Global Journal of Computer Science and Technology

discovering thoughts and inventing future

Volume 10 Issue 1 Version 1.0 | Online ISSN: 0975-4172 | Print ISSN: 0975-4350

12 Technology Changing Ideas

highlights

- Knowledge Management System
- Risk Factors in Software Development
- Lightweight Multipriority Framework
- Detection of IP Address Conflicts

April 2010

ENG

Global Journal of Computer Science and Technology

Global Journal of Computer Science and Technology

Volume 10 Issue 1 (Ver. 1.0)

Global Academy of Research and Development

Editorial Board Members

Dr. Carlos García Pont
Associate Professor of Marketing
IESE Business School, University of Navarra
Doctor of Philosophy (Management), Massachussetts Institute of Technology (MIT)
Master in Business Administration, IESE, University of Navarra
Degree in Industrial Engineering, Universitat Politècnica de Catalunya

Dr. Fotini Labropulu
Mathematics - Luther College
University of Regina
Ph.D., M.Sc. in Mathematics
B.A. (Honors) in Mathematics
University of Windsor

Dr. Lynn Lim
Reader in Business and Marketing
Roehampton University, London
BCom, PGDip, MBA (Distinction), PhD, FHEA

Dr. Mihaly Mezei
ASSOCIATE PROFESSOR
Department of Structural and Chemical Biology
Mount Sinai School of Medical Center
Ph.D., Etvs Lornd University
Postdoctoral Training, New York University

Dr. Miguel Angel Ariño
Professor of Decision Sciences
IESE Business School
Barcelona, Spain (Universidad de Navarra)
CEIBS (China Europe International Business School).
Beijing, Shanghai and Shenzhen
Ph.D. in Mathematics
University of Barcelona
BA in Mathematics (Licenciatura)
University of Barcelona

Philip G. Moscoso
Technology and Operations Management
IESE Business School, University of Navarra
Ph.D in Industrial Engineering and Management, ETH Zurich
M.Sc. in Chemical Engineering, ETH Zurich

Dr. Sanjay Dixit, M.D.
Director, EP Laboratories, Philadelphia VA Medical Center
Cardiovascular Medicine - Cardiac Arrhythmia
Univ of Penn School of Medicine

Dr. Han-Xiang Deng
MD., Ph.D
Associate Professor and Research Department
Division of Neuromuscular Medicine
Davee Department of Neurology and Clinical Neurosciences
Northwestern University Feinberg School of Medicine

Chief Author

Dr. R.K. Dixit (HON.)
M.Sc., Ph.D., FICCT
Chief Author, India
Email: authorind@computerresearch.org

Dean & Editor-in-Chief (HON.)

Vivek Dubey(HON.)
MS (Industrial Engineering),
MS (Mechanical Engineering)
University of Wisconsin
FICCT
Editor-in-Chief, USA
editorusa@computerresearch.org

Er. Suyog Dixit
BE (HONS. in Computer Science), FICCT
SAP Certified Consultant
Technical Dean, India
Website: www.suyogdixit.com
Email:suyog@suyogdixit.com,
dean@computerresearch.org

Sangita Dixit
M.Sc., FICCT
Dean and Publisher, India
deanind@computerresearch.org

Contents of the Volume

From the Chief Author's Desk

The research activities among different disciplines of natural science are backbone of system. The deep and strong affords are the demands of today. Sincere afford must be exposed worldwide. Which, in turns, require international platform for rapid and proper communication among similar and interdisciplinary research groups.

The Global Journal of Computer Science and Technology is to fulfill all such demands and requirements, and functions also as an international platform. Of course, the publication of research work must be reviewed to establish its authenticity. This helps to promote research activity also. We know, great scientific research have been worked out by philosopher seeking to verify quite erroneous theories about the nature of things.

The research activities are increasing exponentially. These great increments require rapid communication, also to link up with others. The balanced communication among same and interdisciplinary research groups is major hurdle to aware with status of any research field.

The Global Journals is proving as milestone of research publication. In view of whole spectrum of Knowledge, the research work of different streams may be considered as branches of big tree. Every branch is of great importance. Thus, we look after the complete spectrum as whole. Global Journals let play all the instruments simultaneously. We hope, affords of Global Journals will sincerely help to build the world in new shape.

Dr. R. K. Dixit
Chief Author
chiefauthor@globaljournals.org

Face and gender Recognition Using Genetic Algorithm and Hopfield Neural Network

C.R Vimal Chand

GJCST Computing Classification
I.4.0, I.2.6, F.1.1, I.1.2

***Abstract*- This paper describes a face recognition system for personal identification and verification using genetic algorithm and Hopfield Neural Network. This FRS system is also being trained for gender identification. Face recognition system consists of three steps. At the initial stage of this system some pre-processing are applied on the input image. Secondly, face features are extracted, which will be taken as the input of the eight parallel Hopfield neural network and genetic algorithm (GA). In the third step, classification is carried out by using Hopfield neural network and GA to identify gender. The proposed approaches can be tested on a number of face images. Sex-recognition in faces is a prototypical pattern recognition task and it appears to follow no simple algorithm. It is modifiable according to fashion (makeup, hair etc).While ambiguous cases exist, for which we must appeal to other cues such as physical build (if visible), voice pattern (if audible) and mannerisms.**

I INTRODUCTION

The literature has some works related to facial expressions recognition, such as Multi-layer Perceptron [1], FRS using Back propagation neural network[2], and also works that apply the Hopfield network to detect face expressions[3,4] and nominal color coding of classified images[5]. Among many recognition subjects, face recognition has drawn considerable interest and attention from many researchers for the last two decades because of its potential applications, such as in the areas of surveillance, secure trading terminals, Closed Circuit Television (CCTV) control, user authentication, HCI Human Computer Interface, intelligent robot and so on. A number of face recognition methods have been proposed [6] [7] and some related face recognition systems have been developed. In this paper we proposed a computational model of face recognition, which is fast, reasonably simple, and accurate in constrained environments such as an office or a household. The proposed approaches have advantages over the other face recognition schemes in its speed and simplicity, learning capacity and relative insensitivity to small or gradual changes in the face image.

