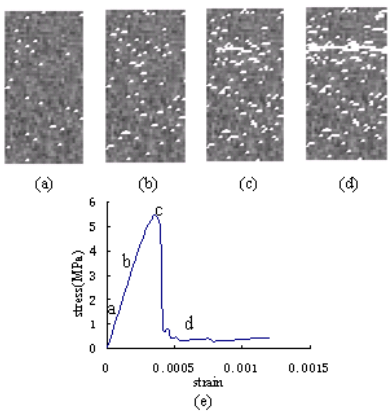


Figure 11. Fracture process of the sample shown in Fig. 9(c)

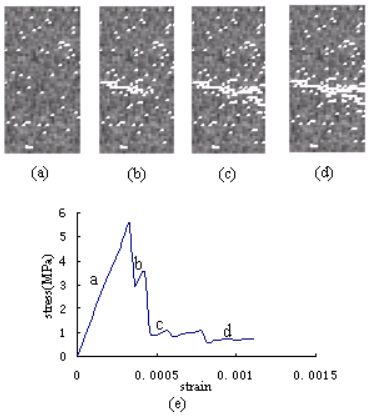


Figure 12. Fracture process of the sample shown in Fig. 9(d)

IV. CONCLUSION

Based on the lattice cellular automata, the influence of heterogeneity and the length of the pre-existing crack on the propagation of cracks are studied. Numerical simulation results are in good accordance with theoretical results and experimental results. Conclusions can be drawn as follows,

1. Cellular automata model is an efficient method to study mode-I crack propagation.

2. Cellular automata can be used to study the influences of heterogeneity on the crack propagation.

3. Cellular automata can be used to study the size effect of quasi-brittle materials.

4. Pre-existing cracks will not lead to failure if the length of crack is controlled in a certain range.

Quasi-brittle materials must be divided into a large number of cell elements if real fracture process is modeled. Thus large-scale computation will be needed. Now parallel computation is considered as an efficient method to solve this problem. Parallelization is easy to be implemented for cellular automata, which is the subject for the next stage.

ACKNOWLEDGMENT

This paper is financial supported by National Natural Science Foundation of China under Grand no.40902083, Shandong Province Natural Science Foundation no.Q2008F03, Promotive research fund for excellent young and middle-aged scientisits of Shandong Province no. 2009BSB01921 and Application Foundation Research of Ministry of Transport of the people's Republic of China.

REFERENCES

[1] M.F. Kaplan, 1961. "Crack propagation and the fracture of concrete," *American Concrete Institute Journa*l, Vol. 58, Nov. 1961, pp. 591-610.

[2] S. Mindess, "Fracture process zone detection," *Fracture Mechanics Test Methods for Concrete.* Edited by P. Shah and A. Carpinteri. Chapman & Hall, London.,1991, pp.231–255.

[3] Z.P. Bažant , M.R. Tabbara, M.T. Kazemi and G. Pijaaudier-Cabot, "Random particle model for fracture of aggregate of fibre composites," *Journal of Engineering Mechanics*, American Society of Civil Engineering, Vol.116, Aug. 1990, pp. 1686 – 1705.

[4] J. A. Hudson, "Soft, stiff and servo-controlled testing machine, a review with reference to rock failure," *Engineering Geology*, Vol.6, Mar. 1972, pp. 159-173.

[5] T.C. Rilem, "Strain softening of concrete in uniaxial compression," *Report of the Round Robin Test*. 1997, pp.195-208.

[6] Z.P. Bažant and B.H. Oh, "Crack band model for concrete," *Material and Structures*.Vol.16, 1983, pp.155 – 177.

[7] Z.P. Bažant and E. P. Chen, "Scaling of structure failure," *Applied Mechanical Review*, Vol.50, Oct. 1997, pp.593-627.

[8] G.V. Guinea , M. Elices and J. Planas, 2000. "Assessment of the tensile strength through size effect curves," *Engineering Fracture Mechanics*, Vol.65, 2000, pp. 189-207.

[9] C.A. Tang, H. Liu, P.K.K. Lee, Y. Tsui and L.G. Tham, "Numerical studies of the influence of microstructure on rock failure in uniaxial compression—Part II: constraint, slenderness, and size effect," *International Journal of Rock Mechanics and Mining Sciences*. Vol.37, Apr. 2000, pp.571 – 583.

[10] E. Schlangen and J.G.M. Van Mier, "Simple lattice model for numerical simulation of concrete material and structures," *Material and Structure*, Vol. 25, 1992, pp. 534-542.

[11] E. Schlangen and E.J. Garboczi, "Fracture simulations of concretes using lattice model: computational aspects," *Engineering Fracture Mechanics*, Vol.57,May 1997, pp. 319-332

[12] J.G.M. Van Mier, A.Vervuurt and M.R.A. Van Vliet, "Material engineering of cement-based composites using lattice models," *Computational Fracture Mechanics in Concrete Technology*, EdiCarpinteri,A. and Aliabadi,M. WIT Press, Boston, Southampton,Computational Mechanics Publications. 1999, pp.1~32.

[13] S. Wolfram, 1986. *Theory and applications of Cellular automata*, World Publishing CO. PTE. LTD., 1986.

[14] K. Chen, and P. Bak, "Self-organized criticality in a crack-propagation model of earthquakes," *Physics Review A*, Vol.43,Apr. 1991,pp. 625 – 630.

[15] T. Tatting and Z. Gurdal ,. "Cellular automata for design of two-dimensional continuum structure," *Proceedings of 8th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, 2000

[16] H. Zhou, Y.J. Wang, Y.L. Tan and X.T. Feng, 2002. "Physical cellular automata to simulate rock fracture—part I: Fundamental," *Chinese Journal of Rock Mechanics and Engineering*, Vol.21, Apr. 2002,pp. 475 – 478.

[17] M. T. Li, X.T. Feng and H. Zhou, 2003. "Evolving cellular automata for simulating rock failure," *Chinese Journal of Rock Mechanics and Engineering*, Vol.22, Oct. 2003, pp. 1656-1660 .(In Chinese)

[18] M.T. Li, X.T. Feng and H. Zhou, "2D vector cellular automata model for simulating fracture of rock under tensile condition," *Key Engineering Materials*, Vol. 261-263, Dec. 2004,pp. 705-710.

[19] M.T. Li, X. T. Feng and H. Zhou, "Cellular automata simulation of Interaction mechanism of two cracks in rock under uniaxial compression," *International Journal of Rock Mechanics and Mining Sciences*, Vol.41,Mar. 2004, pp. 452.

[20] B.Q. Wang, M.T. Li, and Q.Y. Zhang, "Cellular automata model to simulate rock three point bending test," *Proceedings of RaSiM7(2009): Controlling Seismic Hazard and Sustainable Develpoment of Deep Mines*, C.A. Tang(ed), Rinbton Press, Aug. 2009, pp. 619-624.

[21] S.Z. Yin, *Theory and application of fracture mechanics*, Tsinghua University Press, 1992.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15, August 2010, pp. 480-482

Temperature Monitoring System of Generator Stator Based on PN Junction Sensor

Xiaoqi Wang
School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, China
Email: wangxiaoqi@hpu.edu.cn

***Abstract*—The paper focuses on the design and realization of the temperature monitoring system of generator stator based on PN junction sensor. This system is developed for the fault caused by overheating temperature of generator stator in the power generation. By real-time temperature monitoring to generator stator, the fault can be detected early, thus reducing economic losses. In view of the particularity of generator stator's environment and the advantages of PN junction sensor, the system is presented. On this basis, the principle of the system, PN junction sensor, hardware circuit and the communication module with host computer are introduced specifically in this paper. In the system, the real time temperature of stator can be shown in host computer, and the alarm can be answered if temperature anomalies occur.**

***Index Terms*—stator, temperature monitoring, PN junction sensor, microcontroller, communication**

I. INTRODUCTION

In the industrial production, temperature information is a decisive reference factor. Monitoring the temperature of industry equipment immediately, relates to the safety of the whole industrial production[1]. In the industry fields of metallurgy, chemical industry, energy resources and building materials, temperature measurement accuracy and rapidity, makes much sense. And the safety production of power has a relation to the stabilization of the various walks of life, the society, and even people's life. As the equipment of the power generation, generator stator's temperature is over-heat, which may lead to the halt of the power generation. The result is terrible.

The temperature rise is one of the major technical indicators of the generator. If the rise is too high, normal operation of generator will be affected. Further, power interruption may occur, and consequences would be disastrous[2]. In order to ensure the safe and reliable operation of generator, temperature monitoring of generator stator must be done on power generation site at any time. So, the design and realization of the temperature monitoring system of generator stator, is of great practical significance.

II. BRIEF INTRODUCTION OF THE SYSTEM

The temperature monitoring system of generator stator based on PN junction sensor, consists of PN junction sensor, operational amplifier circuit, A/D signal acquisition module, signal processing module, communication module and display or alarm in host computer. Figure 1 shows the schematic diagram of the system.

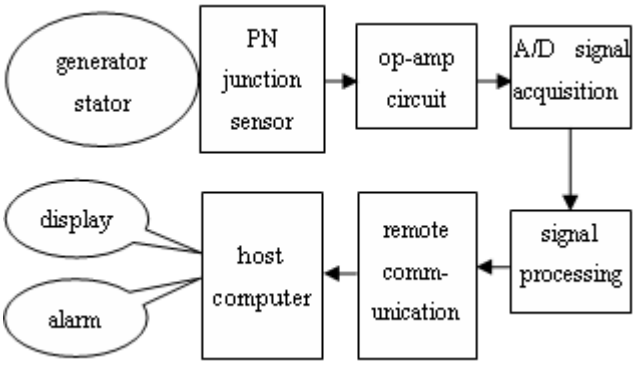


Figure 1. Schematic diagram of the system.

A. Principle of temperature measurement

Distribution of stator temperature is uneven. To ensure that all stator temperature anomalies can be detected, a number of PN junction sensors were placed in the stator wall. PN junction sensor converts the temperature signal of the generator stator into a voltage signal, which is reversed and amplified by the op-amp circuit, with linearly proportional to the temperature signal. Voltage signal through the op-amp circuit is collected into signal processing module, which may consists of microcontroller chip or DSP chip. After being computed and processed, it's converted to digital signal, through remote communication, then displayed and answered the alarm in host computer[4].

B. PN junction sensor

DA linear PN junction temperature sensor is a new type of semiconductor-sensitive devices, which is a negative temperature coefficient temperature-voltage conversion device[3]. It has not only the advantages of thermocouples, platinum resistance, thermal resistance, but also can overcome some defects that the traditional temperature measuring devices are difficult to overcome. As a new measuring apparatus, it is an essential basis component in automatic control technology and instrumentation industries. It is 50 times more sensitive than the thermocouple in K sub-degree, 30 times better than the thermal resistance in linearity, and 20 times faster than more platinum resistance at response speed. It is small, has fast response, good linearity, and good interchangeability. So it is the ideal temperature zone temperature sensing device. In ambient temperature region of -30 ~ 200 ℃, this linear PN junction temperature sensor has an extremely bright future in

applications. And it completely gets rid of the traditional temperature sensor's non-linearity, which bothers electronic designers long. Without compensation of the linear network, complex design and calculation, as well as the complexity debug in the production process, it brings great convenience to the design and production.

The temperature of stator is usually less than 120 ℃, and the space around generator stator is limited[5]. While the temperature measuring range of PN junction sensor is -30 ~ 200 ℃, and it's small, inexpensive, possessing good linearity, interchangeability, of no linear compensation, so PN junction temperature sensor is chosen to be used in this system.

The characteristics of diode related to temperature. The forward voltage of diode changes with temperature. PN junction changes in temperature for every 1 ℃, in voltage for about 2mV. The forward voltage of PN junction drops with the temperature T rising, approximately linearly. Therefore, PN junction temperature sensor is made according to the relationship of voltage and temperature. Silicon transistor can be used as temperature sensor, with measuring rang from -50 ℃ to +100 ℃, accuracy about ± 1%. The characteristics of PN junction in voltage and temperature are shown in Figure 2.

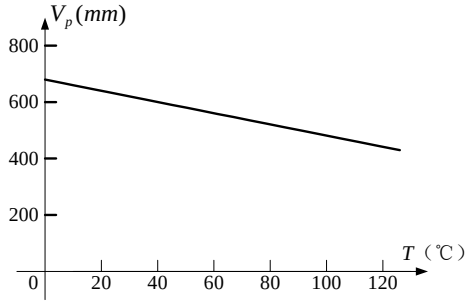


Figure 2. The characteristics of PN junction in voltage and temperature.

III. DESIGN OF HARDWARE CIRCUIT

Taking account of the accuracy of analog information, we choose the avr microcontroller ATmega128[6]. Its precision of AD conversion is 12 bits with 8 channels. It converts the analog voltage signal collected by AD conversion, then computes and processes it. So the digital signal we need is made.

Digital temperature information communicates with host computer by RS485. MAX485 is used here[7].

A. Introduction of RS-485 interface

RS-485 transceiver sends signal in balance and receives differentially, so it has anti-common-mode capability. Coupled with the receiver, it has a high sensitivity, can detect voltage as low as 200mV. So the transmission signal can be restored 1000 meters far away. Using the RS-485 bus, a pair of twisted-pair will be able to achieve the multi-station networked to form a distributed system. Because of the advantages of simple equipment, inexpensive charge and long distance communication, it is used widely.

B. Introduction of MAX485

Pin diagram of MAX485 is shown in Figure 3, and all pin functions are as follows:

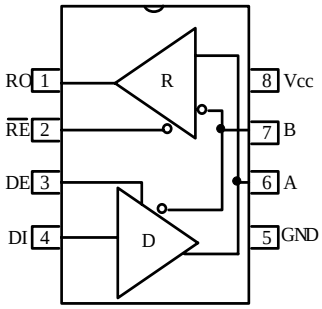


Figure 3. Pin diagram of MAX485.

(1)RO: receiver output. A-B ≥ +0.2 V, RO=1; A-B ≤ -0.2V, RO =0.

(2) $\overline{RE}$: Receiver Output Enable. $\overline{RE}$ =0, receiver output permitted; $\overline{RE}$ =1, receiver output prohibited, and RO is high impedance.

(3)DE: Driver Output Enable. DE=1, driver working permitted; DE = 0, driver working prohibited, and the output terminal A, B is high impedance.

(4)DI: Driver Input. DI = 1, A high output, B low output; DI = 0, A low output, B high output.

(5)GND: ground.

(6)A: receiver input in phase and driver output in phase.

(7)B: receiver input inverting and driver output inverting.

C. Design of circuit

Signal processing and 485 communication circuit in the system[8] is shown in Figure 4.

Amplified signal, processed into digital temperature information by ATmega128, is then sent to host computer. In order to isolate bus and microcontroller system, coupler isolation is needed between the asynchronous communication port of ATmega128 and MAX485, (TLP521-1 and TLP521-2 used here), thus increasing security, and reducing the circuit interference.

IV. CALIBRATION OF THE SYSTEM

Place the PN junction temperature sensor in constant temperature water tank, adjust the liquid medium temperature 0 ℃ (or in the ice water mixture), adequately stir to reach balance and then adjust corresponding potentiometer, so that host computer displays 00.0 (calibrated 0 ℃); then set the thermostat sink temperature of 100 ℃ (or in the boiling water environment), a few minutes after reaching 100 ℃, adjust corresponding potentiometer, so that host computer displays 100.0. After repeating the above-mentioned adjustment

operation several times, the calibration of the system is completed.