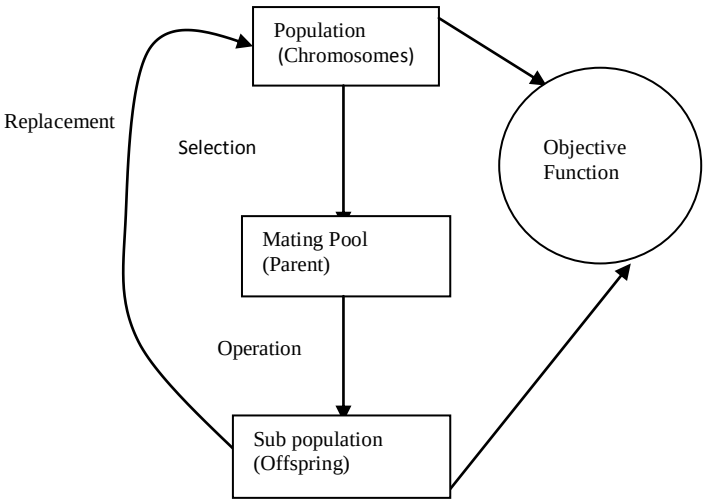
Ma et al. [1] propose facial expression recognition using an MLP architecture allied to image compression techniques, optimization algorithms such as quasi-Newton and pruning methods. In an attempt to simplify this approach, and yet keeping the rate results between 95% and 100% of recognition, a Hopfield neural network can be used for further processing. Ma et al.'s work recognizes the facial expression on each image that is given to the neural network. This FRS system attempts to, beyond recognize, to recover the image of a face without facial expression corresponding to the image of a face with an expression, using Hopfield neural network like associative memory.

II OUTLINE OF THE SYSTEM

The design and implementation of the Face Recognition System (FRS) can be subdivided into two main parts. The first part is *image processing* and the second part is *recognition techniques*. The image processing part consists of Face image acquisition through scanning, Image enhancement, Image clipping, Filtering, Edge detection and Feature extraction. The second part consists of the artificial intelligence which is composed by Genetic Algorithm and Hopfield Neural Network.

The first part of FRS consists of several image processing techniques. Firstly, face's image acquisition then image clipping is performed using start-point and end-point detection algorithm. Then the edges are detected using high-pass filter, high-boost filter, median filter or several edge detection methods. Finally, the features are extracted. These extracted features of image are then fed into Genetic algorithm and Hopfield Neural Network.

In the second part two techniques are used one is based on Genetic algorithm and another one is based on Hopfield neural network. In the first techniques, the extracted features are saved into memory and using genetic algorithm; the recognition of unknown face image is performed by comparing this special pattern to the pattern for which an image module is already built. In the second technique extracted features are given as input to the Hopfield network and it is trained to create a knowledge base for recognition.



III METHODOLOGY

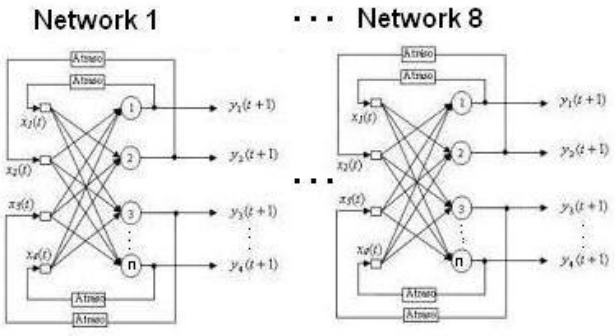
After the acquisition of the image the features are extracted. To extract features of a face at first the image is converted into a binary. From this binary image the centroid of the face image is calculated using the below equations.

$$X = \frac{\sum mx}{\sum m} \qquad Y = \frac{\sum my}{\sum m}$$

Where x, y is the co-ordinate values and m=f(x,y)=0 or1.Then from the centroid, only face has been cropped and converted into the gray level and the features have been collected.

Extracted features of the face images have been fed in to the Genetic algorithm and Hopfield Neural Network for recognition. The unknown input face image has been recognized by Genetic Algorithm and Hopfield Neural Network. The recognized image is compressed and given as input for gender identification.

The face images to be presented to the neural model receive previous processing: first, a dimension reduction, from 350x275 to 32x32 pixels (this do not cause great losses in the image characteristics), and next, the creation of a vector formed by the lines of the reduced image matrix (32x32), lined up in a single vector column with 1024 positions. The initial idea was to use the vector as an input to a single Hopfield network composed by 1024 neurons. This model was tested and proved its inefficiency due to confusion in relation to the gray levels, providing unsatisfactory results. It was necessary to separate the gray levels and to create a network for each level. Thus, the vector with 1024 positions was sliced into 8 bits, one for each gray. Since we now had eight times more data than before, changes on the network topology were needed. Following this idea, we decided to test a model with 8 parallel networks, one for each gray level. Each network works only in one level of the image and retrieves only the same level of the image. The inverse process to the previous processing is then applied to reconstruct the images that were recovered from the neural networks. The model of the Hopfield network is given model.



IV CONCLUSION

In this paper present model of Face Recognition System using the concept of Genetic algorithm and Step Error Tolerance Hopfield Neural Network and digital image processing has been discussed. Here a static Face Recognition system has been developed. The maximum efficiency is 82.61% for Face Recognition System by using Genetic algorithm and the maximum efficiency is 91.30% for Face Recognition System by using 8 parallel Hopfield network is expected. The efficiency can be increased by better technique of scaling, efficient technique of edge detection such as advanced edge detection technique and feature extraction. The same network is trained to identify the gender based on the features extracted.

V REFERENCES

1) Ma, L.; Khorasani, K.; "Facial expression recognition using constructive feed forward neural networks".Systems, Man and Cybernetics, Part B, IEEETransactions, June 2004
2) Rosenblum, M.; Yacoob, Y.; Davis, L. S. "Human expression recognition from motion using a radial basisfunction network architecture". IEEE Trans. Neural Networks, September 1996.
3) Yoneyama, M.; Iwano, Y.; Ohtake, A.; Shirai, K."Facial Expressions Recognition Using Discrete Hopfield Neural Network". International Conference on Image Processing (ICIP '97) 1997.
4) Ying, D; Shibata, Y.; Nakano, Y.; Hashimoto, K. "Recognition of facial expressions based on the Hopfield memory model". Multimedia Computing and Systems, IEEE International Conference, June 1999.
5) Campadelli, P.; Mora, P.; Schettini, R. "Using Hopfield networks in the nominal color coding of classified images". Conference B: Computer Vision & Image Processing., IEEE CNF Proceedings of the 12th IAPR
6) R. Chellappa, C.L. Wilson, S. Sirohey (1995), "Human and machines recognition of faces: a survey", Proc. IEEE 83(5): 705-740.
7) A.Samal and P.A.Iyengar (1992): "Automatic recognition and analysis of human faces and facial expressions: A survey. Pattern Recognition".
8) Rafael C. Gonzalez and Richard E Woods, "Digital Image Processing", Person Education Asia.
9) Rafael C. Gonzalez and Richard E Woods," Digital Image Processing Using MATLAB7", Person Education Asia.
10) Goldberg, D.E., "Genetic Algorithms in Search, Optimization and Machine Learning", Pearson Education, Singapore, 2002.
11) Sarawat Anam, Md. Shohidul Islam, M.A. Kashem, M.N. Islam, M.R. Islam, M.S. Islam, "Face Recognition Using Back Propagation Neural Network", Proceedings of the International MultiConference of Engineers and Computer Scientists 2009 Vol I
12) Tranel, D., Damasio, A.R., & Damasio, H." Intact recognition of facial expression, gender, and age in patients with impaired recognition of face identity".

An Analysis of LSB & DCT based Steganography

Dr. Ekta Walia [a], Payal Jain [b], Navdeep [c]

GJCST Computing Classification
F.2.1 & G.2.m

***Abstract*- This paper presents analysis of Least Significant Bit (LSB) based Steganography and Discrete Cosine Transform (DCT) based Steganography. LSB based Steganography embed the text message in least significant bits of digital picture. Least significant bit (LSB) insertion is a common, simple approach to embedding information in a cover file. Unfortunately, it is vulnerable to even a small image manipulation. Converting an image from a format like GIF or BMP, which reconstructs the original message exactly (lossless compression) to a JPEG, which does not (lossy compression), and then back could destroy the information hidden in the LSBs. DCT based Steganography embed the text message in least significant bits of the Discrete Cosine (DC) coefficient of digital picture. When information is hidden inside video, the program hiding the information usually performs the DCT. DCT works by slightly changing each of the images in the video, only to the extent that is not noticeable by the human eye. An implementation of both these methods and their performance analysis has been done in this paper.**

Keywords- Least Significant Bit (LSB), Discrete Cosine Transform (DCT), Steganography

I INTRODUCTION

Steganography comes from the Greek words Steganós (Covered) and Graptos (Writing). Steganography in these days refers to information or a file that has been concealed inside a digital picture, video or audio file. If a person or persons view the object that the information is hidden inside, he or she will have no idea that there is any hidden information; therefore the person will not attempt to decrypt the information.

[a] Professor, Department of Information and Technology, Maharishi Markandeshwar College of Engineering Maharishi Markandeshwar University, Mullana, Ambala(Haryana)
E-mail: wekta@yahoo.com, Tel No: 91-9416551292[a]
[b] Lecturer, Department of Information and Technology, Maharishi Markandeshwar College of Engineering Maharishi Markandeshwar University, Mullana, Ambala (Haryana)
payaljain2006@gmail.com[b], Tel No: 91-9466742552[b]
[c]Student, Department of Information and Technology, Maharishi Markandeshwar College of Engineering Maharishi Markandeshwar University, Mullana, Ambala(Haryana)

A. Steganographic Techniques

i. Physical Steganography

Physical Steganography has been widely used. In ancient time people wrote message on wood and then covered it with wax. Message was written on the back of postage stamps. Message was written on paper by secret inks.

ii. Digital Steganography

Digital Steganography is the art of invisibly hiding data within data. It conceals the fact that message exists by hiding the actual message. In this, secret data can be hidden inside the image, text, sound clip which can be represented in binary.

iii. Printed Steganography

Digital Steganography output can be in the form of printed documents. The letter size, spacing and other characteristics of a cover text can be manipulated to carry the hidden message. A recipient who knows the technique used can recover the message and then decrypt it.

II METHODS OF CONCEALING DATA IN DIGITAL IMAGE

A. Least Significant Bit (Lsb)

LSB is the lowest bit in a series of numbers in binary. e.g. in the binary number: 10110001, the least significant bit is far right 1.
The LSB based Steganography is one of the steganographic methods, used to embed the secret data in to the least significant bits of the pixel values in a cover image. e.g. 240 can be hidden in the first eight bytes of three pixels in a 24 bit image.

PIXELS:	(00100111 11101001 11001000)
	(00100111 11001000 11101001)
	(11001000 00100111 11101001)
240 :	011110000
RESULT:	(00100110 11101001 11001001)
	(00100111 11001001 11101000)
	(11001000 00100110 11101000)

Here number 240 is embedded into first eight bytes of the grid and only 6 bits are changed.

B. *Discrete Cosine Transform (Dct)*

DCT coefficients are used for JPEG compression. It separates the image into parts of differing importance. It transforms a signal or image from the spatial domain to the frequency domain. It can separate the image into high, middle and low frequency components.

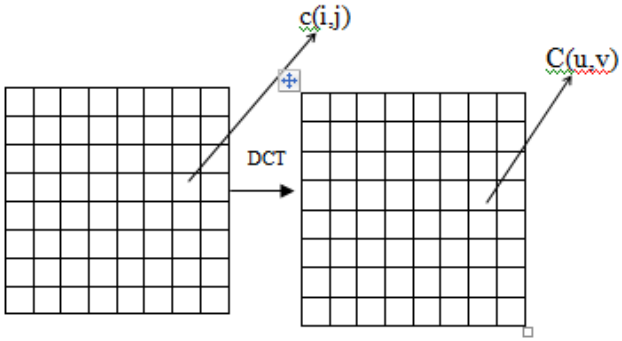


Fig. I Discrete Cosine Transform of An Image

The general equation for a 1D (N data items) DCT is defined by the following equation:

$$C(u) = \alpha(u)\sum_{x=0}^{N-1} f(x)\cos\left[\frac{(2x+1)u\pi}{2N}\right] \quad (1)$$

for u = 0, 1, 2, . . . , N-1.

The general equation for a 2D (*N* by *M* image) DCT is defined by the following equation:

$$C(u,v) = \alpha(u)\alpha(v)\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} f(x,y)\cos\left[\frac{(2x+1)u\pi}{2N}\right]\cos\left[\frac{(2y+1)v\pi}{2N}\right] \quad (2)$$

for u,v = 0, 1, 2, . . . , N-1

Here, the input image is of size N X M. c(i, j) is the intensity of the pixel in row i and column j; C(u,v) is the DCT coefficient in row u and column v of the DCT matrix.