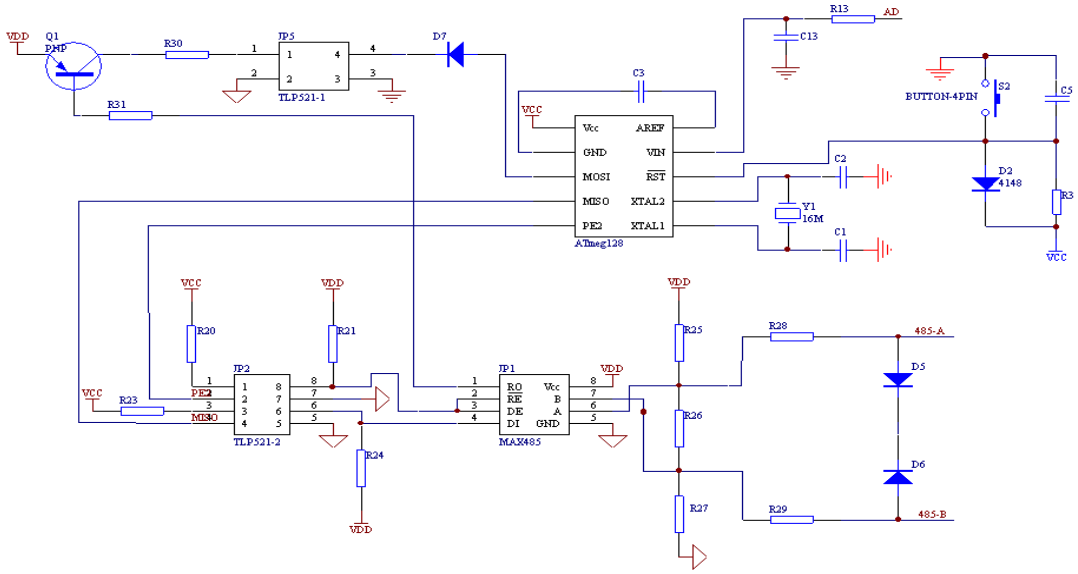


Figure 4 Signal processing and 485 communication circuit in the system

V. CONCLUSIONS

In this system, PN junction sensor is used to extract the temperature signal of the generator stator, which is processed by amplification circuit and microcontroller to form the signal we need. Then it communicates with host computer. This way can spread to more industries. If this real-time temperature monitoring system is used in other industrial machinery equipment, production failures and potential danger can be reduced, and it will be of great practical significance and has bright prospects for development.

REFERENCES

[1] Zhang Xiubo, Monitor and Protection of Winding Temperature of Water Cooling Stator of Generator. Electric Power Construction,Jul.2002:Vol. 23,No. 7.

[2] Zhang Zhenhan, On-line Temperature Detecting Technique for Elecctrical Machines. Shanghai Jiaotong University, 2003.6.

[3] http://cwyq.diytrade.com/

[4] Zhang Lei, Ren Xuejuan, PN junction sensor-based high-precision digital temperature wells Miriam. China Science and Technology Information Mar.2009.

[5] Zhao Hongtao, The Principle and Application of PN Junction Temperature Sensor. Electron IC Engineer, Jul.2006:Vol.32,No.7.

[6] Liu Haicheng, AVR microcontroller theory and measurement and control engineering applications-based on ATmega48/ATmega16. Beihang University Press.

[7] Hu Shishun, Wang Qin, A kind of master-slave microcomputer infrared temperature measurement system[J]. Infrared Technology, 1994, (11):31-34.

[8] H. Ziegler, A low-cost digital temperature sensor system. Sensors and Actuators, Volume 5, Issue 2, February 1984, Pages 169-178.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15, August 2010, pp. 483-486

An E-Learning Resource Pool for the International Promotion of Chinese Language

Huang Xiao-chun
College of Chinese Language and Literature, Wuhan University, Wuhan 430072, Hubei, China
Email: xiaochun.huang@gmail.com

***Abstract*—Chinese is becoming more welcome as a foreign language in the world. Traditional learning resource cannot meet the urgent needs from Chinese language learners in the Internet era. To compensate the serious lack of e-Learning resources for Chinese teaching and learning, a Chinese e-Learning resource pool was introduced, and its system functions, architecture and knowledge system were described. E-Learning technologies were utilized in the resource pool to support collaborative, active study in multimedia network environment. Natural language processing technologies like speech recognition and text information processing can improve the effect of e-Learning so that they were integrated in the resource pool too.**

***Index Terms*—e-Learning, natural language processing, Chinese language learning resources**

I. Introduction

Nowadays there are more than 3,000 institutions of higher learning in 109 countries and regions than have offered Chinese courses. The number of foreigners learning Chinese as a second language around the world is greatly increased in recent years and is over 40 millions now. To satisfy the vast demand of learning and to attract more learners, enormous resources for teaching and learning Chinese have been developed. Confucius Institute online [1] is an official website of Hanban, the executive body of the Chinese Language Council International, aimed to provide overseas Chinese learners information about Confucius Institute news and affairs, resources of Chinese Language and culture as well as interaction center. There are also abundant resources in Learning Chinese with FLTRP [2], such as Confucius classroom, Chinese words, interactive spoken Chinese and many training programs etc... China the Beautiful [3] gives an overall introduction about China including Chinese Science, literature and Language knowledge within simple frame pages. The websites all have rich traditional resources for teachers and learners; however, they are incapable to supply enough e-Learning resources efficiently for oversea learners.

Teaching Chinese to overseas is one of the basic contents of Chinese promotion, meanwhile, learning Chinese the basis for people who want to know China first-hand, and both of them are difficult. Most of the students want to learn Chinese anywhere at any time, and their native languages may be totally different. Their cultural backgrounds, education levels, study motivations may be different, too. Besides of the sophistic situation, Chinese is too hard and mysterious to learn for many foreigners, especially for Indo-European students. To improve the situation, new types of resources and e-Learning technologies are to be introduced to make resources more customized, more humanized and "smarter" than ever.

The paper presents a well-designed large-scale e-Learning resource pool which is aimed to provide what users need as much as possible. Then, as [2] using tone recognition technology to help students master differentiate tones, to make the use of the resource functions more convenient and more intelligent, natural language processing technologies are going to be used to manage the resources. Additionally, modern learning technologies are to be adapted to support Chinese e-Learning activities.

II. System Architecture

At the end of 2009, 282 Confucius Institutes and 272 Confucius Classrooms have been built by Hanban. In this case, furnishing cost-price feasible e-Learning resources is a more pressing issue than ever. The goal of the resource pool is to offer one-stop intelligent resources and services for Chinese language teachers, learners and other enthusiasts.

It is important to choose and apply proper implement technologies. Comparing with traditional learning, e-Learning is more flexibility in time and location, and is more cost-effective for learners. Learners at any place in the world can access shared e-Learning information unlimitedly, and control study pace by themselves [4]. Moreover, natural language processing, especially Chinese information processing technologies will help people release from many hard works.

Taking all the above matters into account, the resource pool should realize several functions within a system architecture that consists of a communication platform and a learning resource platform, including a carefully constructed knowledge system as follows:

A. System Functions

- Multi-goal driven. Teachers can find teaching aids like Chinese textbooks, courseware, reference books, pedagogic tools and the other things that can help them to prepare lessons. Students can receive real-time teaching and need not to face the teacher, and they can do self-study under the instructions in the pool. Other non-Chinese native spoken learners can learn Chinese language and culture as they like.

- Multimedia. Optimized combination of multimedia learning resources would be utilized to improve the efficiency of teaching and learning. Modern pedagogic medium make learning less monotonous and thus more effective.
- Multilingual. The learners are from different countries, and they use several dozens of languages. To meet learners' needs, especially for beginners, multilingual learning resources are to be deployed in the pool. Meanwhile, to make deep impress on learners, the resources would be developed with full consideration of learners' cultural background and unique characterizes of their native languages.
- Interactive. Users would be able to submit feedback and to participate in enriching and optimizing the content of the resource pool. The instruction for them would be explicit and easy to follow. The pool would provide various communication channels including email, forum, blog, instant message system, etc.
- Intelligent. E-Learning environments can provide explanation, tutoring and intelligent diagnosis of students' understanding [5]. Personalization techniques, that is, adaptive hypermedia techniques and filtering and recommendation techniques are used to customize the user learning style and content [6]. On top of that, the paper noticed that natural language processing techniques could help people to manage resources based on dealing with content and to use resources in more convenient and efficient ways. Chinese information processing techniques have made great progress in recently years, and not been applied in e-Learning as they could. The Chinese e-Learning resource pool is to synthesize these techniques to organize Chinese learning resources and offer intelligent services for users.

B. System Architecture

The Chinese e-Learning resource pool is composed of two parts: learning resource platform and communication platform. The pool's system architecture is in Fig. 1.

Users can access the pool by local area network, Internet or mobile devices, and gain or supply Chinese learning resources, or receive video conference teaching. To serve for that, instant messaging, shared whiteboard, custom input facilities are to be utilized in the system. There are two main databases in the system: user profile database and learning resource database. The former stores user information; the latter stores a large amount of learning resources such as archived video lectures, documents, flashes and so on. User management, learning management, learning resource management and system management are the four management modules of the system.

C. Resource Organization

Learning resources are divided into three lays: material, component and course, and resources of each lay consist of various categories as Table 1 shows. The service objects of resources are recorded in the learning resource database such as the teacher's or student's language, level of Chinese education, age group and so on. Users can choose their favorite resources or accept ones that the resource recommendation module chooses for them with consideration of their user profile. Resources in the lower lay could be united into some kind of resources in the upper lay, so that a user can select a few of materials and the system would generate a PowerPoint courseware automatically, or select some components and receive a teaching case from the automatic generating system. Teachers can save time and energy in such a way to some extent.

TABLE I. RESOURCE LAYS AND CATEGORIES

lays	categories
course	one-to-many/one-to-one/self-study, synchronous/asynchronous
learning component	courseware, knowledge document, textbook, exercise, test paper, teching case, literature
material	audio, video, cartoon, text, picture, graph, game, teaching and learning software, webpage material

D. Knowledge System Architecture

There are mass resources structured by a knowledge system in the website of the resource pool. The principles of building such a knowledge system are authority, complete and explicit. The knowledge system covers nearly every branch of Chinese language and culture that are usually taught in Chinese Language classes.

- Language knowledge: Chinese character library, Chinese vocabulary library, Chinese phonetics library, Chinese grammar library, Law library of Chinese language etc.
- Culture knowledge: religion culture, philosophy culture, utensil culture, folk culture, science and technology culture, aesthetic culture, academic culture and political culture.

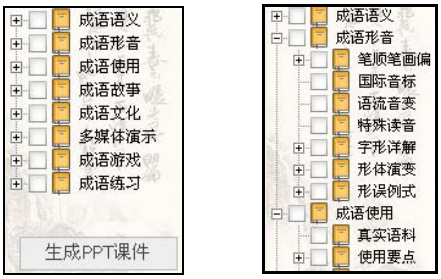


Figure 2. Structure of the Chinese idiom bank

Chinese idiom bank is one branch of the Chinese vocabulary library, which structure is showed in Fig. 2.

III. SUPPORTING TECHNOLOGIES

E-Learning technologies and natural language processing technologies can do favor to the resource pool in different ways.

A. E-Learning technologies

1) Annotations. Initiate interactive learning on specific topics or terminologies is often difficult for learners. That is, annotations are usually used in strategy learning [7]. Content of resources and procedure of teaching and learning could be annotated for advanced application. For example, considering the interanational annotation standards, an particular annotation set is designed to describe the structure of modern Chinese grammar in the knowledge system and tag the content of each resource related to modern Chinese grammar. With the annotations, semantic information retrieval system could find resources on some topic rapidly and correctly.

2) Distributed components. Distributed component technology can improve the performances of information processing, system collabration, system robust and extensibility as well as accelerate application development [8]. Sun's Enterprise Java Beans have been deployed in the resource pool to support distributed learning and problem solving.

3) Interactive activities. It is necessary to know learners' oppinions about resources and services as well as to arouse their enthusiasm of collabrative learning.

a) Users defined tags. Users can define tags to annotate resources offered by themselves as long as the tags meet with the specifications of the standard annotation set.

b) Evaluation and recommendation. For each accesible resource in the resource pool, users can give feedbacks about resources and services to help other users to make decision.

c) Communication channels. Users can communicate with the pool staffers or other users throught email, phone, blog, forum, instant message and online question answering system in the website of the resource pool.

B. Natural language processing technologies

Most contents and content descriptions of learning resources are in natural languages. Natural language processing technologies are essential for comprehensive applications [9].

1) Machine Translation. The resources and services in the resource pool must be multilingual translated. Since the students are from dozens of countries and speak different languages, traditional human translation are too expensive and time-consuming. Machine translation could do a great favor to solve the problem, even though the technology is still far from perfect now.

a) Translate shared content. Resources about Chinese basic knowledge points, web pages, instructions and help information are needed by all of the users.

b) Real time interpretation. Even if there is only one teacher in a online classroom that is utilized with a real-time interpretation system, Students with different native language can study together.

2) Speech recognition. Phonetic technologies are useful in language learning.

*a) Phonetic learning.*When a learner studies how to pronounce by himself, speech recognition software can compare his pronounciation with the standard one, then the student can correct pronunciation by the result. In addition, the learner can make dictation without a teacher.

b) Captioning for videos. There are large amount of videos which content are not described in text, and thus can not be retrieved effectively. To solve the problem, usually videos should be viewed and described by human. By automatic captionning, the laborious work can be done by a software that combines speech recongnition and automatic summarization technologies.

c) Situation Dialogur exercises. A speech recognition system is capble of "talking with" learners to help them practise oral Chinese. Learners can master a larger vacabulary and learn more about Chinese tone and flow changes in real spoken situations.

3) Text information Processing. The following technologies are quite useful in e-Learning systme:

a) Text categorization: resources newly coming can be classified automatically accroding to their text content, and learners be grouped according to their profiles. A customized recommendation system can push a proper kind of resources to the right group of users.

b) Text-to-Speech: Text will be read by a text-to speech system. In such a way, students can do listening practice or learn how to pronounce. In some cases, the technology can be used as screen reading.

c) Semantic information retrieval. With the help of annotation system, resources are semantically annotated, so that information retrieval system can search out what are the user really need.

d) Question-answering system. Frequent questions and answers can be recorded in the system. whenever a user want to get help about the resource pool, he will get an answer by the question-answering system. With it, Time-zone would not be a problem.

In addition to the above natural language processing technologies, there exists plenty of other technologies that have been or can be used in Chinese e-Learning such as Chinese input methods and automatic summarization technology. For instance, to force learners to memorize Chinese tones, Chinese character Pinyin input methods could be modified to display character only after the right tone is given.

IV. CONCLUSION

The Chinese e-Learning resource pool aims to provide one-stop learning and intelligent services for Chinese language teachers, students and enthusiasts. Besides well-designed Chinese knowledge system, modern e-Learning technologies and natural language processing technologies are to be utilized to permit collaborative learning and reduce the cost of Chinese promoting around the world. Parts of these have been applied in the resource pool, which has already been built and deployed on line [10]. With the mature of science and technology, more humanized applications and teaching medium will be expected in the pool, such as teaching and learning resources facing the disabled, touchable learning system and so on.

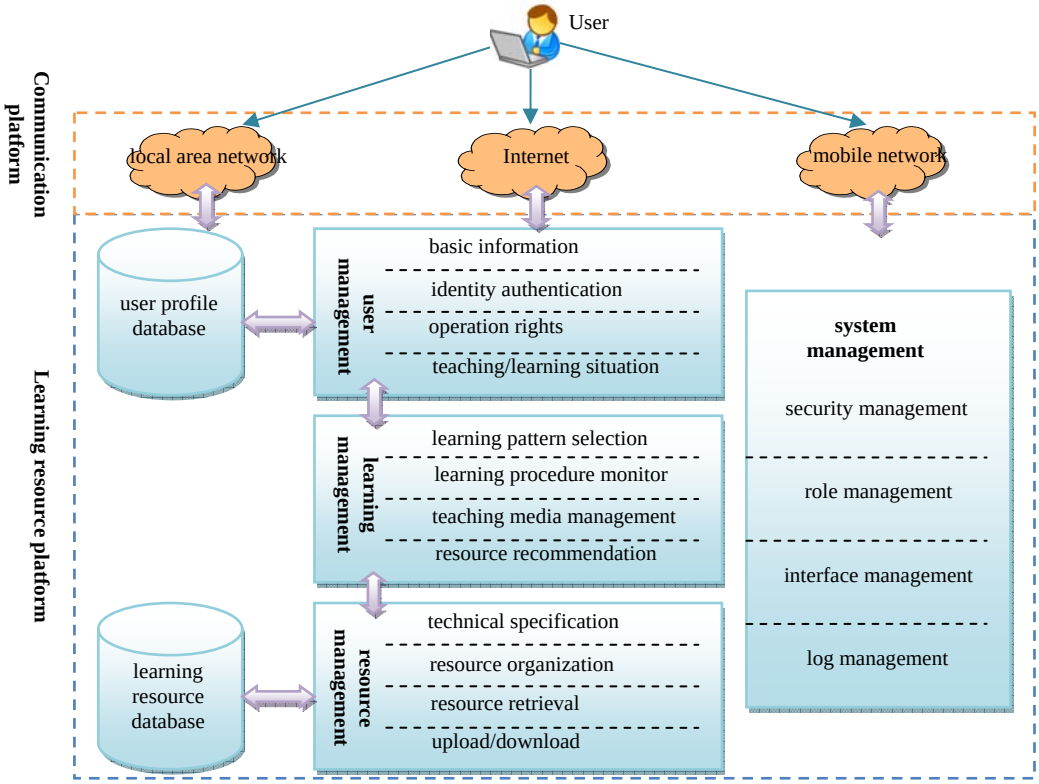


Figure 1. System architecture of the e-Learning resource pool

ACKNOWLEDGMENT

This work has been supported by Hanban project WDJD0003 and 211 young teachers' scientific research project fund sponsored by the College of Chinese Language and Literature, Wuhan University.