Signal energy lies at low frequency in image; it appears in the upper left corner of the DCT. Compression can be achieved since the lower right values represent higher frequencies, and generally small enough to be neglected with little visible distortion.

DCT is used in steganography as-

Image is broken into 8×8 blocks of pixels.

Working from left to right, top to bottom, the DCT is applied to each block.

Each block is compressed through quantization table to scale the DCT coefficients and message is embedded in DCT coefficients.

III LITERATURE SURVEY

A lot of Research has been carried out on Steganography because it is important to know how much data can be concealed without image distortion. Their description is as follows:

Ken Cabeen and Peter Gent [1] have discussed the mathematical equations of Discrete Cosine Transform (DCT) and its uses in image compression. Andrew B. Watson [2] has discussed Discrete Cosine Transform (DCT) technique for converting a signal into elementary frequency component. He developed simple function to compute DCT and show how it is used for image compression. Jessica Fridrich et. al [3] have discussed a reliable and accurate method for detecting least significant bit (LSB) non sequential embedding in digital images. The secret message length is derived by inspecting the lossless capacity in the LSB and shifted LSB plane. Mohesen Ashourian, R.C. Jain and Yo-Sung Ho [4] have proposed a data hiding scheme to embed a signature image in the host image. They selected a gray scale host image of 512×512 pixels and signature image of 256×256 pixels. They developed image data hiding scheme on dithered quantization and a modified baseline JPEG coding scheme. A test of system performance has been done by JPEG compression, addition of Gaussian noise, and Gaussian and Median filtering of host image. J.R.Krenn [5] has proposed a method to embed message in LSB of DC coefficients of cover image. He proposed a simple pseudo-code algorithm to hide a message inside a JPEG image. Ren-Junn Hwang et. al[6] have proposed data hiding based on JPEG technique. They proposed a method of compressing the stego image by lossy compression method to reduce the image size. The receiver then extracts complete data correctly from lossy compressed image.

H. W. Tseng and C. C. Chang *[7]* have proposed a novel high capacity data hiding method based on JPEG. They proposed a method that employs a capacity table to estimate the number of bits that can be hidden in each DCT component so that significant distortions in the Stego-image can be avoided. Youngran Park et. al *[8]* have proposed a new image steganography method to verify whether the secret information had been deleted, forged or changed by attackers. They proposed a method that hides the secret information into special domain of digital image. Neeta Deshpande et. al *[9]* have embedded data in least significant bits of cover image. They explained the LSB embedding technique and presented the evaluation results. Aneesh Jain and Indranil Sengupta *[10]* have proposed a scheme, which hides data in bitmap images, in a way that there is almost no perceptible difference between the original image and new image, and this is also resistant to JPEG compression. M. Chaumont and W. Puech *[11]* have proposed a method to hide the color information in a compressed grey-level image, allow free access to the compressed gray level image, and give color image access only if you own a secret key. KokSheik Wong, Xiaojun Qi, and Kiyoshi Tanaka *[12]* have proposed Mod4 steganography method in discrete cosine transform (DCT) domain. Mod4 is capable of embedding information into both uncompressed and JPEG compressed image. Takayuki Ishida et. al *[13]* have discussed a modified QIM-JPEG2000 steganography which improve the previous JPEG2000 steganography using quantization index modulation (QIM).

IV ALGORITHMS OF STEGANOGRAPHY

A. *Lsb Based Steganography*

Algorithm to embed text message:-

Step 1: Read the cover image and text message which is to be hidden in the cover image.
Step 2: Convert text message in binary.
Step 3: Calculate LSB of each pixels of cover image.
Step 4: Replace LSB of cover image with each bit of secret message one by one.
Step 5: Write stego image

Algorithm to retrieve text message:-

Step 1: Read the stego image.
Step 2: Calculate LSB of each pixels of stego image.
Step 3: Retrieve bits and convert each 8 bit into character.

B. *DCT Based Steganography*

Algorithm to embed text message:-
Step 1: Read cover image.
Step 2: Read secret message and convert it in binary.
Step 3: The cover image is broken into 8×8 block of pixels.
Step 4: Working from left to right, top to bottom subtract 128 in each block of pixels.
Step 5: DCT is applied to each block.
Step 6: Each block is compressed through quantization table.
Step 7: Calculate LSB of each DC coefficient and replace with each bit of secret message.
Step 8: Write stego image.

Algorithm to retrieve text message:-
Step 1: Read stego image.
Step 2: Stego image is broken into 8×8 block of pixels.
Step 3: Working from left to right, top to bottom subtract 128 in each block of pixels.
Step 4: DCT is applied to each block.
Step 5: Each block is compressed through quantization table.
Step 6: Calculate LSB of each DC coefficient.
Step 7: Retrieve and convert each 8 bit into character.

V PERFORMANCE & RESULTS

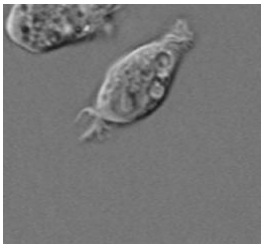
Comparative analysis of LSB based and DCT based steganography has been done on basis of parameters like PSNR. Both grayscale and colored images have been used for experiments. Peak signal to noise ratio is used to compute how well the methods perform.
PSNR computes the peak signal to noise ratio, in decibels, between two images. This ratio is used as a quality measurement between two images. If PSNR ratio is high then images are best of quality.

$$PSNR(x,y) = \frac{10\log_{10}(\max(\max(x),\max(y))^2}{|x-y|^2}$$

A. *LSB Based Steganography*

Fig. II Original Cameraman.bmp **Fig III Stego cameraman.bmp**

PSNR between Fig II and Fig III = 51.0870 dB



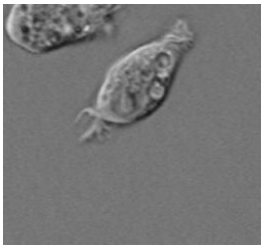
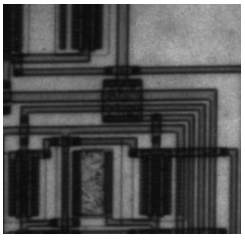