REFERENCES

[1] Hanban. "Confucius Institute online". http://www.chinese.cn/. Acessed May 30, 2010.

[2] Foreign Language Teaching and Research Press. "Learning Chinese with FLTRP". http://www.chineseplus.cn/ . Acessed May 30, 2010.

[3] Pei Ming-long. "China the Beautiful". http://www.chinapage.com/main2.html. Acessed May 30, 2010.

[4] Sarmad Mohammad. "Effectiveness of E-Learning System". *2009 International Conference on Computer Engineering and Technology*, January 2009, pp. 390-394, doi: 10.1109/ICCET.2009.150.

[5] Brusilovsky, P., Eklund, J., & Schwarz, E.1 67-207. "for developing adaptive courseware". Seventh International World Wide Web Conference. *Computer Networks and ISDN Systems*, 30 (1-7), pp. 291-300.

[6] Magoulas, G.D., Chen Sherry, Y. (Eds), "Individual Differences in Adaptive Hypermedia", *Proceedings of the AH 2004 Workshop*, 2004.

[7] Ping-Lin Fan, Su-ju Lu, Han-Jang Wu, Chih-Ming Chen. "High Interactive Web-Annotation Based e-Learning Platform", *2010 IEEE International Conference on Digital Game and Intelligent Toy Enhanced Learning*, April, 2010, pp. 198-201, doi: 10.1109/DIGITEL.2010.58.

[8] Ed Roman, *Mastering Enterprise JavaBeans Third Edition.* John Wiley & Sons, Inc., January 2005.

[9] Tajudeen A. Atolag. "E-Learning: The Use of Components Technologies and Artificial Intelligence for Management and Delivery of nstruction", *24th Int. Conf. information Technology lnterfaces*, June 24-27, 2002, Cavtat,Croatia, pp. 121-128.

[10] Wuhan University Teaching Resource Research and Development Base for International Promotion of Chinese. "Resource Pool for Teaching and Learning Chinese as A Foreign Language". Http://www. hy123.org. Acessed May 30, 2010.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15, August 2010, pp. 487-491

A Web-based Agricultural Decision Support System on Crop Growth Monitoring and Food Security Strategies

Wang Zhi-Qiang, Chen Zhi-Chao
Henan Polytechnic University, Jiaozuo, 454000,China

***Abstract*—Food shortage has been among the most threatening problems to the world since the beginning of the new century. And it means more to the countries with a large amount of citizens such as China. Chinese governments at different levels have being taken different kinds of actions to stabilize and increase the yields of grains. A major premise of making right decisions is the ability to accurately assess crop growth and food supply, and an scientific decision-making process to provide appropriate strategies or countermeasures based on them. This can be accomplished partly by using the decision support system (DSS) that provide accurate and detailed information about crop growth and food supply. In this paper, an agricultural spatial DSS (ADSS) frame was studied and developed to meet the increasing demands. The ADSS was aimed at suggesting efficient strategies for problems in crop growth and food safety as well as providing timely and accurate information about crop growth and food supply. The system, based on the spatial information technologies and crop growth simulation methods, contains three parts: (1) a spatial agricultural resources data warehouse has been constructed; (2) a crops monitoring and simulation package was studied and developed; (3) a spatial decision support package for food-supply security developed. The ADSS has been applied to the Northeast China and been proven to be a successful tool for crop growth monitoring and food security strategies.**

***Index Terms*—Crop growth monitoring; food supply security; ADSS; Northeast China**

. INTRODUCTION

The latest years has seen mankind confronted with the problem of food security worldwide. China, with 7% of total arable land to feed 22% of total population of the world, has long been a focus of public attention (Cheng, 2005a). It has always been, therefore, the primary importance among priorities for Chinese governments at different levels to keep national or regional grain balances of demand and supply among all the grain macro-management activities. While timely and accurate information about regional crop growth and grain production is among the premises of food (grain) supply security macro-management for the governments. Moreover, with the pace of the regionalization of the food production bases accelerating, the major food-producing areas is playing increasingly important role in the national food security system(Cheng, et al. 2005b). The capability to accurately predict the crop growth and food production in these districts can contribute a lot to better scientific decision-making on the food-supply security at regional and national level (Wu Bingfang, 2004).

The application of decision support system(DSS) in the fields of agriculture provides a new and efficient method for improving the management of regional food production mode and decision-making on the management of food security (Cao Weixing, et al. 2007). The space technologies such as RS, GIS, now increasingly powerful tools, give support to agricultural DSS (ADSS). Crop-growth monitoring with RS can monitor the growth of crops and predict grain production more objectively and rapidly. RS, thereat, become a fundamental component and tool of DSS. GIS, being powerful in the data compilation, storage, management and spatial analysis, are widely used as the platform of an agricultural DSS, various crop growth models, grain production estimation models and other models extend extensively the functions of the agricultural DSS. Furthermore, with the rapid development of internet/intranet technology, Web technology is growing up to a new branch in the development of ADSS (Mei Fangquan, 2003). The application domains of agricultural DSS are mainly restricted to a specific field(Rui Xijie, et al.2005; Zhu Yan, 2004), while it comes to regional agricultural macro-management, little DSS research has been reportedly done (Zhao QianJun, 2005). This paper presents the development of a DSS for the macro-management of the regional agricultural activities and food security in Northeast China, which has been one of the key grain production bases of China. The proposed DSS has the ability to suggest solutions for common problems in agricultural practices, which include crops growth monitoring, simulation, grain production estimation and the macro-management of agricultural production activities and food security.

. THE STUDY AREA,RESEARCH SINGIFIC ANCE AND OBJECTIVES

Northeast China has been being among the key food-production bases of China since the foundation of PRC, serving as one of the main sources of merchandized food supply for rice, corn, soybean. Northeast China is playing a increasing important role in food security with the situation of food security is getting worse. And it is also regarded as the largest strategic spare food base in China or a stabilizer of domestic food market (Cheng, et al. 2005b). The irrational agricultural land use structure, disarrangement of agricultural activities, unreasonable

dispose of crop planting structure and frequency of natural disasters , however, resulted in fluctuations of the regional dramatic grain production, hampering the development of the regional agricultural industry and threatening regional and national food supply security. More advanced methods and technologies are needed to manage all the information with different sources, and interruptedly analyze all the information to get better strategies for administration. An ADSS can, obviously, meet the needs. But up to now, there is no such decision support system available in the field.

The paper aimed at developing an ADSS dealing with the development of a DSS for crops growth monitoring and simulation, grain production prediction, providing the decision-making strategies and the early warning of food supply and the macro-management of agricultural activities for the agricultural administration in Northeast China. Therefore, the main objective is to develop a monitoring system for watching crop growth, finding problems in crop growth, selecting appropriate choices. To achieve this objective, the following tasks were accomplished:

(1) Conducting an extensive literature review;

(2) Interviewing experts in agriculture, information system;

(3) Constructing the spatial data warehouse of agricultural resources in northeast China;

(4) Experimental data were acquired by conducting a series of experiments, and then validated and modified with them the models from the literature or other source;

(5) Developing crops monitoring, crops growth simulation and grain production estimation system;

(6) Developing the DSS for food security in Northeast China(including the knowledge base and models base).

Number footnotes separately in superscripts [1, 2, …]. Place the actual footnote at the bottom of the column in which it was cited, as in this column. See first page footnote as an example.

. THE FRAME WORK OF THE ADSS

Based on the characteristics of data and users, a mixed structure of the system was adopted which is a combination of B/S and C/S structure. The architecture of the system is shown in the Fig. 1, based on the traditional 3-tiers architecture of "browser/application server/database server, the application services is divided into two parts according to the logic functions of the system: application server and Web server. The B/S structure is made up of four components: Web browser/Web server/application server/database server. The Web browser brings users UI ands transmit the requests from users; the Web sever deals with and distributes the requests; database server provides data retrieval, storage, modification, etc; while application sever contains various models such as agricultural land use disposal models, crop growth models, grain prediction and estimation models, data access models, etc.

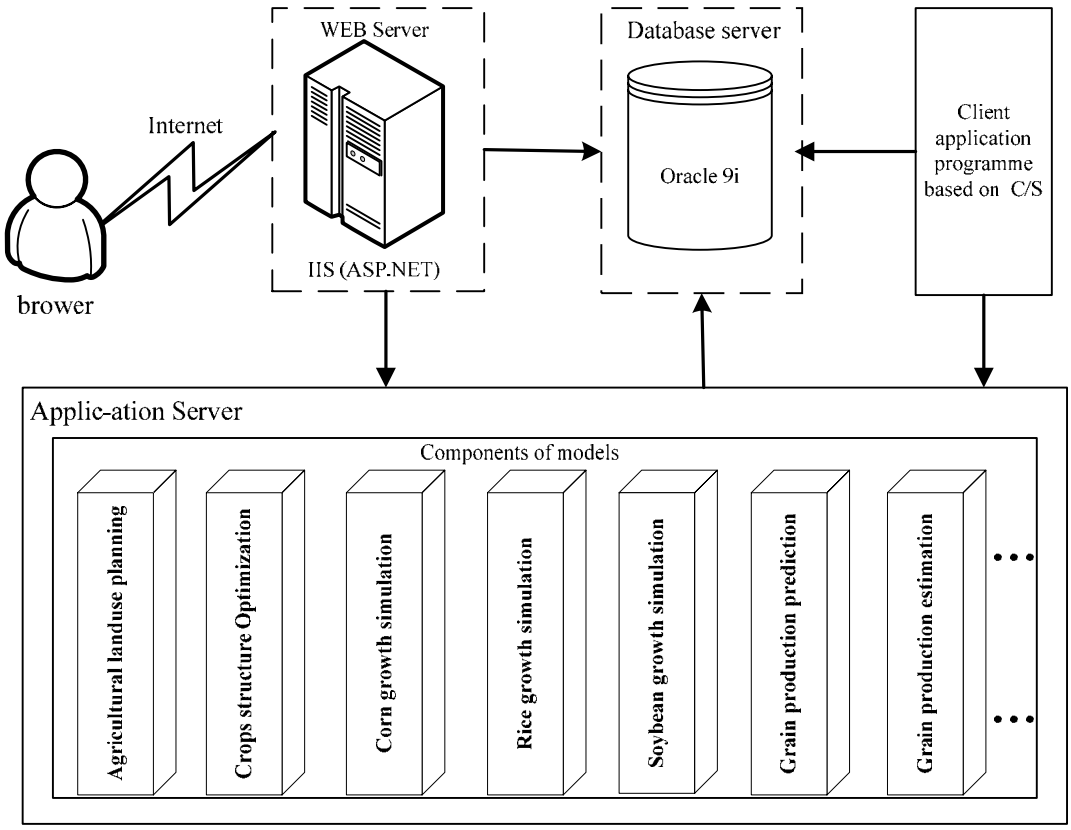


Fig. 1 The framework of the agricultural decision support system(ADSS)

For the system server or other special users, B/S mode may be less convenient and a bit slower, especially when processing a huge amount of images and other operational processes. While C/S architecture makes data processing and other operations run across multiple hosts and brings more efficiency and scalability to the application. In the paper, a C/S mode was added into the system's architecture. The client-server applications in the paper are split into three components: the client component of the application requests or sends information to a central application server or database server, and then application server or database server deals with all the processes and returns the results to the client.

Ⅳ. DEVELOPMENT OF THE ADSS

A Construction of the spatial data warehouse of agricultural resources

The spatial database includes satellite RS images (Land sat TM, MODIS, AVHRR), statistical data (about the regional agricultural economic index, grain production, etc.), agricultural meteorological data, agronomy parameters, agricultural land resources, and data about the environment, information about the demand and supply of grain at various scale.

A database management system has been developed for the spatial data warehouse with the following operations: inputting, storing, managing, querying and indexing data in the data warehouse, regular spatial analysis with data, data and thematic map outputting and database maintaining.

B Development of the crops monitoring system based on RS and corresponding simulation models base

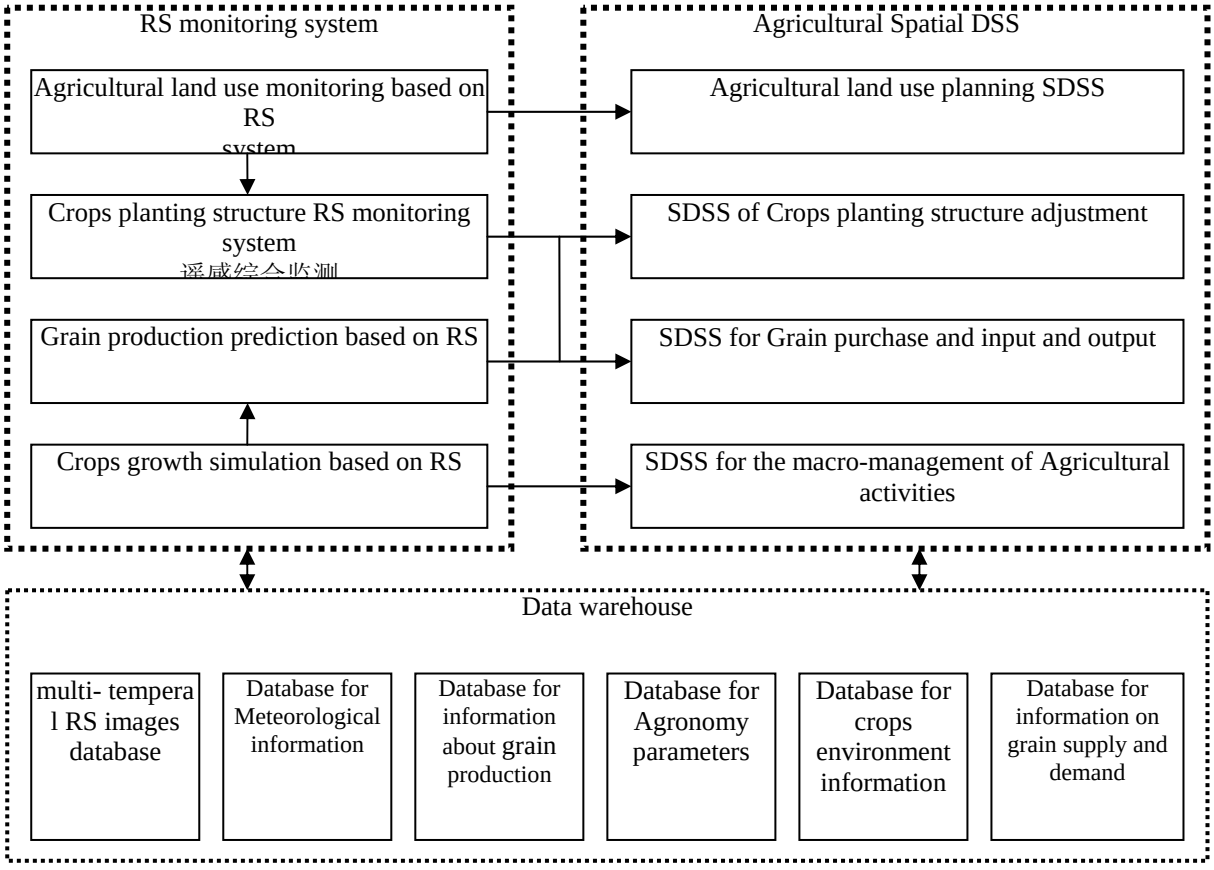


Fig. 2 A sketch of the functions of the system

The system contains monitoring, acquisition, and simulation of crop growth and condition, and prediction of grain harvest based on RS. It consists of 6 components: crop condition monitoring， crop acreage estimation, crops growth simulation and yield prediction， grain production estimation，cropping index and crop structure monitoring, drought monitoring.

C Development of the ADSS for food security and macro-management of agricultural activities

1 Methods

The system was developed using a rule-based shell software available in the market. The basic building blocks of the decision support system are the rules. Rules

were made from the data available from the survey and the literature review.

The typical rules are generally composed of the following components:

Qualifiers: These are questions that the user is asked to provide answers to within an "If-then" frame. The value of the qualifier is fixed and cannot be changed.

Variables and mathematical models: Variables are created to accept input from the users while the mathematical models are used to evaluate and direct the decision process.

Choices: Choices are the product of specific scenarios in the decision support system. The recommendations at the end of a session are the choices. In the decision support system presented in this paper, many measures such as adjustment of agricultural land types, adjustment of regional corps planting structure and other macro-management measures of agricultural activities are the choices.

2 Contents of the Sub-DSS in the ADSS

The sub-DSS aimed at serving the management and agricultural technicians with decision-making choices. It includes the following components:

(1) Spatial decision support on the adjustment of agricultural land-use structures

This module is used for evaluating the adaptability of agricultural land resources with GIS, and analyzes the major factors to bring the environment ecological risks, and quantize the types, amount and spatial distribution of arable land which need to ex-farming.

(2) Spatial decision support on the adjustment of regional corps planting structure

This module analyzes information about the demand and supply of each kind of grain and predict the tendency home and abroad , and then evaluates the regional crops planting structure at present with RS monitoring of crops planting structure, and quantify the types, the quantity and spatial distribution of crop planting structure needed to adjust.

(3) Spatial decision support of macro-management of agricultural activities

This module analyzes grain production potentials with crop growth simulation models and GIS technology; it also analyzes quantitatively the types and effects of hazards which lead to underproduction of crop

(4) Early warning system of Regional grain supply risk.

Based on the early warning model of grain supply risk, it predicts the probable changes regional grain demand on the basis of the prediction of regional population changes and the grain production at different years, then evaluates the status of grain shortage, gives decision-making choices in advance.

. CONCLUSTIONS

A major component of an ADSS is the ability to accurately assess the condition of crop growth and food supply to provide appropriate strategies or countermeasures. This can be accomplished in part by using new space technologies such as RS,GIS. A agriculture DSS for crop growth monitoring, simulation and food security was studied and developed. RS, GIS, crop growth simulation models and early warning model of grain supply risk were incorporated into the package of the system. These technologies and simulation models are discussed in the literature. Yet, a limited number of these methods are utilized together to serve regional micromanagement of agricultural activities.

The ADSS for crop watch and food security discussed in this paper was developed after: (1) Conducting an extensive literature review; (2) Interviewing experts in agriculture, information system, carrying out field survey of various agriculture zones; (3) acquiring experimental data by conducting a series of experiments, and then validating and modifying the models.

However, the system has still some limitations such as (1) limited statistical data about the crops condition and growth in the past years are included in the system's database; (2) most of monitoring data are to be interpreted manually; (3) the scalability of models. The models in the system were only validated by the data from several spots (such as Hailun Station for ecological agriculture), but when they are used in other agricultural zones, deviations or errors will inevitably occur. More detailed validation should be done in the future work.

ACKNOWLEDGEMENTS

This work was jointly supported by Foundation item: national natural fund (No. 40401003), Scientific Research Foundation of Jiangsu Key Laboratory of Resources and Environmental Information Engineering (No: 20080202) and Key Items for Henan science and technology (No: 102102310364).

REFERENCES

[1] Cao Weixing, Pan Jie, Zhu Yan, Liu Xiaojun, (2007). Growth model and Web application-based decision support system for wheat management , Transactions of the Chinese Society of Agricultural Engineering, 23(1): 133-138

[2] Cheng Yeqing Zhang Pingyu, (2005a). Regional Patterns Changes of Chinese Grain Production and Response of Commodity Grain Base in Northeast China , Scientia Geographica Sinica. 25(5): 513-520

[3] Cheng Yeqing, Zhang Pingyu, (2005b). Construction and distribution of grain commodity bases in northeast China-problems and countermeasures, System Sciemces and Comprehensive Studies In Agriculture, 21(4): 264-267

[4] Mei Fangquan, (2003). Analysis of Development and Strategy for Agricultural Inform ation Technology , Review of China Agricultural Science and Technology, 5(1): 13-17

[5] Rui Xijie, Yang Changbao, Ma Shengzhong, (2005). Research 0f macr0sc0pical agriculture decision-making supporting system: Taking corn as sample, Journal of Northwest University(Natural Science Edition), 35(2)163-166

[6] Sherif Yehia, Osama Abudayyeh, Imran Fazal, Dennis Randolp, (2008). A decision support system for concrete bridge deck maintenance, Advances in Engineering Software, In press,39: 202 – 210.

[7] Wu Bingfang. (2004). China Crop Watch System with Remote Sensing，Journal of Remote Sensing, 8(6): 481-497

[8] Zhao QianJun, Xie Gaodi, Li Jun. (2005). Design of management decision support system for agricultural resources on a county scale. Transactions of the CSAE, 21(4): 123—126

[9] Zhu Yan, Shen Weixiang, Cao Weixing, (2004). Dynamic knowledge model and decision support system for rapeseed cultivation management, Transactions of the CSAE, 20(6): 141-144

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15, August 2010, pp. 492-494

Study on the Characteristics of Dielectric Barrier Discharge and Dielectric Barrier Corona Discharge

Zeng Mi[1], Lu Yan[2] ,Sun Yan-zhou[3]
[1]School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo 454000, China
[2] JiaoZuo Coal Corporation, Henan Coal and Chemical chemical Group, Jiaozuo 454000, China
[3] School of Electrical Engineering, Henan Polytechnic University, Jiaozuo 454000, China
E-mail:pqdong@hpu.edu.cn

***Abstract*—The dielectric barrier discharge characteristics of two electrode configurations of flat-flat and line-tube are studied and compared by means of measured voltage and current oscillograms and discharge photos. Based on discharge mechanism , the experiment results are analyzed. Results show that: the flat-flat electrode DBD displays filamentary discharge mode. Meanwhile, the line-tube electrode DBCD is relatively homogeneous and stable because of the corona discharge.**

***Index Terms*—dielectric barrier discharge; flat -flat electrode; line-tube electrode ; corona; under normal atmospheric pressure**

I. INTRODUCTION

The low temperature plasma and its application produced by glow discharge under normal atmospheric pressure (here, the "glow discharge under normal atmospheric pressure" is referred simply to as "APGD") are research hotpots attended extensively by the scholars both abroad and at home. It overcomes the shortcoming of the vacuum chamber. And the non-equilibrium homogeneous plasma produced by APGD has broad application prospects on material surface modification, thin film deposition, etching, medical appliance sterilization, fiber modified, etc[1]. Research on the dielectric barrier discharge(referred simply to as "DBD") adopted different electrode configurations is a quite effective method to realize APGD, which devises specific poweres and mediums with different frequencies to construct APGD environment in some specific gases and gas mixtures. This method gets a long general persistent research on a global scale.

Dielectric barrier discharge is a kind of gas discharge which inserts a thin dielectric into discharge space. It also is called silent discharge because the medium inserted between the electrodes which can prevent local sparks or arc discharge in discharge space, and also can achieve other stable discharges under normal atmospheric pressure[2]. Because DBD can work in great pressure and frequency range and realize APGD under specified conditions, many people engaged in this field research since 1987. Corona discharge is one of the effective means among APGD to gain the low temperature plasma under normal atmospheric pressure, which adopts inhomogeneous electric field, can easy generate discharge, but has the disadvantages of smaller discharge area and lower power density.

Therefore, if corona discharge and dielectric barrier discharge can be combined with a suitable ways, we should gain a better glow discharge under normal atmospheric pressure. Here, we call this kind of discharge as dielectric barrier and corona discharge (referred simply to as "DBCD")

II. TYPICAL CONFIGURATIONS OF DBD AND DBCD

Figure 1-a is common configuration of flat - flat electrode configuration in DBD, which adds a layer of insulating medium between the two electrodes. This configuration adopts ac power supply with voltage of 20kV[3].

Research shows that this kind of electrode configuration will occur many short current filaments when appears electric breakdown with short cycle. This phenomenon is called u-EDM. Depend on the type of gases, medium surface properties and operation conditions, dielectric barrier discharge can get several known different kinds of discharging mode, including filamentous discharge, pattern discharge and complete barrier discharge.

Figure 1-b is a typical electrode configuration of DBCD adopted line - tube configuration electrode, which fine lines electrode is a high voltage electrode of diameter 0.2-1mm and fixed in the center of barrier medium glass tube. The outside diameter of the tube is 30～40mm. The thickness of medium is 1～5mm. Barrier medium can adopt tube of glass, resin, polytetrafluoroethylene, etc[4].

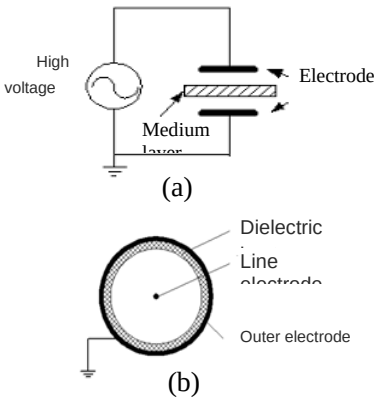


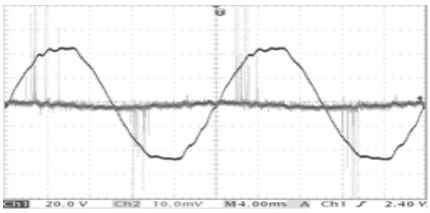
Fig.1 Typical electrode configuration of dielectric barrier discharge and barrier corona discharge

III. CHARACTERISTICS AND MECHANISM OF DISCHARGE

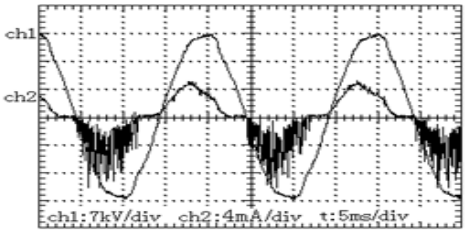
Fig.2-a is the variation waveform curve of current and voltage of flat-flat electrode DBD and fig.2-b is the waveform curve of line-tube electrode DBCD. From Fig.2-a, it shows that many micro-electrical discharge pulses compose the current in every half cycle of voltage. The phenomenon of u-EDM begins initial discharge voltage and ends up with the maximum voltage. And the current waveform almost distributes symmetrically in voltage positive and negative half cycle separately.

From Fig.2-a, the line-tube electrode DBCD current waveform appears as distinct polarity effect. In the positive half cycle of voltage, it appears as continuous discharge current mode. But in negative half cycle of voltage, it appears as Trichel pulse mode.

This kind of phenomenon can be explained as follows. The diameter of outer barrel discharge cathode is larger than the inner line cathode's, and this factor leads to the volume of outer cathode's ionization region of cathode sheath layer is larger than the inner line cathode's. so in whole voltage cycle, the current of upper half cycle appears larger than the current of under half cycle, even though the electric field of the inner line cathode is more power than the outer.



(a) Voltage-current oscillogram of DBD

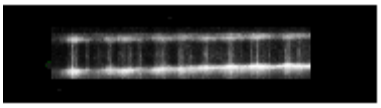


(b)

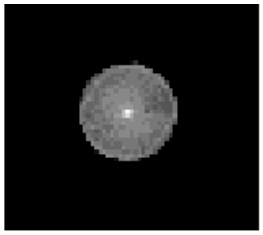
Fig.2 Voltage-current oscillogram

Fig.3 is illuminated diagrams of two different electrode configurations shoot by camera. Fig.3-a is illuminated diagram of flat-flat electrode DBD. It shows that many filamentous luminances homogeneously distribute in the discharge space. Fig.3-a is illuminated diagram of line-tube electrode DBCD. It shows the air gap of the tube appears stable glow when the 14kV~20kV of voltage is brought on the electrode.

The DBD of flat-flat electrode can be explained by Stream Theory as follows. Because its discharge space is approximately homogeneous electronic space, and when the ionization coefficient α of stronger area of discharge electronic space is up to an enough value, the α of most of area in the gap will also up to a corresponding value. Then the initial electron avalanche will develop rapidly in high electric area, and will soon form streams in the gap. Moreover, the space charges are accumulated on dielectric because of existence of barrier dielectric. The accumulated charges form an additional electric field in opposition to the applied electric field, and it neutralizes a part of effect of the applied electric field. With the accumulation of charges on the dielectric, the effect of the additional electric field is escalating and the decrease in whole electrical field strength is increasing in gas gap. When field strength in gas gap decreases to less than the gas gap breaking down field strength of gas, the discharge will interrupt. And because DBD adopts ac power, the discharge still appears in current of under half cycle though the discharge is interrupted[5]. Therefore, the DBD of flat-flat electrode is approximate match the discharge of DBD in homogeneous electric field. And its current waveforms approximately express many pulse modes in voltage positive and negative half cycle symmetrically and alternately.



(a) Discharge of flat-flat electrode DBD



(b) Discharge of line-tube electrode DBCD

Fig.3 Discharge of different electrodes

It is necessary to limit electron avalanches increase if want to get a stable diffused mode discharge in DBD. In a high atmospheric pressure, electronic collisions are inevitable. To limit ionization coefficient α of collision is a feasible measure to limit the development of electron avalanches. And because α increases with field strength of air gap, it is necessary to bring down the breaking down field strength of gas, that is those electrons must be gain only in low AC electric field[6].Normally, the breaking down field strength of air under normal atmospheric pressure is very high, and the average strength is about 30kV/cm. So it is very difficult to limit the development of electron avalanche in this environment. So some other measures should be taken to gain secondary electrons or form preionization, and gain a homogeneous diffused mode discharge. The line-tube electrode configuration of DBCD adopts the preionization of line electrode, and realizes the diffused mode discharge preferably. The electric fields of line-tube electrode configuration are very inhomogeneous, especially those

electric fields around the line electrode. Therefore, when voltage gets to a certain extent and before gas gap being broken down, a corona luminous layer will appear nearby the center of these line electrodes. With the increase of voltage, the homogeneous corona layer expands unceasingly, and forms a comparatively stable diffusion mode discharge eventually.

IV. SUMMARY

Both of DBD and DBCD are available to realize APGD. the current of flat-flat electrode DBD is composed by many micro-electrical discharge pulses in every half cycle of voltage. The current waveform of flat-flat electrode DBD almost distributes symmetrically in voltage positive and negative half cycle separately. Comparatively, The current waveform of line-tube electrode DBCD shows polarity effect distinctly. In the positive half cycle of voltage, it appears as continuous discharge current mode. And in negative half cycle of voltage, it appears as Trichel pulse mode. Through comparing, the discharge of configuration of DBCD is more stable than DBD'.

REFERENCE

[1] Wang Yan, Zhao Yanhui, Bai Xiliang. Plasma of DBD application technology development[J]. Nature magazine, 2002, 24(5): 277-282.

[2] Gherardi N, Massines F. Mechanisms controlling the transition from glow silent discharge to streamer discharge in nitrogen[J].IEEE Trans on Plasma Science, 2001, 29(3): 536-544.

[3] Li Xue-cheng, Ying Zeng-qian, etal. Study on the Characteristics of Dielectric Barrier Discharge at Atmospheric Pressure [J]. Journal of Hebei University (Natural Science Edition) 2002, 22(1): 16-18.

[4] Liao Min-fu. Duan Xiong-ying, Li Jin. Research on Modality of Pulsed Corana Plasma in a Coaxial-Electrode System[J]. Journal of Electrical technology, 2002, 17(4): 26-30.