Fig. IV Original cell.bmp **Fig. V Stego cell.bmp**

PSNR between Fig. IV and Fig. V = 49.7214 dB



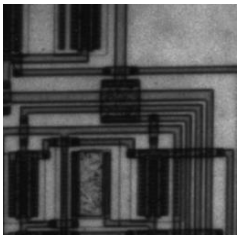


Fig. VI Original circuit.bmp **Fig . VII Stego circuit.bmp**

PSNR between Fig. VI and Fig.VII = 48.3476 dB

i. Using Color Images

Fig .VIII Original army.bmp **Fig. IX Stego army.bmp**

PSNR between Fig VIII and Fig IX =51.0872 dB


Fig .X Original lasercolor.bmp **Fig .XI Stego lasercolor.bmp**

PSNR between Fig X and Fig XI = 51.0881 dB


Fig. XII Original kufte.bmp **Fig. XIII Stego kufte.bmp**

PSNR between Fig XII and Fig XIII = 51.0451 dB

B. DCT Based Steganography

i. Using Grayscale Images


Fig. XIV Original cameraman.bmp **Fig. XV Stego cameraman.bmp**

PSNR between Fig XIII and Fig. XIV = 55.3865 dB


Fig. XVI Original coins.bmp **Fig. XVII Stego coins.bmp**

PSNR between Fig. XVI and Fig. XVII = 55.3049 dB

ii. Using Color Images


Fig. XVIII Original army.bmp **Fig. XIX Stego army.bmp**

PSNR between Fig. XVIII and Fig. XIX = 57.2172 dB


Fig. XX Original ilexvert.bmp **Fig. XXI Stego ilexvert.bmp**

PSNR between Fig. XX and Fig. XXI = 57.0530 dB

VI Conclusion

LSB based steganography embed the text message in LSB of cover image. DCT based steganography embed the text message in LSB of DC coefficients. This paper implements LSB based steganography, DCT based steganography and computes PSNR ratio. PSNR is the peak signal to noise ratio, in decibels, between two images. This ratio is used as a quality measurement between two images. If PSNR ratio is high then images are better of quality. Comparison of LSB based and DCT based stego images using PSNR ratio shows that PSNR ratio of DCT based steganography scheme is high as compared to LSB based steganography scheme for all types of images- (Grayscale as well as Color). DCT based steganography scheme works perfectly with minimal distortion of the image quality as compared to LSB based steganography scheme. Even though the amount of secret data that can be hidden using this technique is very small as compared to LSB based steganography scheme still, DCT based steganography scheme is recommended because of the minimum distortion of image quality.

VII References

1) Ken Cabeen and Peter Gent, "Image Compression and Discrete Cosine Transform", College of Redwoods. *http://online.redwoods.cc.ca.us/instruct/darnold/LAPROJ/Fall98/PKen/dct.pdf*

2) Andrew B. Watson, "Image Compression Using the Discrete Cosine Transform", NASA Ames Research Center , Mathematica Journal, 4(1), p.81-88,1994
3) Jessica Fridrich, Miroslav Goljan, and Rui Du, "Detecting LSB Steganography in Color and Gray-Scale Images", Magazine of IEEE Multimedia, Special Issue on Multimedia and Security, pp.22-28, October-December 2001.
4) Mohesen Ashourian, R.C. Jain and Yo-Sung Ho, *"Dithered Quantization for Image Data Hiding in the DCT domain"*, in proceeding of IST2003, pp.171-175, 16-18 August, 2003 Isfahan Iran.
5) J.R.Krenn, "Steganography and Steganalysis", January 2004.
6) Ren-Junn Hwang, Timothy K. Shih, Chuan-Ho Kao, *"A Lossy Compression Tolerant Data Hiding Method Based on JPEG and VQ."* Journal of Internet Technology Volume 5(2004).
7) Hsien – Wen Tseng and Chin – Chen Chang, " High Capacity Data Hiding in JPEG Compressed Images", Informatica, Volume 15 , Issue 1 (January 2004) 127-142, 2004,0868-4952
8) Youngran Park, Hyunho Kang, Kazuhiko Yamaguchi and Kingo Kobayashi, "Integrity Verification of Secret Information in Image Steganography", Symposium on Information Theory and its Applications, Hakodate, Hokkaido, Japan, 2006.
9) Neeta Deshpande, Kamalapur Sneha, Daisy Jacobs, "Implementation of LSB Steganography and Its Evaluation for various Bits" Digital Information Management, 2006 1st International Conference on. 06/01/2007; DOI: 10.1109/ICDIM.2007.369349
10) Aneesh Jain, Indranil Sen Gupta, "A JPEG Compression Resistant Steganography Scheme for Raster Graphics Images", TENCON 2007 - 2007 IEEE Region 10 Conference, vol.2
11) M. Chaumont and W. Puech, "DCT-Based Data Hiding Method To Embed the Color Information in a JPEG Grey Level Image", 14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP
12) KokSheik Wong, Xiaojun Qi, and Kiyoshi Tanaka, "A DCT based Mod4 Steganography Method" Signal Processing 87, 1251-1263, 2007.
13) Takayuki Ishida, Kazumi Yamawaki, Hideki Noda, Michiharu Niimi, *"Performance Improvement of JPEG2000 Steganography Using QIM"*, Department of System Design and Informatics, Journal of Communication and Computer, ISSN1548-7709, USA, Volume 6, No. 1(Serial No. 50), January 2009.
14) Edward Neuman, "MATLA B Tutorials", Department of Mathematics, Board of Trustees, Southem Illinois University,
15) [15] Rafael C. Gonzalez, Richard E. Woods, "Digital Image Processing", 2nd Edition.

Analysis of Malicious Detection in Bluetooth Enabled Devices Exploiting Wireless Personal Area Networks

GJCST Computing Classification
D.4.6, K.6.5 & C.2.1

M.Latha[1] S.Arockiasamy[2]
1. Asst.Prof, Department of Computer Applications, SNR Sons College, Coimbatore.
Email:lathamurali@yahoo.com
2. Head, Department of Information Systems, University of Nizwa, Sultanate of Oman.
Email : arockia99@gmail.com