[5] Sun Yanzhou, Qiu Yuchang, Yuan Xing-cheng, Study of SO2 Removal Using Dielectric Barrier Corona Discharge [J]. High Voltage Apparatus, 2004, 40(4): 253-254.

[6] Sun Yanzhou, Qiu Yuchang, Yu Fashan, etal. Application of DBD and DBCD in SO2 Removal[J]. Plasma Science and Technology, 2004, 6(6):2589-2592.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15, August 2010, pp. 495-497

Advance improvement on the simple authentication key agreement protocol

Chao Deng[1], Shaoyi Deng[2]
[1] Institute of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, China
Email: super@hpu.edu.cn
[2] Institute of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, China
Email: shaoyideng@163.com

***Abstract*—Yang et al. proposed a cryptanalysis of simple authenticated key agreement protocols that they broke the SAKA-related scheme in 2004. We improved a new scheme which it based on the Tseng's scheme, and it would withstand previous attack such as the modification attack and the dictionary attack.**

***Index Terms*—Key agreement, Middleperson attack, Key exchange, Authenticated**

I. Introduction

Diffie and Hellman proposed key agreement protocol in which two parties can establish a secret session key over an insecure channel. In 1999, Seo and Sweeney proposed a simple authenticated key agreement algorithm (SAKA) that is based on a pre-share password method when two parties want to connect each others. So far so many authenticated key agreement schemes had proposed in [1-4]. Tseng first showed SAKA existing weakness problem and gave an improvement. Ku and Wang pointed out a vulnerable and presented two attacks in Tseng's scheme.

In this paper, we proposed an advanced scheme which it would defense middleperson attack and reduce computation performance. In section 2, we will review of SAKA algorithm. In section 3, we discussed some attack methods to SAKA and proposed an advance version, then give a security analysis for our scheme. The conclusion will discuss in final section.

II. Review Of Saka Algorithm

The SAKA scheme modifies from Diffie-Hellman key agreement scheme. For the remainder of this paper, we assume that Alice and Bob have agreed on a session key for secure communications. The public values are p and g, where p is a large prime and g is a primitive element in $GF(p)$. The steps in the SAKA are described as follows:

Step 1. Before the protocol, Alice and Bob agree to pre-share $a$common password $Q$where $gcd(Q, p-1) = 1$.

Step 2. Alice randomly selects her private key a where $1 \leqslant a \leqslant p-2$, and computes $X_1 \equiv g^{aQ} \pmod p$.Then, Alice sends X_1 to Bob.

Step 3. Bob randomly selects his private key b where $1 \leqslant b \leqslant p-2$ and computes $Y_1 \equiv g^{bQ} \pmod p$.Then, Bob sends Y_1 to Alice.

Step 4. When Alice received Y_1, she can computes $Y \equiv Y_1^{Q^{-1}} \pmod p$ and $Key_1 \equiv Y^a \pmod p$.

Step 5. When Bob received X_1, he can computes $X \equiv X_1^{Q^{-1}} \pmod p$ and $Key_2 \equiv X^b \pmod p$. Therefore, Alice and Bob obtain a common secret session key $g^{ab} \pmod p$.

Step 6. To verify the validity of the session key and authenticate one another, Alice sends Key_1^Q to Bob and Bob sends Key_2^Q to Alice. Therefore, Alice and Bob can check the validity of the session key by applying Q^{-1} and compare this with his/her own session key.

III. Some Attacks

A. Tseng's methods

In the Seo-Sweeney protocol, although an attacker (Eve) cannot impersonate Bob to compute a common session key shared with Alice ,the verifying process of the session key in their protocol has the following weakness; Eve may re-send it to Alice after receiving the message $Key_1^Q \pmod p$ sent by Alice, Alice then computes the key $(Key_1^Q)^{Q^{-1}} \pmod p$ where $Q \cdot Q^{-1} \equiv 1 \pmod p$. Therefore, although Alice obtains a wrong session key and Eve cannot compute the same wrong session key, Alice will still believe it .that is, verification of the session key cannot be achieved using this protocol.

B. Tseng's improvement

To overcome the above weakness, the verification of session key is modified as follows:

(i) Alice sends Y to Bob.

(ii) Bob sends X to Alice.

(iii)Alice and Bob check whether and hold or not, respectively.

With the modified protocol, when Eve receives from Alice, Eve has compute and then sends it to Alice in the verification step of the session key. If Eve wants to obtain a and from or , then she must face discrete logarithm problem. Therefore Eve can not compute the password. Moreover, in the modified protocol, and are computed in the session key establishment phase. Compared to the original protocol, the modified protocol reduces the computational time by two exponentiations in the key verification phase.

C. Ku-Wang method

Key validation:
(v.1) Alice sends Y to Bob.
(v.2) Bob sends X to Alice.
(v.3) Alice and Bob checks whether $X \equiv g^a \pmod p$ and $Y \equiv g^b \pmod p$ and holds or not.
Modification attack [5]: Upon seeing X_1 sent by Alice, Eve can replace it with any number $\in [1, p-1]$ say X_1'. Bob sends Y_1 to Alice, and then Alice sends the corresponding response *Y* to Bob in step (v.1). In step (v.2), Bob will send ,which equals $(X_1')^{Q^{-1}} \pmod p$, to Alice. Because $X' \not\equiv g^a \pmod p$, Alice will not believe Key_1. However, since $Y \equiv g^b \pmod p$ holds, Bob will believe the wrong session key Key_2' which equals ${X_1'}^{Q^{-1}} \pmod p$. .Although Eve cannot compute Key_2', she can still fool Bob into believing this wrong session key . Note that if *step (v.1)* and *step (v.2)* are exchanged, the protocol is still vulnerable to the modification attack in the opposite direction, i.e. it is Alice rather than Bob who will be fooled into believing a wrong session key.

D. Ku-Wang improvement

Enhanced key validation steps:
(v.1) Alice computes $Y_2 \equiv (Key_1)^Q \equiv g^{abQ} \pmod p$ and and then sends Y_2 to Bob.
(v.2) Bob checks whether $Key_2 \equiv Y_2^{Q^{-1}} \pmod p$ holds or not. If it holds, Bob believes that he has obtained the correct X_1 and Alice has obtained the correct Y_1, i.e. Bob is convinced that key_2 is validated, and hen sends X to Alice.
(v.3) Alice checks whether $X \equiv g^a \pmod p$ holds or not. If it holds, Alice believes that she has obtained the correct Y, and Bob has obtained the correct, i.e. Alice is convinced that Key_1 is validated.
The weakness of the Seo-Sweeney protocol is due to the same values of the two key validation messages [5]. One problem within Tseng's modified protocol is that the values of the two key validation messages will be the same once $Y_1 = X_1$. Another problem within Tseng's modified protocol is that Bob cannot judge the correctness of X_1 from the received Y. In the enhanced key validation steps, the first key validation message is directly inherited from the Seo-Sweeney protocol while the second key validation message is adopted from Tseng's modified protocol. The use of asymmetric messages in the enhanced key validation steps is one of the methods of resisting the attack of backward replay without modification. In addition, the first key validation message, Y_2 can alternatively be generated from $Y_2 \equiv (Y_1)^a \pmod p$ and verified by checking whether $Y_2 \equiv (X_1)^b \pmod p$. This alternative is useful if the protocol is implemented in hardware. As the generation (or verification) of Y_2, can be performed in parallel with the session key generation, the computation delay can be reduced.

IV. THE PROPOSED SCHEME

We follow the original of Tseng SAKA which definition same notations. Alice and Bob want to communicate in a insecure channel. Then the protocol is described in below.
Here the $\oplus$ symbol represents the bitwise exclusive-or operation.
Step 1. Alice randomly chooses her secret key a and computes $Q_0 = g$, $Q_{i+1} \equiv Q_i^Q \pmod p$, $i \in [0,4]$, $X_1 = Q_1 \oplus (g^{aQ} \pmod p)$, then sends X_1 to Bob.
Step 2. Bob also chooses his secret key b and computes $Q_0 = g$, $Q_{i+1} \equiv Q_i^Q \pmod p$, $i \in [0,4]$, $Y_1 = Q_2 \oplus (g^{bQ} \pmod p)$, then sends the Y_1 to Alice.
Step 3. When Y_1 is received, Alice computes $Y = (Q_2 \oplus Y_1)^{Q^{-1}} \pmod p$
and $Key_1 \equiv Y^a \equiv g^{ab} \pmod p$.
Step 4. When X_1 is received, Bob computes $X = (Q_1 \oplus X_1)^{Q^{-1}} \pmod p$
and $Key_2 \equiv X^b \equiv g^{ab} \pmod p$.
Key verification phase
(v.1) Alice computes $k_1 = Q_3 \oplus ((Key_1)^Q \equiv g^{abQ} \pmod p)$.
Then, Alice sends K_1 to Bob.
(v.2) When K_1 is received, Bob checks whether $Key_2 \equiv (K_1 \oplus Q_3)^{Q^{-1}} \pmod p$. If it holds, Bob believes that he has obtained the correct X_1 and Alice has obtained the correct Y_1 . Then, he sends $X' = (X \oplus Q_4)$to Alice.
(v.3) When X' is received, Alice checks whether $(X' \oplus Q_4) \equiv g^a \pmod p$. If it holds, Alice believes that she has obtained the correct Y_1 and Bob has obtained the correct X_1 .The protocol procedure is described in figure 1.

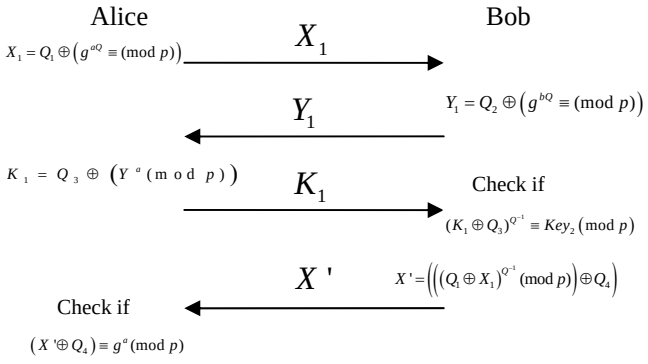


Figure 1. Advance Tseng SAKA protocol.

Security Analysis:

Scenario 1**:** If adversary Eve wants to impersonate Alice and Bob into believing a wrong session key in the scheme. Eve pre-computes a value z where $z \pmod{(p-1)}$ and its inversion element $z^{-1} \pmod{(p-1)}$. Fortunately, Eve does not know the pre-share password Q and she could not compute and replace $X_1 = Q_1 \oplus (g^{aQ} \pmod{p})$ with $X_1' = Q_1 \oplus (g^{aQz} \pmod{p})$, Eve therefore could not obtain Q_1 or $(g^{aQ} \pmod{p})$. Thus, our scheme can withstand the modification attack[7].

Scenario 2: Assume Eve reconstruct the values Q and Q^{-1}shared between Alice and Bob. Eve intercepted $X_1 = Q_1 \oplus (g^{aQ} \pmod{p})$ and sent by Alice in Step 1. Eve could not compute Q_2 and masquerade Bob by $Y_1 = Q_2 \oplus (g^z \pmod{p})$. The pre-share password Q is sequent changed for each sessions, it therefore does not a fixed value on connection. Our scheme could withstand the dictionary attack[8-11]. In the mean time, it also could defense replay and backward replay attack.

V. Conclusions

As we have proved, our scheme can resist all the attacks mentioned in [5-8]. Eve cannot fabricate which protected by a random number sequence since and are random chosen. Our scheme is secure for previous version.

Acknowledgment

I am grateful to Dr. Chenglian Liu (Information Security Group, Royal Holloway, University of London). This research was partially supported by the Doctor Fund of Henan Polytechnic University.

References

[1] D. Seo and P. Sweeney, “Simple authenticated key agreement algorithm”, Electron. Letters, 1999, vol. 35, no. 13, pp. 1073–1074.

[2] Y. M. Tseng, “Weakness in simple authenticated key agreement protocol”, Electron. Letters, 2000, vol. 36, no. 1, pp. 48–49.

[3] W. C. Ku and S. D. Wang, “Cryptanalysis of modified authenticated key agreement protocol”, Electron. Lette., 2000, vol. 36, no. 21, pp. 1770–1771.

[4] Iuon-Chang Lin, Chin-Chen Chang, Min-Shiang Hwang, Security Enhancement for the Simple Authentication Key Agreement Algorithm, COMPSAC 2000, pp. 113-115.

[5] Chou-Chen Yang, Ting-Yi Chang, Min-Shiang Hwang, Cryptanalysis of Simple Authenticated Key Agreement Protocols, IEICE TRANS, 2004, VOL. E87-A, NO.8.

[6] I. E. Liao, C. C. Lee, M. S. Hwang, A password authentication scheme over insecuer networks, Journal of Computer and System Sciences, 2006, Vol. 72, No.4, pp. 727-740.

[7] C. S. Bindu, P. C. S. Reddy, B. Satyanarayana, Improved remote user authentication scheme preserving user anonymity, International Journal of Computer Science and Network Security, 2008, Vol. 8, No. 3, pp.62-65.

[8] W. S. Juang, S. T. Chen, H. T. Liaw, Robust and efficient password-authenticated key agreement using smart cards, IEEE Transactions on Industrial Electronics, 2008, Vol. 55, No. 6, pp.2551-2556.

[9] T. H. Chen, W. B. Lee, A new method for using hash functions to solve remote user authentication, Computers & Electrical Engineering, 2008, Vol. 34, No. 1, pp.53-62.

[10] E. J. Yoon, K. Y. Yoo, Improving the novel three-party encrypted key exchange protocol, 2008, Computer Standards & Interfaces, Vol. 30, No. 5, pp. 309-314.

[11] Y. Liao, S. S. Wang, A secure dynamic ID based remote user authentication scheme for multi-server environment, Computer Standards & Interfaces, 2009, Vol. 31, No. 1, pp. 24-29.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15, August 2010, pp. 498-501

Construction of Basis Algebra in L-fuzzy Rough Sets

Zhengjiang Wu[1,2]
[1] School of Information Science and Technology, Southwest Jiaotong University, Chengdu, China
[2] School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, China
Email: jiang2021987@163.com

***Abstract*—The approximation operator play a vital role in rough set theory. About the approximation operator, usually, we discuss three fundamental elements, including the binary relation in the universe, the basis algebra and the property of approximation operators. In the constructive approach of approximation operators, the properties of the approximation operators depend on the basis algebra and the binary relation. In this paper, we introduce the basis algebra construction approach, named as the basis algebra choosing approach, the basis algebra is built with properties of approximation operators and other certain binary relations.**

***Index Terms*—L-fuzzy rough sets, approximation operators, basis algebra, binary relation, residuated lattice**

I. INTRODUCTION

Modeling uncertain information, including fuzziness, randomicity, incompleteness and uncomparativity, is one of the main research topics in knowledge representations. Most existing approaches are based on the extensions of classical set theory such as fuzzy set theory and rough set theory.

The concept of rough set [1] was originally proposed by Pawlak as a formal tool for modeling and processing the incomplete information in information systems. In the rough set theory, the core idea of rough set is to approximate the knowledge with uncertainty by using two "certain" definitions. The two "certain" definitions are named as the lower and the upper approximation sets.

The lower and the upper approximation operators are the two unary mappings in the universe. To real applications, we select a certain rough set model to approximate the uncertain information. For instance, if there is the fuzzy information, we take the fuzzy rough set into account.

Many new types of rough sets theories had been put forward, such as fuzzy rough sets, L-fuzzy rough sets and general rough sets. In the axiomatic approach, such as ref. [1,2], the approximation operators had been defined as two unary operators which satisfy some axioms in the universe. In the axiomatic approach of approximation operators it is vital to find the binary relation to construct the approximation space.[3]

In the constructive approach of rough approximation operators, many work focused on the properties of approximation operators in different binary relations. By compared with the rough sets[1], fuzzy rough sets[4-6] and L-fuzzy rough sets[3,7,8], it can be found that the properties of approximation operators based on different basis algebras (such as boolean algebra, [0,1] and lattice) are different, even in the same binary relations and the forms of approximation operators. In this paper, we discuss the L-fuzzy rough sets based on residuated lattice, IMTL algebra and Boolean algebra. The differences of these rough sets will be shown under the influence of basis algebra.

II. PRELIMINARIES

Definition 1[9]. By a residuated lattice, we mean an algebra $(L,\vee,\wedge,\otimes,\to,0,1)$ such that

(1) $(L,\vee,\wedge,0,1)$ is a bound lattice with the top element 1 and the bottom element 0.

(2) $\otimes: L\times L\to L$ is a binary operator and satisfies for $\forall a,b,c\in L$,

$$a\otimes b=b\otimes a,\ a\otimes(b\otimes c)=(a\otimes b)\otimes c,$$

$$1\otimes a=a,\ a\le b\Rightarrow a\otimes c\le b\otimes c.$$

(3) $\to: L\times L\to L$ is a residuum of $\otimes$, i.e. $\to$ satisfies for all $a,b,c\in L$,

$$a\otimes b\le c\Leftrightarrow a\le b\to c.$$

A residuated lattice $(L,\vee,\wedge,\otimes,\to,0,1)$ is called complete iff the underlying lattice $(L,\vee,\wedge,0,1)$ is complete. Given a residuated lattice L, we define the precomplement operator $\sim: L\to L$ as follows: $\forall a\in L$, $\sim a=a\to 0$.

Theorem 1[9] Suppose $(L,\vee,\wedge,\otimes,\to,0,1)$ is a residuated lattice, and $\sim$ is the precomplement operator on L. Then $\forall a,b,c\in L$,

(1) $a\otimes b\le a$, $a\to b\ge b$.

(2) $a\to(b\to c)=(a\otimes b)\to c$,

$$a\to(b\to c)=b\to(a\to c).$$

(3) If $a\le b$ and $c\le d$, then $d\to a\le c\to b$.

(4) $a\le b\Rightarrow a\to b=1$.

(5) $a\le\sim\sim a$, $\sim\sim\sim a=\sim a$.

(6) $a\to\sim b=b\to\sim a=\sim(a\otimes b)$.

(7) $a\to b\le\sim(a\otimes\sim b)$, $a\otimes b\le\sim(a\to\sim b)$.

(8) If L is a complete lattice, then

$$\left(\bigvee_{i\in I}a_i\right)\otimes b=\bigvee_{i\in I}(a_i\otimes b),\ a\to\left(\bigwedge_{i\in I}b_i\right)=\bigwedge_{i\in I}(a\to b_i),$$

$$\left(\bigvee_{i\in I} a_i\right)\to b=\bigwedge_{i\in I}(a_i\to b),$$

$$a\to\left(\bigvee_{i\in I} b_i\right)\geq\bigwedge_{i\in I}(a\to b_i)$$

$$\left(\bigwedge_{i\in I} a_i\right)\to b\geq\bigvee_{i\in I}(a_i\to b).$$

Definition 2[9] The residuated lattice $(L,\vee,\wedge,\otimes,\to,0,1)$ is called an MTL-algebra iff it satisfies the following prelinery condition, for $\forall a,b\in L$, $(a\to b)\vee(b\to a)=1$.

Definition 3[11] The MTL-algebra L_{MTL} is called an IMTL-algebra iff it satisfies the following condition: for $\forall a\in L_{MTL}$, $\sim\sim a=a$.

Theorem 3[11] Suppose $(L,\vee,\wedge,\otimes,\to,0,1)$ is an IMTL algebra, and $\sim$ is the precomplement operator on L. Then for $\forall a,b\in L$,

(1) $a\to b=b\to a$.

(2) $a\to b=\sim(a\otimes\sim b)$, $a\otimes b=\sim(a\to\sim b)$.

Definition 5[7] Suppose U is a non-empty finite universe and R_L is an L-fuzzy binary relation on U based on residuated lattice. (U,R_L) is called an L-fuzzy approximation space based on the residuated lattice. For any set $A\in F_L(U)$, the lower and upper approximation $\underline{R}_L(A)$ and $\overline{R}_L(A)$ of A with respect to the approximation space (U,R_L) are L-fuzzy sets on U whose membership functions are respectively defined by

$$\underline{R}_L(A)(x)=\inf_{y\in U}\left(R_L(x,y)\to A(y)\right)$$

$$\overline{R}_L(A)(x)=\sup_{y\in U}\left(R_L(x,y)\otimes A(y)\right)$$

The pair $\left(\underline{R}_L(A),\overline{R}_L(A)\right)$ is referred to as an L-fuzzy rough set. $\underline{R}_L,\overline{R}_L:F_L(U)\to F_L(U)$ are referred to as lower and upper L-fuzzy approximation operators.

Radzikowska et al. have proved some properties of L-fuzzy rough approximation operations based on residuated lattice in [5]. These properties of L-fuzzy rough sets have been collected in Theorem 3, the others as the supplement of [5] are in Theorem 4[3].

Theorem 4[5] Let $(L,\vee,\wedge,\otimes,\to,0,1)$ be a complete residuated lattice, (U,R_L) is L-fuzzy approximation space. Then for $\forall A,B\in F_L(U)$,

(1) $\underline{R}_L(U)=U$, $\overline{R}_L(\varnothing)=\varnothing$.

(2) If $A\subseteq B$, then $\underline{R}_L(A)\subseteq\underline{R}_L(B)$, $\overline{R}_L(A)\subseteq\overline{R}_L(B)$.

(3) $\underline{R}_L(A\cap B)=\underline{R}_L(A)\cap\underline{R}_L(B)$,

$\overline{R}_L(A\cup B)=\overline{R}_L(A)\cup\overline{R}_L(B)$.

(4) $\underline{R}_L(A\cup B)\supseteq\underline{R}_L(A)\cup\underline{R}_L(B)$,

$\overline{R}_L(A\cap B)\subseteq\overline{R}_L(A)\cap\overline{R}_L(B)$.

Theorem 5[3,12] Let $(L,\vee,\wedge,\otimes,\to,0,1)$ be a complete residuated lattice, (U,R_L) is L-fuzzy approxim-ation space. Then for $\forall A,B\in F_L(U)$,

(5) $\underline{R}_L(A)\subseteq\sim\overline{R}_L(\sim A)$, $\overline{R}_L(A)\subseteq\sim\underline{R}_L(\sim A)$.

(6) $\sim\overline{R}_L(A)=\underline{R}_L(\sim A)=\sim\overline{R}_L(\sim\sim A)=\sim\sim\underline{R}_L(\sim A)$.

Theorem 6[3,12] Let $(L,\vee,\wedge,\otimes,\to,0,1)$ be a complete IMTL-algebra, (U,R_L) is L-fuzzy approximation space. Then for $\forall A,B\in F_L(U)$,

(1) $\underline{R}_L(A)=\sim\overline{R}_L(\sim A)$, $\overline{R}_L(A)=\sim\underline{R}_L(\sim A)$.

In Theorem 6, we just list the properties which are different from the the residuated lattice-fuzzy rough set.

The major difference between the L-fuzzy rough sets based on residuated lattice and IMTL-algebra is the duality of the lower and the upper approximation operators. The duality is important for the axiomatic approach. Usually, in the axiomatic approach, the relation is defined by the upper approximation operator. Through the duality (or the semi-duality) of the approximation operator, the relation is loaded in the lower approximation operator. Without the duality, the process add some other conditions to load the relation by the lower approximation operator, such as the condition "$\{\sim a\mid a\in L\}=L$".

III. The Basis Algebra in L-Fuzzy Rough Sets

In an IMTL algebra-fuzzy rough set, we prove the lower and the upper approximation operators are dual, but it does not hold in a residuated lattice-fuzzy rough set. One question arises: Which property in the IMTL-algebra is virtual for the duality of approximation operators? Another question is: If we want the approximation operators to satisfy some properties, such as the duality in the axiomatic approach, which properties should the corresponding basis algebra satisfy?

Definition 6 *Let* $(L,\vee,\wedge,\otimes,\to,0,1)$ be an algebra, where $(L,\vee,\wedge,0,1)$ is a complete lattice. $\otimes,\to$: $L\times L\to L$ are the binary operators. If L satisfies:

(1) $\sim: L\to L$ is a unary operator. For all $a\in L$, $\sim a=a\to 0$;

(2) $\forall\alpha_i,\beta\in L$, $\bigvee_{i\in I}\alpha_i\otimes\beta=\bigvee_{i\in I}(\alpha_i\otimes\beta)$,

(3) $\forall\alpha_i,\beta\in L$, $\bigvee_{i\in I}\alpha_i\to\beta=\bigwedge_{i\in I}(\alpha_i\to\beta)$

Then the L is named as D-algebra.

By selecting the certain L-fuzzy sets, such as 1_{y_0} and $B_{\{1,y_0\}}$, we can prove all of the following theorems, where for all $a\in L$, $x\in U$,

$$\alpha_{y_0}(x)=\begin{cases}\alpha & x=y_0\\ 0 & x\neq y_0\end{cases},\ B_{\alpha,y_0}(x)=\begin{cases}\alpha & x=y_0\\ 1 & x\neq y_0\end{cases}.$$

Theorem 7 Let L be a D-algebra and U be a non-empty universe. If for all L-fuzzy approximation spaces

(*U*,*R*), the upper approximation operators $\bar{R}_{\mathcal{L}}$ satisfy $\bar{R}_{\mathcal{L}}(\varnothing)=\varnothing$, then

(r1) for all $a\in L$, $\alpha\otimes 0=0$.

Theorem 8 Let *L* be a D-algebra and *U* be a non-empty universe. If for all L-fuzzy approximation spaces (*U*,*R*), the upper approximation operators $\underline{R}_L$ satisfy $\underline{R}_L(U)=U$, then

(r1') for all $a\in L$, $\alpha\to 1=1$.

Theorem 9 Let *L* be a D-algebra and *U* be a non-empty universe. If for all L-fuzzy approximation spaces (*U*,*R*), the upper approximation operators $\bar{R}_L$ satisfy for all $A,B\in F_L(U)$,

$$A\subseteq B\Rightarrow \bar{R}_L(A)\subseteq \bar{R}_L(B),$$

then

(r2) for all $a,b,c\in L$, $a\le b\Rightarrow c\otimes a\le c\otimes b$.

Theorem 10 Let *L* be a D-algebra which satisfies (r1) and *U* be a non-empty universe. If (*U*,*R*) is an L-fuzzy approximation space, the L-fuzzy binary relation *R* is reflexive, and the upper approximation operators $\bar{R}_L$ satisfy for all $a\in L$, $\bar{R}_L(\hat{a})=\hat{a}$, then

(r3) for all $a\in L$, $1\otimes a=a$.

Theorem 11 Let *L* be a D-algebra which satisfies (r1) and (r3), and *U* be a non-empty universe. If (*U*,*R*) is an L-fuzzy approximation space, the L-fuzzy binary relation *R* is reflexive, and the upper approximation operators $\bar{R}_L$ satisfy for all $A,B\in F_L(U)$,

$$\bar{R}_L(A\otimes B)=\bar{R}_L(B\otimes A),$$

then

(r4) for all $a,b\in L$, $a\otimes b=b\otimes a$.

Theorem 12 Let *L* be a D-algebra which satisfies (r1) and (r3), and *U* be a non-empty universe. If (*U*,*R*) is an L-fuzzy approximation space, the L-fuzzy binary relation *R* is reflexive, and the upper approximation operators $\bar{R}_L$ satisfy for all $A,B,C\in F_L(U)$, $\bar{R}_L((A\otimes B)\otimes C)=\bar{R}_L(A\otimes(B\otimes C))$ then

(r5) for all $a,b,c\in L$, $a\otimes(b\otimes c)=(a\otimes b)\otimes c$

Theorem 13 Let *L* be a D-algebra which satisfies (r1) and (r1'), and *U* be a non-empty universe. If for all L-fuzzy approximation spaces (*U*,*R*), the approximation operators $\bar{R}_L,\underline{R}_L$ satisfy for all $A,B\in F_L(U)$,

$$\bar{R}_L(A)\subseteq B\ \Leftrightarrow A\subseteq \underline{R}_L(B),$$

then

(r6) for all $a,b,c\in L$, $a\otimes b\le c\Leftrightarrow a\le b\to c$.

By the definition of the residuated lattice, if the D-algebra satisfies (r1)-(r6) and (r1'), then the D-algebra is a residuated lattice. Following these steps, we can construct the MTL-algebra and the IMTL-algebra.

Theorem 14 Let *L* be a D-algebra which satisfies (r1) and (r3), and *U* be a non-empty universe. If (*U*,*R*) is an L-fuzzy approximation space, the L-fuzzy binary relation *R* is reflexive, and the upper approximation operators $\bar{R}_L$ satisfy for all $A,B\in F_L(U)$,

$$\bar{R}_L\left(\left(B_{a,y_0}\to b_{y_0}\right)\cup\left(B_{b,y_0}\to a_{y_0}\right)\right)=U$$

then

(r7) for all $a,b,c\in L$, $(a\to b)\vee(b\to a)=1$.

Theorem 15 Let *L* be a D-algebra which satisfies (r1) and (r1'), and *U* be a non-empty universe. If for all L-fuzzy approximation spaces (*U*,*R*), the approximation operators $\bar{R}_L,\underline{R}_L$ satisfy for all $A\in F_L(U)$,

$$\bar{R}_L(A)=\sim\underline{R}_L(\sim A)$$

then

(r8) for all $a\in L$, $\sim\sim a=a$.

By the definition of the IMTL-algebra, if the D-algebra satisfies

(r1)-(r8) and (r1'), then these D-algebra is an IMTL-algebra.

In this section, as the sufficient conditions, the properties of the approximation operators are the special ones.

IV. Conclusions

The choosing process of the basis algebra is similar to the axiomatic process of the approximation operators. The axiomatic approach is the process that finds the binary relation based on the axiom set and the basis algebra. The basis algebra choosing approach is the process that constructs the basis algebra with the properties of the approximation operator and the binary relation. As the axiomatic approach[3,6,10], in the basis algebra choosing approach, the condition set of the basis algebra choosing approach is not unique, such as the fuzzy rough set (r1)-(r8), (r1') aren't our only choice. We can find other conditions to replace them. For example, the condition "(r3') For every binary relation *R*, $\bar{R}_L(U)=U$" can replace (r3).

Acknowledgment

This work was supported by the National Natural Science Foundation of China (Grant No. 60873108), Henan Province Key Scientific and Technological Project (082102210079).

References

[1] Z. Pawlak: Rough sets, *International Journal of Computer and Information Science*, 1982, 341-356.

[2] L.A. Zadeh, Fuzzy sets, *Information Control*. **8**, 1965, 338-353.
[3] Zh. J. Wu, W. F. Du, K. Y. Qin: The properties of L-fuzzy rough set based on complete residuated lattice. *2008 International Symposium on Information Science and Engieering*, Shanghai, China, 2008, 617-621.
[4] D. Dubois, H. Prade: Rough fuzzy sets and fuzzy rough sets, *Internat. J. General Systems*. **17**(2-3), 1990, 191-209.
[5] D. Dubois, H. Prade: *Putting fuzzy sets and rough sets together, In Intelligent Decision Support, (Edited by R. Slowinski)*, Kluwer Academic, Dordrecht, 1992, 203-232.
[6] N. N. Morsi, M. M. Yakout: Axiomatics for fuzzy rough sets, *Fuzzy Sets and Systems*. **100**, 1998, 327-342.
[7] A. M. Radzikowska, E. E. Kerre: A comparative study of fuzzy rough sets, *Fuzzy Sets and System*, **126**, 2002, 137-155.
[8] A. M. Radzikowska, E. E. Kerre: An algebraic characterization of fuzzy rough sets, *Fuzzy Systems, 2004 IEEE International Conference*, 109-114.
[9] J. Pavelka: On fuzzy logic I: Many-valued rules of inference, II: Enriched residuated lattices and semantics of propositional calculi, III: Semantical completeness of some many-valued propositional calculi. *Zeitschr. F. Math. Logik und Grundlagend. Math.* **25**, 1979, 45-52, 119-134, 447-464.
[10] F. Esteva, L. Godo: Monoidal t-norm-based logic: towards a logic for left-continuous tnorms, *Fuzzy Sets and Systems*, **124**, 2001, 271-288.
[11] D. Pei: On equivalent forms of fuzzy logic systems NM and IMTL, *Fuzzy Sets and Systems*, **138**, 2003, 187-195.
[12] Zh. J. Wu, L. X. Yang, T. R. Li, K. Y. Qin: The basis algebra in L-fuzzy rough sets. *2009 International Conference on Rough Set and Knowledge Technology*, The Gold Coast, Austrlia, 2009, 320-325.