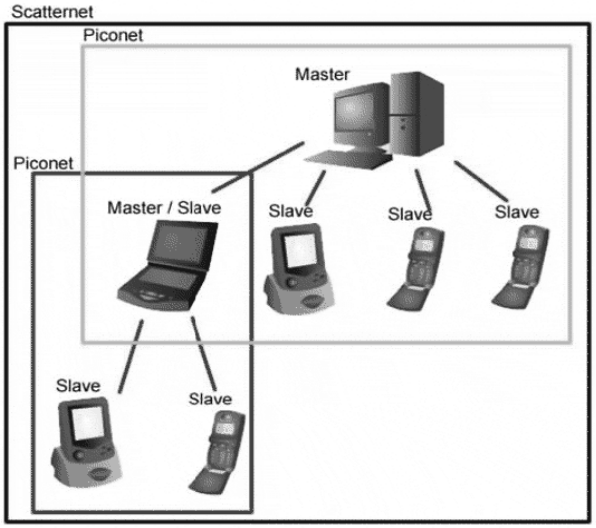
***Abstract*- The growing popularity of mobile devices (smart phones, handsets, PDAs) along with 3G technology brings the mobile internet services on these devices. The wireless devices with messaging capabilities attracted the malware writers to target the hand held devices. Even though the mobile device have numerous benefits like mobility, compact size and ease of their connectivity, the open nature increases the threats and risks being posed. The mobile viruses so far discovered exploited vulnerabilities in Bluetooth by infecting nearby devices and then propagate through SMS to other devices in the mobile network. The problem becomes worse with the growth of MMS (Multimedia Messaging Service), mobile game, and mobile commerce in near future. This paper investigates the propagation of mobile worms and viruses that spread primarily via SMS/MMS messages and short range radio interfaces – Bluetooth.**

Keywords- Bluetooth, Bluetooth Security, Mobile Threats, Radio frequency, and Worms.

I INTRODUCTION

A. *Bluetooth Architecture*

Bluetooth technology is a low cost and low power technology. It is primarily used in short-range radio frequency (RF) communication. It is a protocol used for connecting set of wireless devices, ranging from PDAs, Mobile phones, Notebook computers, Microwave ovens, Refrigerators etc,. The Bluetooth specification was developed by the Bluetooth Special Interest Group (SIG), an industry consortium founded by Ericsson, IBM, Intel, Nokia, and Toshiba. Bluetooth radio operates in the 2.4 GHz unlicensed ISM band (Industrial, Scientific, and Medical). Two or more devices sharing the same channel form a piconet. There is one master device and upto seven active slave devices in a piconet. The
devices can be in any one of the state action, shift, hold, and park. Multiple piconet with overlapping coverage areas form a scatter net. Bluetooth Network – piconet and scatter net is shown in figure1 [8].



The Bluetooth supports both point-to-point connection and point-to-multipoint connection. In point-to-multipoint connection, the channel is shared among several Bluetooth devices. The channel is divided into time slots, each 625µs in length where each slot corresponds to an RF hop frequency. The hop rate is 1600 hops/s, Bluetooth uses frequency hopping for low interference and fading, uses TDD (Time-Division-Duplex) scheme for full duplex transmission and transmits using GFSK (Gausian Frequency Shift Keying) modulation.

The protocol uses a combination of circuit and packet switching. Bluetooth protocol stack can support asynchronous connection-less (ACL) link for data and up to three simultaneous synchronous connection oriented (SCO) links for voice or a combination of asynchronous data and synchronous voice (DV packet type). There are Transport Protocol group, Middleware protocol group and Application group in the protocol stack. The Transport group protocols are used to manage physical and logical links with higher layer protocols and applications and allow Bluetooth device to locate each other. The Radio, Baseband, Link Manager, Logical link control and Adaption (L2CAP) layers and the Host controller and Interfaces (HCI) are included in the transport protocol group. Third party industry standard protocols and Bluetooth SIG developed protocols are included in middleware protocol group. Industry standard protocols include point-to-point, Internet protocols. TCP, WAP and Object Exchange Protocol, RFCOMM, Telephony

control signaling protocol and service discovery protocol comes under Bluetooth SIG developed protocols. The Bluetooth protocol stack is shown in Figure 2 [15]. The IEEE standard 802.15.1 standard for Bluetooth was published in 2002. The IEEE standard defines only the physical (PHY) and MAC (Medium Access Control) layer in its standards.[7]

B. *Device Discovery in Bluetooth*

In Bluetooth environment before a connection is established, device discovery procedure has to be performed before packets start flowing in the wireless links between the master and slave devices and vice versa. When a Bluetooth device wants to find other devices in its vicinity, it broadcast inquiry packets by hopping 3200 times per second along a 32-channel inquiry hopping sequence. A nearby device in the discoverable mode listens on the same frequency sequence but moves

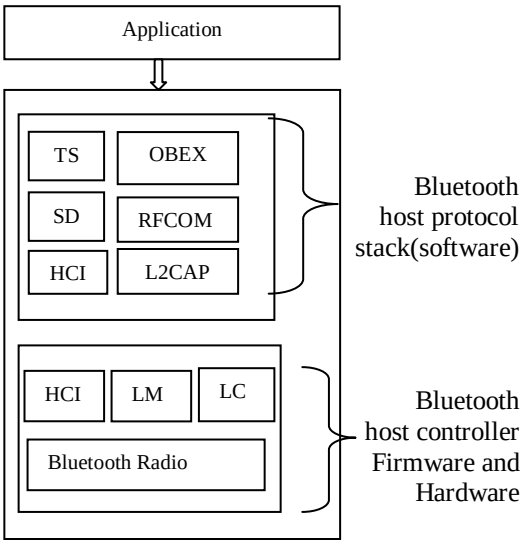


Figure 2: Bluetooth Protocol stack

forward its listening carrier every 1.28 seconds. When a device hears an inquiry packet it backs off for a random period of time and then reenters the scanning state. When it receives another inquiry packet, it responds with a frequency hop synchronization (FHS) packet. On the arrival of this packet, the inquirer device discovers the responders. Once a device has discovered its neighboring devices, it may want to establish a connection with one or more of them. In order to set up a Bluetooth link with a neighbor device, it goes through the paging process. This process is similar to the inquiry process, except that the paging device explicitly specifies the receivers address to indicate which device it wants to set up a connection with. After a connection is established, the pager device and the paged device are called the master and slave of the new link respectively. In the connected state, the master and the slave can exchange normal data packets by hopping 1,600 times per second along a 79-channel frequency sequence decided by the master's local clock and its device address. [12]

C. *Bluetooth Security*

Bluetooth provides security in three ways:

i. It uses pseudo-random frequency hoping to solve the problem of interference from other signals after transmitting or receiving a packet.
ii. Authentication to restrict connectivity to devices. Authentication is initiated when the device is in security mode 2 or in security mode 3.
iii. It uses encryption to employ secret keys, where only authorized users can make data intelligible again.