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15, August 2010, pp. 502-505

Study on Localization Algorithm of Mine Personnel Positioning System Based on Zigbee

Chen Yanli[1], Xu Xiaoling[2], Liu Xiaoyan[1]
[1]School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo,China
E-mail：{yanlichen,xyanliu}@hpu.edu.cn
[2]School of Mechanical and Electrical Engineering,Maoming University,Maoming,China
E-mail：xiaolingxu@163.com

Abstract-**For the demand of lower localization costs, less communications costs and higher accuracy of objects tracking under mine pit, weighted centroid localization algorithm based on RSSI is introduced to localize the node. This algorithm shows relatively high accuracy by simulation and is suitable for the localization of mine personnel.**

Index terms-**weighted centroid localization algorithm, RSSI,mine personnel localization**

I. INTRODUCTION

For the poor working conditions, complexity of the tunnel and being prone to happen of malignant events such as the gas explosion, water inrush, landslides under mine pits, lives of the staff are seriously threatened, and the development of mining enterprise and social stability are also restricted. With stricter safety production requirement in the domestic coal mine, it is most practically important to establish reliable personnel positioning system under mine pit for improving safety management. With the mine personnel positioning system based on Zigbee, we can know each person's real-time location and activity scope[1,2,3]. Location information, such as events occurred and sensor nodes, is very important for mine personnel positioning system, and monitoring information is unimportance without location information for monitoring. Therefore, it is crucial for mine personnel positioning system to ascertain accident position or acquire location node position.

Special nodes are arranged under coal mine in network in proportion, which energy is powerful and equipped with GPS system, or acquire themselves coordinate by other specific way. These special nodes announce their own position information to other nodes, and offer information to ascertain location for others. Calculate unknown node coordinate by measuring distance or angle between special nodes and other nodes, or the relative location for each other, and network connectivity.

II. MEASURING METHODS OF DISTANCE BETWEEN NODES OR ANGLE

Measuring methods of distance between nodes are TOA,TDOA,AOA, RSSI and so on. Although the first three methods have high measuring precision in theory, because of constraints in hardware, it is difficult to use under coal mine. Wireless transceiver is wireless sensors node own resources, by adopting the method of measuring RSSI to measure distance,and need not to add extra hardware. This is realistic and feasible method of sensor node positioning in mine personnel positioning system.

Based on RSSI positioning, firstly, test received power of receiving node, together with launching power of the launching node, convert propagation loss into the distance between nodes by signal propagation attenuation model, and then ascertain unknown node location by using localization algorithm. This technology mainly use RF signal [4]. For example, in free space, the antenna, which far from transmitter is *d*, receive signal intensity is illustrated by (1) as below:

$$P_r(d) = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 d^2 L} \tag{1}$$

P_t is launching power; $P_r(d)$ is receiving power on the distance of d point; G_t, G_r separately refers to gain of transmitting antenna and receiving antenna; d refers to distance, its unit is meter; L is system loss factor that is nothing to do with propagation; λ is wavelength, its unit is meter. Like this, by measuring receiving signal power and using formula (1), we can calculate approximate distance between receiving and launching node.

But the RSSI measured distance shows great instability by the analysis of RSSI localization algorithm, and it maybe arise location error in ±50°[4]. Therefore, we need to choose appropriate node localization algorithm aiming at its shortcoming, and improve in order to reduce localization error.

III. NODE LOCALIZATION ALGORITHM

Usually use connectivity to approximately estimate distance, among node localization algorithm of low cost priority. Measuring based on connectivity only simply show that whether two nodes are very near or connecting. Although receiving successfully data packets are random variation given received signal power and noises, to some extent connectivity provide node locations messages through binary variable.

Localization algorithm base on centroid is a simple centroid location algorithm base on connectivity.

Centroid location algorithm is an outdoor location algorithm based on network connectivity [5]. It used all beacon nodes in unknown node communication range as geometric centroid to measure unknown node position. Location process is as follows: beacon nodes transfer a beacon signal at intervals, and this signal includes its ID and location messages. When signal quantity which unknown node received from beacon nodes at a period of time is extend some presupposed threshold, the node will locate in a polygon centroid net which is consist of these connected beacons.

The most merit of centroid location algorithm absolutely base on the connectivity of network. It carries out simply and little calculation, but it needs more beacon nodes. In the large WSN distributing dense beacon nodes, this algorithm has these advantages. The distributing dense beacon bodes can increase probability of forming polygon between beacons and unknown nodes. It diminish location granularity and then improve of the accuracy of location estimation. In addition, little calculation can save power dissipation, and increase the effectiveness of network nodes.

IV. Wireless sensor location algorithms based on ZigBee

Combining denotation based on receiving signal intensity with weighted centroid localization algorithms to locate object, it make location very precise. Position of determinate event occurs or localization of acquiring node takes vital effect for the effectiveness of sensor network application.

Weighted centroid algorithms, mainly based on RSSI figure between fixed beacon nodes and unknown nodes, calculate weights of each fixed beacon node. Show the degree that fixed beacon node decision for centroid coordinate by weights, also speak influence that fixed beacon node for centroid localization, and reflect inner relationship between them.

There are n fixed beacon nodes in the network, and the i fixed beacon node of B_i with known coordinate (x_i, y_i) , $1 \le i \le n$. Then, estimated coordinate of unknown node M is (x_e, y_e). Thus, the formulas of weighted centroid algorithm are illustrated by (2):

$$x_e = \frac{\sum_{i=1}^{n} w_i \times x_i}{\sum_{i=1}^{n} w_i}, \quad y_e = \frac{\sum_{i=1}^{n} w_i \times y_i}{\sum_{i=1}^{n} w_i} \qquad (2)$$

w_i refers to weight of each fixed beacon node, which usually should be a function of distance between unknown node and fixed beacon node. If the unknown node cannot connect with beacon node B_i , w_i is zero.

For wireless sensor network, RSSI is affected by environment in a great degree, even RSSI are very different for the same node at same position in different environment. In addition, at the same circumstance, RSSI may be different if the node is at a different region or different direction although the distance is equal .This means that distance during different node is different at the same RSSI in the same network topology distribution. Corresponding weight should be also different. If calculate unknown node location just consider RSSI from unknown node to some fixed beacon node, without adding other modified method, it may lead to algorithm in a great error because RSSI is influenced greatly by environment. Therefore, we must consider distance of fixed beacon node and signal intensity information, and take the both kinds of information as reference to revise weight of each fixed beacon node.

V. Experimental Analysis

The function of algorithm is verified by changing the density of beacon nodes to verify the location error and comparing with the algorithm of the weighted centroid localization. Using Matlab integrated mathematics tools of software as the basic plate form of algorithm simulation, and using it can evaluate the performance of the wireless sensor network's location algorithm. Simulation area is a rectangle by 30m×40m .

(1) Influence on the location error in limited areas by increasing the number of beacon nodes

First, by a fixed unknown node (20, 20), through increase the number of beacon nodes gradually in rectangular areas to detect the location error between the weighted centroid algorithm based on RSSI and the weighted centroid algorithm.

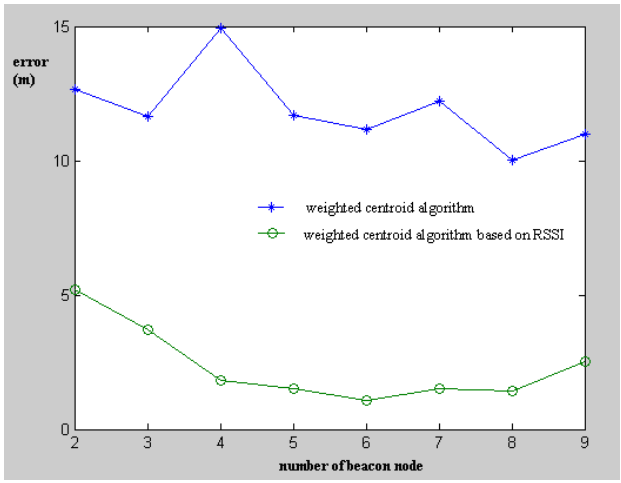


Fig.1 Localization error comparison on the increasing number of beacon node

In Fig. 1, it shows the location error of the weighted centroid algorithm based on RSSI less than the location of the weighted centroid algorithm. Moreover, when the number of the beacon nodes greater than two nodes,

the error of nodes location below 5m. Therefore, during the actual measurement, it requires the unknown nodes to maintain communication with less than three beacon nodes at least thus improving the position accuracy.

(2) Influence on the location error by changing the position of the fixed number of the beacon nodes

In reality, it is impossible to increase the beacon nodes arbitrarily within the restricted scope to improve the accuracy. Thus, it is necessary to get a balance between the number of beacon nodes and the location accuracy. By using a fixed number of beacon nodes and changing the location of the beacon nodes, the influence on location error can be got as following.

In the simulation processing, through fixing two beacon nodes (0, 0), (40, 30), moving two beacon nodes (0, 30) and (30, 40), meanwhile, changing the position of the beacon nodes (0, 30) to the beacon nodes (-10, 30), (-20, 30), … , (-60, 30) towards the left. Meanwhile, changing the position of the beacon nodes (40, 0) to the beacon nodes (50, 0), (60, 0) …, (100, 0) toward the right. That is to say, the beacon nodes composed a chart which changes from the rectangle into the serration, when it extends without restriction, it will change into a straight line. With the density of a fixed number of beacon nodes changes, finally we will get the location error of the unknown beacon nodes through the Matlab simulation as in Fig. 2.

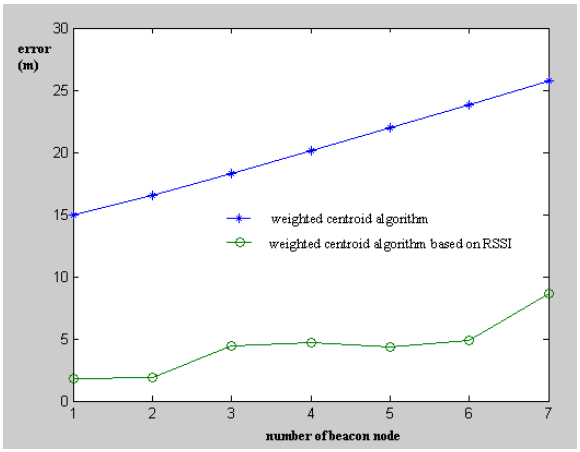


Fig.2 Localization error comparison on the increasing beacon node density

Fig.2 shows that the locating error increases gradually with the distance of the beacon nodes increases. By comparison, the accuracy of the algorithm of the Moreover, when the number of the beacon nodes weighted centroid based RSSI is better than the algorithm of the weighted centroid. Especially, in coal mine well, the environment is abominable, and the arrangement of the beacon nodes of is very difficult, the advantage of the algorithm of the weighted centroid based on RSSI is more prominent. In the practical application, adjusting the arrangement of the beacon nodes properly within the scope of the error permitted can reduce the use of the beacon nodes.

（3)Influence of the unknown nodes location on the nodes locating error

Fixing the beacon nodes on the apex of rectangles which is 40m by 30m, in this area we select six vertex at random which are (10, 20), (20, 20), (30, 20), (10, 10), (20, 10) and (30, 10) as the location of the unknown nodes, and calculate. Figure 3 shows the relationship of the location error where fixing the unknown nodes on the six vertex.

When the number of the beacon nodes is unchanged, the centroid algorithm has bigger influence on the location error of unknown nodes at different positions. Contrasting improved algorithm to the algorithm of the weighted centroid, for increase field density of the fixed beacon nodes as weighting, which makes the algorithm become more stable and accurate after improving.

By analysis on localization error from the above three aspects, we realize that weighted centroid algorithms based on RSSI has better localization precision than weighted centroid algorithm. For localization tracking applications in mine pit, weighted centroid algorithms based on RSSI introduces a logarithmic function with a normal distribution to describe electromagnetic wave propagation under mine, which is more suitable to practical surrounding of application. At the same time, when we select localization algorithms, we directly adopt receiving field density denotation of node itself to location measuring and localization calculation by weighted centroid algorithms; it is simple and easy to satisfy the requirement of wireless sensor network tracking.

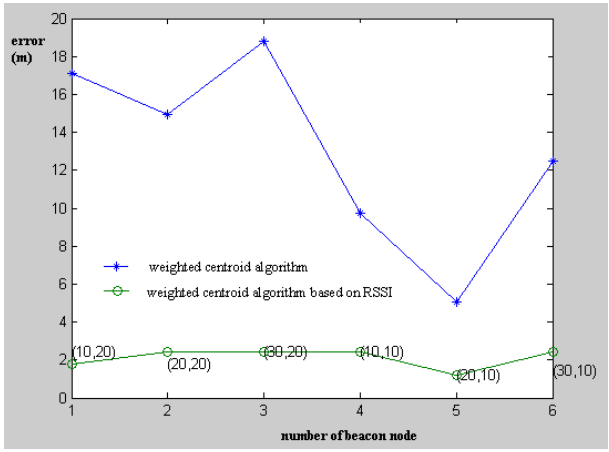


Fig.3 Localization error comparison on the fixed number of beacon node

VI.CONCLUSION

According to targets tracking requires location for lower costs and lower communications costs and higher accuracy, for miner personnel positioning methods making detailed study, raising the algorithm of the weighted centroid localization based on RSSI. It has a higher precision, which is a lower costs of locating solution suitable for miner personnel location and proved the advantages of the algorithm by stimulation experiments.

REFERENCE

[1] Zhan Haomin, Sun Changsong, Wu Shan,Li Dongyan. The Application of the ZigBee Technology in Coal Mine Rescue System[J]. Computer Engineering and Applications, Vol.24, pp:181-183, Jun, 2006 (in Chinese)

[2] Zhang Zhibin, Xu Xiaoling, YAN Lianlong. Underground localization algorithm of wireless sensor network based on Zigbee [J]. Journal of China Coal Society, vol.34, pp125-128, Jan,2009 (in Chinese)

[3] Wang Yufen, Zhang Zhibin, Li Changjiang. Wireless sensor network applied to coal mine gas monitoring and measuring system [J]. Coal Science and Technology,Vol.35, pp:34-36,Aug,2007(in Chinese)

[4] Duan Weijun,Wang Jiangang,Wang Fubao. Research and Development of Localization Systems and Algorithms for Wireless Sensor Networks[J]. Information and Control,Vol.35, pp239-244, April,2006 (in Chinese)

[5] Bulusu B, Heidemann J, Estrin D. Dendity adaptive algorithms for beacon placement in wireless sensor networks. In:IEEE ICDCS'01, Phoenix AZ. April,2001

ISBN 978-952-5726-10-7
Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10)
Jiaozuo, P. R. China, 14-15, August 2010, pp. 506-509

Semi-Supervised Dimensionality Reduction

Yongmao Wang[1,2], Yukun Wang[1]
[1]School of Computer Science and Technology,Henan Polytechnic University,454000 Jiaozuo, Henan, China
E-mail:wymyjs2000@hpu.edu.cn
[2] School of Information Engineering ,University of Science and Technology Beijing ,10083 Beijing, China
E-mail:wyk@hpu.edu.cnt

***Abstract*—In many domains, dimensionality reduction is taken into account in order to avoid the curse of dimensionality. This paper proposes a semi-supervised dimensionality reduction algorithm (SSDR) by combining Locality Preserving Projection (LPP, an unsupervised dimensionality reduction algorithm) and Local Fisher Discriminant Analysis (LFDA, a supervised dimensionality reduction algorithm). SSDR can not only preserve the intrinsic structure of the unlabeled data but also solve the problem of multimodal in the projected low-dimensional space. Experiments show that SSDR is superior to many established dimensionality reduction methods.**

***Keywords*—multimodal,LPP,LFDA,Dimensionality Reduciton, Semi-supervisedg**

I. INTRODUCTION

In many domains such as content based image retrieval (CBIR), face recognition, one is confronted with high dimensional data which maybe result in dimensionality curse.

The curse of dimensionality has a bad effect on the classification, visualization. However, it is assumed that high dimensional data is embedded in a lower dimensional manifold. Dimensionality reduction is the transformation of high-dimensional data into a meaningful representation of reduced dimensionality. According to whether supervised information is available or not, existing dimensionality reduction methods can be roughly categorized into supervised ones and unsupervised ones.

There are many unsupervised dimensionality reduction methods such as Principal Components Analysis (PCA) [1], Locality Preserving Projection (LPP) [2] and Neighborhood Preserving Embedding (NPE) [3] . PCA tends to preserve the global structure of data. LPP and NPE tend to preserve the local structure of data. PCA finds a linear transformation such that the amount of variance of the data in reduced lower dimensional space is maximal. LPP seeks an embedding transformation such that nearby sample pairs in the original high-dimensional space are kept close in the embedding space. NPE preserves the local structure which means that each data point can be represented as a linear combination of its neighbors.

In the supervised learning scenario where data samples are accompanied with supervised information such as class labels, pair wise constraints or other prior information. Here we focus on the supervised information in form of class labels. Fisher Discriminant Analysis (FDA) [4] is a popular dimensionality reduction method. FDA seeks a linear transformation such that the between-class scatter is maximized and the within-class scatter is minimized. However, but FDA tends to give undesired results if samples in a class form several separate clusters, which is called multimodal. Local Fisher Discriminant Analysis (LFDA) [5] is proposed to improve the performance of FDA. LFDA effectively combines the ideas of FDA and LPP. If the samples with class label are sufficient, the supervised methods can achieve the better performance than the unsupervised ones. But, unlabeled training examples are readily available but labeled ones are fairly expensive to obtain. Therefore, the covariance matrix of each class may not be accurately estimated if labeled training examples are insufficient.

In order to solve the problem in supervised and unsupervised methods, semi-supervised dimensionality reduction is proposed, which is a new issue in semi-supervised learning, which learns from a combination of both labeled and unlabeled data. In this paper, we propose a new semi-supervised dimensionality reduction algorithm (SSDR) combining LFDA and LPP, which can trade between-class separation with locality preservation. SSDR can be performed by Eigen decomposition, which is efficient and reliable.

The rest of this paper is organized as follows. In Section II, we provide a brief review of FDA, LFDA and LPP. In Section III, we introduce the concept of SSDR. The experimental results are presented in Section IV. Finally, we conclude the paper and provide suggestions for future work in Section V.

II. FDA, LFDA AND LPP

A. FDA

FDA is a popular supervised dimensionality reduction method. Suppose we have a set of l samples $x_1, x_2, \cdots, x_l \in R^n$, belonging to c classes. FDA seeks the transformation w which maximizes the following objective function

$$J(w) = \frac{w^T S_B w}{w^T S_W w} \tag{1}$$

$$S_B = \sum_{k=1}^{c} n_k (u_k - u)(u_k - u)^T \tag{2}$$

$$S_W = \sum_{k=1}^{c} \sum_{i=1}^{n_k} (x_i - u_k)(x_i - u_k)^T \tag{3}$$

where S_W and S_B are respectively the within-class scatter matrix and the between-class scatter matrix, u is the total sample mean vector, n_k is the number of samples in the k-th class, u_k is the average vector of the k-th class, and x_i is the i-th sample in the k-th class.

The optimal **w**'s are the eigenvectors corresponding to the non-zero eigenvalue of eigen-problem:

$$S_B w = \lambda S_w w \tag{4}$$

Since the rank of S_B is bounded by c-1, there are at most c-1 eigenvectors corresponding to non-zero eigenvalues.

FDA can reformulate FDA in a *pairwise* manner. Therefore, (2) and (3) can be expressed as

$$\begin{aligned} S_W &= \frac{1}{2}\sum_{i,j=1}^{l} W_{ij}^{w}\left(x_i - x_j\right)\left(x_i - x_j\right)^T \\ &= X(D^w - W^w)X^T \end{aligned} \tag{5}$$

$$\begin{aligned} S_B &= \frac{1}{2}\sum_{i,j=1}^{l} W_{ij}^{b}\left(x_i - x_j\right)\left(x_i - x_j\right)^T \\ &= X(D^b - W^b)X^T \end{aligned} \tag{6}$$

where

$$W_{ij}^{w} = \begin{cases} 1/n_k & x_i \text{ and } x_j \textit{ are in the same class} \\ 0 & x_i \text{ and } x_j \textit{ are in the different class} \end{cases} \tag{7}$$

$$W_{ij}^{b} = \begin{cases} 1/n - 1/n_k & x_i \text{ and } x_j \textit{ are in the same class} \\ 1/n & x_i \text{ and } x_j \textit{ are not in the same class} \end{cases} \tag{8}$$

Either of D^w *and* D^b is a diagonal matrix whose entries are column (or row, since W^w and W^b are all symmetric) sum of W^w and W^b respectively, $D_{ii}^{w} = \sum_j W_{ij}^{w}$, $D_{ii}^{b} = \sum_j W_{ij}^{b}$. So, (5) can be express as

$$X\left(D^b - W^b\right)X^T w = \lambda X\left(D^w - W^w\right)X^T w \tag{9}$$

B. LFDA

In LFDA, we define a affinity matrix $A_{i;j}$, which weight the values for the sample pairs in the same class. The main concept of LFDA is that far apart sample pairs in the same class have less influence on S_W and S_B. Based on the above pairwise expression in part A of section II, the local within-class scatter matrix S_{lw} and the local between-class scatter matrix S_{lB} are defined as

$$\begin{aligned} S_{lW} &= \frac{1}{2}\sum_{i,j=1}^{l} W_{ij}^{lw}\left(x_i - x_j\right)\left(x_i - x_j\right)^T \\ &= X(D^{lw} - W^{lw})X^T \end{aligned} \tag{10}$$

$$\begin{aligned} S_{lB} &= \frac{1}{2}\sum_{i,j=1}^{l} W_{ij}^{lb}\left(x_i - x_j\right)\left(x_i - x_j\right)^T \\ &= X(D^{lb} - W^{lb})X^T \end{aligned} \tag{11}$$

where

$$W_{ij}^{lw} = \begin{cases} A_{ij}/n_k & x_i \text{ and } x_j \textit{ are in the same class} \\ 0 & x_i \text{ and } x_j \textit{ are in the different class} \end{cases} \tag{12}$$

$$W_{ij}^{lb} = \begin{cases} A_{ij}\left(1/n - 1/n_k\right) & x_i \text{ and } x_j \textit{ are in the same class} \\ 1/n & x_i \text{ and } x_j \textit{ are in the different class} \end{cases} \tag{13}$$

Either of D^{lw} *and* D^{lb} is a diagonal matrix whose entries are column (or row, since W^{lw} and W^{lb} are all symmetric) sum of W^{lw} and W^{lb} respectively, $D_{ii}^{lw} = \sum_j W_{ij}^{lw}$, $D_{ii}^{lb} = \sum_j W_{ij}^{lb}$.

LFDA finds the transformation w which maximizes the objective function

$$J(w) = \frac{w^T S_{lB} w}{w^T S_{lW} w} \tag{14}$$

The optimal **w**'s are the eigenvectors corresponding to the non-zero eigenvalue of eigen-problem:

$$X\left(D^{lb} - W^{lb}\right)X^T w = \lambda X\left(D^{lw} - W^{lw}\right)X^T w \tag{15}$$

C. LPP

The primary consideration of LPP is to preserve the neighborhood structure of the data set. LPP seeks an embedding transformation w such that nearby sample pairs in the original high-dimensional space are kept close in the embedding space. Given a set of examples, we can use a p-nearest neighbor graph G to model the relationship between nearby data points. Specifically, we put an edge between nodes i and j if x_i and x_j are "close", i.e., x_i and x_j are among p nearest neighbors of each other. Let the corresponding weight matrix be S, define by

$$S_{ij} = \begin{cases} 1 & \textit{if } x_i \in N_p\left(x_j\right) \textit{ or } x_j \in N_p \in \left(x_i\right) \\ 0 & \textit{otherwise} \end{cases} \tag{16}$$

where $N_p(x_i)$ denotes the set of p nearest neighbors of x_i. Therefore, LPP finds a transformation w to make the following objective function minimized.

$$\begin{aligned} J(w) &= \frac{1}{2}\sum_{i,j}\left(w^T x_i - w^T x_j\right)S_{ij} \\ &= w^T X\left(D - S\right)X^T w \end{aligned} \tag{17}$$

where D is a diagonal matrix whose entries are column (or row, since S is symmetric) sums of S, $D_{ii}=\sum S_{i,j}$

The optimal **w**'s are the eigenvectors corresponding to the non-zero eigenvalue of eigen-problem:

$$X(D-S)X^T w = \lambda XDX^T w \tag{18}$$

III. Semi-Supervised Dimensionality Reduction (SSDR)

If only a small number of labeled samples are available, supervised dimensionality reduction methods tend to overfit the embedding space to the labeled samples; thus their performance can be heavily degraded. In such cases, it is effective to utilize unlabeled samples which are often available abundantly. Based on the above idea, in this section, we propose a new semi-supervised dimensionality reduction method by combining LFDA and LPP.

We combine the objective function of LFDA and LPP to form a new objective function. SSDR seeks the transformation w which maximizes the objective function.

$$J(w) = \frac{w^T S_{lB} w}{w^T S_{lW} w + \alpha w^T X(D-S)X^T w} \tag{19}$$

The optimal **w**'s are the eigenvectors corresponding to the non-zero eigenvalue of eigen-problem:

$$X(D^{lb} - W^{lb})X^T w = \lambda X(D^{lw} - W^{lw} + \alpha(D-S))X^T w \tag{20}$$

LFDA is defined only for labeled samples. Therefore, when computing the S_{lB} and S_{lW}, we assign zero to W_{ij}^{lb} and W_{ij}^{lw} if at least one of x_i and x_j is unlabelled. If both of them are labeled, we compute W_{ij}^{lb} and W_{ij}^{lw} by (12) and (13) as usual. Similarly, LPP is originally defined only for unlabeled samples. When computing S, we treat all samples as unlabeled.

IV. Experiment and Result

Our experiments are running under Processor: Intel Core(TM) Duo T2400 with 1 GB RAM and Matlab version 7.0 using Database toolbox.

To test our algorithm, we use natural images mostly from "Corel Image Gallery" [6] , where are about 67000 images. In this paper, we select 1000 from 10 categories. Some sample images are shown in Figure 1.

We use 48-dimensional Color Texture Moment

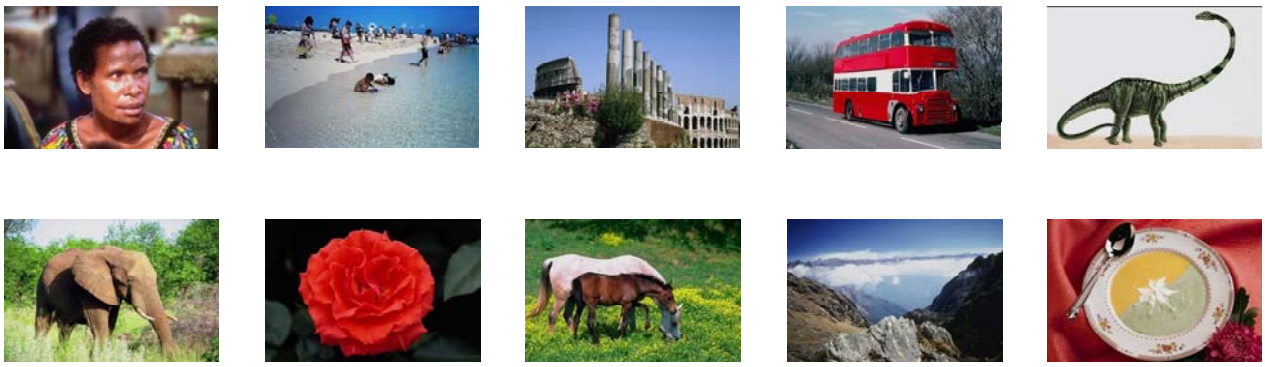


Figure 1. Sample image from Corel Image Gallery

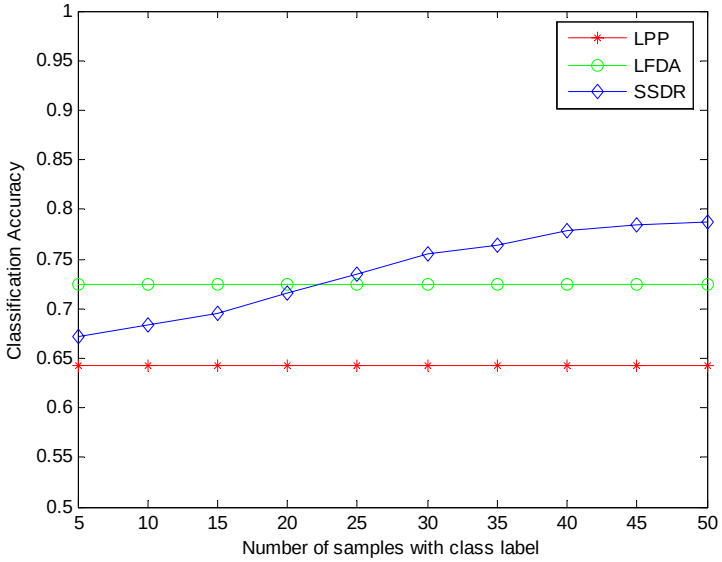


Figure 2. Comparation of LPP,LFDA and SSDR

(CTM) [7] to represent the images. CTM integrates the color and texture characteristics of the image in a compact form. CTM adopts local Fourier transform as a texture representation scheme and derives eight characteristic maps for describing different aspects of co-occurrence relations of image pixels in each channel of the (SVcosH, SVsinH, V) color space. Then CTM calculates the first and second moment of these maps as a representation of the natural color image pixel distribution.

We use classification accuracy to evaluate the performance of SSDR. We compare SSDR with LPP and LFDA when the number of samples with class label is changed. After dimensionality reduction, nearest neighborhood (1-NN) classifier is employed for classification. For each data set, we use the first half of the data for training (learning the projections) and the remaining data for testing. We conduct the experiment for twenty times and take the average accuracy percentage as the result. Figure 2 shows SSDR can achieve the high accuracy.

V. CONCLUSION

In this paper, we propose a simple but efficient semi supervised dimensionality reduction algorithm called SSDR, which combines LPP and LFDA. SSDR can not only preserve the intrinsic structure of the unlabeled data but also solve the problem of multimodal in the projected low-dimensional space. Experiments show that SSDR leads to considerable improvements in embedding, classification over conventional dimensionality reduction methods.

REFERENCES

[1] Turk MA and Pentland AP. “Face recognition using eigenfaces”. Proc. IEEE Conference on Computer Vision and Pattern Recognition. Madison: IEEE Computer Society, 1991. pp.586−591

[2] X. He and P. Niyogi. “Locality preserving projections”. *Advances* in Neural Information Processing Systems 16.Vancouver,British Columbia, Canada, 2003

[3] X. He, D. Cai, S. Yan and H. Z. “Neighborhood Preserving Embedding”. Proc. IEEE International Conference on Computer Vision (ICCV 05), IEEE Press, Oct. 2005, pp. 1-6

[4] Martinez AM and Kak AC. PCA Versus LDA. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2001,23(2):pp.228−233

[5] Masashi Sugiyama. “Dimensionality Reduction of Multimodal Labeled Databy Local Fisher Discriminant Analysis”. Journal of Machine Learning Research,2007,5:pp.1027-1061

[6] Y. Wang and Z. Xu. “Image retrieval using the color approximation histogram based on rough set theory”. Proc. IEEE International Conference on Information Engineering and Computer Science (ICIECS 09), IEEE Press, Dec. 2009, pp. 1452-1455

[7] H. Yu, M. Li, H. Zhang, and J. Feng. “Color texture moments for content-based image retrieval”. Proc. IEEE International Conference on Image Processing (ICIP 02), Dec.2002,pp.24–28

Author Index