Frequency –hopping scheme

Bluetooth users use Frequency Hopping Spread Spectrum (FHSS) when transmitting signals. A channel is used for a very short period (e.g. 625μs for data/voice links) followed by a hop marked by a predetermined pseudo-random sequence to another. The frequency-hopping scheme enables the Bluetooth device to avoid interference with other devices. Bluetooth also allows for radio link power control, a low power consumption adaption output scheme, where devices can negotiate and adjust their radio link power consumption relative to the transmitted signal intensity. The combinations of a frequency hopping scheme and radio link power control provide Bluetooth with some additional protection from eavesdropping and malicious access. Spread spectrum transmission are less affected by outside signal interference since any noise interference is likely to affect only a small portion of the signal and not impact the entire signal.

Security Modes

All Bluetooth enabled device implement the Generic Access profile. The profile defines a security model that includes three security modes:

Security Mode 1

i. Mode 1 is as insecure mode of operation. It provides no security.
ii. When a Bluetooth device is in security mode 1, no security procedure is initiated.
iii. Device operating in this mode are able to pair with devices operating in the same mode because neither device implements security controls.

Security Mode 2

i. Mode 2 is known as service level enforced security, provides security at the service level after the channel has been established. This mode enables applications to run in parallel and have different access policies.
ii. When a Bluetooth device is in security mode 2 no security procedure is initiated before a channel establishment request has been received or a channel establishment procedure has been initiated by itself.
iii. Device operating in this mode enforce service level security at the L2CAP layer and above by involving authorization and authentication scheme.

Security Mode 3

i. Mode 3 known as link level enforced security provides security at the link level before the channel is established.
ii. Link encryption is enforced by devices operating in mode 3 at the LMP layer.[10]

II BACKGROUND STUDY

A. *Bluetooth Security Issues*

Bluetooth protocol is a PAN protocol used by devices that communicate wirelessly with one another within 300 feet. Bluetooth is designed to run in a peer-peer short-range wireless network. If the security of Bluetooth is compromised and if one or more devices in the network are used as gateways to other connected network, it could expose the devices or their attached networks. If the network is adhoc there is no access point, there is no centralized mechanism for security administration as there with a WAN where MAC address filtering and other security mechanisms are used to provide protection against rogue access.

Additional to this there are security concerns for Bluetooth mobile phones, where information stored on them such as the addresses and phone numbers of contacts, calendar information and Emails can be stolen. Blue bugging in a hacking technique that access the phone's commands, allowing the hacker to make phone calls, and delete contact information or eavesdrop on the phone owner's conversation. Cell phone worms take advantage of the Bluetooth technology to propagate to other Bluetooth devices.

Everyday more and more viruses and worms have been surfaced on the mobile devices. Bluetooth worms are significantly different from Internet worms in three ways. A Bluetooth enabled device controlled by the worm can only infect neighbors within its radio range. This differs from Internet worms that often scan the entire IP address space for susceptible victims. Second the bandwidth available to Bluetooth devices is much narrower that those of Internet links. Finally the mobility of Bluetooth worm is dynamic when compared with Internet worm. In future the attack on the Bluetooth enabled devices may be severer in the form of handset downtime, service description due to Denial of Service (DoS) attacks, physical damage in device hardware and theft of sensitive data on the device [3][4][5][6][9][11].

B. *Study On Mobile Malwares*

The earliest versions of malicious codes are harmless and they didn't spread from device to device. The recent malicious malwares are capable of spreading to nearby devices via Bluetooth and pose serious threats on enterprise networks. The Table.1 shows the classification of Malwares, its spreading mechanism, target platforms etc.[1][2][13][14][16][17][18].

III RESULTS AND ANALYSIS

This study focuses on the initial investigation of the Bluetooth worm's nature and its characters. The main difference between Blue tooth worm and the Internet worm is the mechanism adapted by the worms to infect the devices. The Blue tooth worm uses proximity scanning process to infect the nearby devices.

A simulation environment is setup with the basic RWP (Random Way Point) mobile model. It is a simple model in which an entity randomly selects a destination in the rectangle area. It moves straight from the current position towards the destination at a uniform speed. When it reaches the destination, it is idle for a period of time, which is again uniformly distributed. Once it enters into active state from idle state, it chooses another destination and the process repeats.

Cabir & Comwarrior are the most popular worms that affects Bluetooth enabled devices.Cabir replicates over Bluetooth connection and install its payload as Symbian System Installation file(SSI).It drains the power of infected phones and starts scanning for next Bluetooth devices for infection. Comwarrior spreads through the messages sent, in which the payload is attached. When it reaches a MMS enabled phones it randomly choose a phone number from the device address book and resets the infected device on the 1st hour of 14th of any month. After infecting a phone, it searches for nearby Bluetooth enabled devices for sending infected files.

The simulation window is set for 3600 seconds. The first worm is chosen randomly for a set of 200 devices and the infection starts at simulation time of 0 seconds. Fig.3 shows the Cabir worms attack ratio. With increase of time (seconds) the graph shows the increase of attack vector created by the Cabir worm. Fig.4 shows the Comwarrior attack ratio. With increase of time (seconds) the graph shows the increase of attack vector created by the Comwarrior worm.Fig.5 shows the average infection caused by the Cabir worm. The infection rate is higher or lower according to the number of Bluetooth enabled devices / applications that are within the affected device vicinity.Fig.6 show the average infection caused by the Comwarrior worm. The infection rate increases with time.

The simulation result shows the impact of Cabir worm and Comwarrior worm on Bluetooth enabled applications.

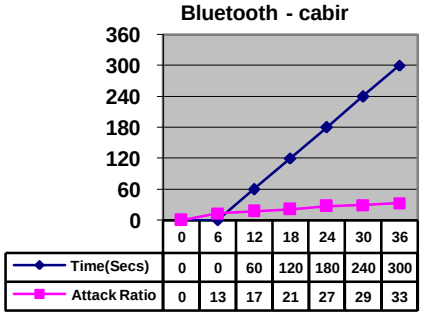


Figure 3. Attack Ratio

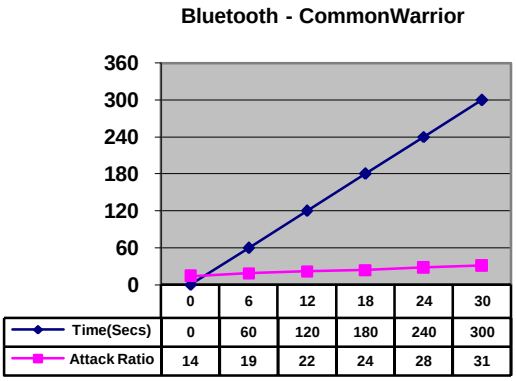


Figure 4. Attack Ratio

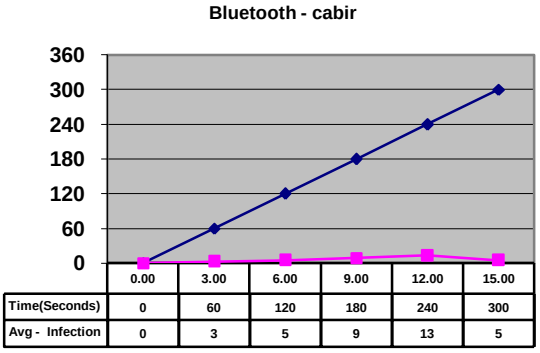


Figure 5. Avg – Infection

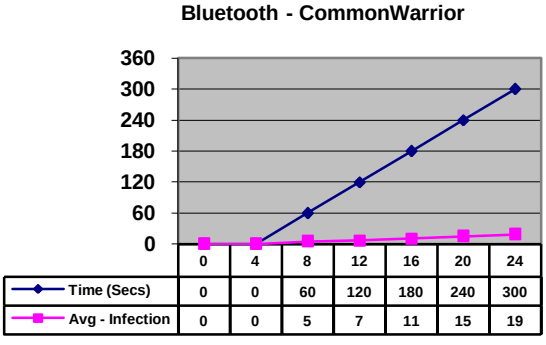


Figure 6. Avg – Infection

Table1. Year wise findings of Malwares in Bluetooth enabled devices

Year	Malware Type	Malware Name	Vulnerabilities
2000 [1]	Virus	Phage	Spreads from PDA to PDA if infected files are shared
2000 [1]	Worm	The Spanish Timophonica	It sends message to random GSM phone numbers via SMS gateway. It modifies MS outlook settings and device registry of an infected device. It also deletes CMOS memory and Master boot records of the device. It propagates through email.
2000 [1]	Worm	The Japanese 110	Affects .NIT DO CO MO i-mode mobile phones. Individual phone numbers in the address book can become victims of Dos attack.
2001 [1]	Virus	Liberty A	It affects hand held devices like palm PDAs, palm OS. To activate this virus it has to be manually installed. It deletes all applications and databases on palm OS-compatible device.
2004 [1]	Worm	Mabir	Spreads by selecting addresses of newly received MMS messages. It transmits MMS messages.
2004 - 2005 [1]	Worm	Cabir	Its variants replicate over Bluetooth connections and install the worm payload as a symbian system Installation (SSI) file.
2005 [1]	Worm / Virus	Lasco	It propagate by transferring its payload to any device in range. It combines the self replication of virus with the self propagation capability of worms.
2005 [1]	Worm	Commwarrior	Propagate by sending messages (along with the payload) to an MMS-enabled phone number randomly chosen from the compromised device's address book and resets the infected device on the first hour of 14th of any month.
2005 [1]	Trojan	Skulls	Propagates by sending both SMS and MMS messages and overwrites address books, e-mail viewer etc.
2005 [1]	Trojan	Drever	Propagates by prompting a user to install an update for symbian OS. Its primary damage is to disable symbian antivirus programs.
2005 [1]	Trojan	Locknut	It propagates like Lasco and overwrites ROM binaries and may crash the OS.
2005 [1]	Trojan	Cardblock	It attacks multimedia cards flash memory of mobile phones.
2005 [1]	Virus	WinCE.Duts	Protect Pc Virus infecting ARM based device. It appends itself to the executable files in the root folder and modifies the program execution leader.
2006 [1]	Virus	Crois over	Spreads from windows desktop PCs to mobile devices running on windows mobile pocket Pc.
2007 [16][17]	Worm	Cabir,I,cabir.H	Spread through specially formatted symbian OS distribution file distinguished as a security management utility. When the infected file, is launched the mobile phone's screen displays the word "velasco".
2008 [16][17]	Worm	Boselo.A	This worm is transmitted thru MMS users are deceived by the extension and unknowingly install the piece, the worm. A targets symbian s60-enabled/malicious device includes Nokia-6600, 6630, 6680, 7610, N70 and N72 handsets.
2009 [16][17]	Worm	Yxes.A	Gathers mobile phone nos and repeatedly send sms. The message feature a web malicious web address (URL) upon clicking of the address in the received message, the recipients will download a copy of the worm.

IV CONCLUSION & FUTURE WORK

The RWP model takes time (seconds) as a control parameter. It takes more time for propagation and less time for infection. The study of mobile virus / worms exploiting Bluetooth and messaging networks needs an enhanced model. Agent based Malware Model (AMM) is the most popular mobility model for analyzing the epidemic growth rate of Bluetooth worms. AMM focuses on the worm that spreads only with the help of human intervention. Further study has to be focused on the propagation speed and infection rate caused by the worms that spreads automatically in the Bluetooth network.

V REFERENCES

1) Abhijit Bose and Kang G. Shin, On Mobile viruses Exploiting messaging and Bluetooth services, IEEE, 2006.
2) Ahonen et al, Information security Threats and solutions in Mobile world, The service Developer perspective 2005, VTT, Espoo, VTT Research notes 2308.
3) Guanhua yan and S.Eidenbeny, "Bluetooth worms, Models, Dynamics, and Defense implications", proc. 22nd Ann. Computer security Applications conf (ACSAC) 2006.
4) Guanhua yan, L.Cuellar, S.Eidenbeng, H.D.Flores, N.Hengartner and V.Vu, "Bluetooth worm Mobility pattern Matters", Proc. Acm symp, Information, computer and communication, security (ASIACCS'07) Mar, 2007.
5) Guanhua yan and S.Eidenbeng, "Modeling propagation Dynamics of Bluetooth worms", Proc.27th IEEE, International conference Distributed computing systems (ICDCS'07) June 2007.
6) Guanhua yan and Stephan Eidenbenz, Modeling propagation dynamics of Bluetooth Worms (Extended version), IEEE, Mar 2009.
7) Guide to Bluetooth security (http://www.bluetooth.com/)
8) Jasonlam, Introduction to Bluetooth & J2ME, contributing writer, 2004
9) Jingsu, A.G.Miklas, K.K.W.Chan, K.PO, A.Akhavan, S.Saroiu, E.d.Lara and A.Goel, "A preliminary Investigation of worm Infections in a Bluetooth Environment", Proc.Fourth ACM workshop Recuring Malcode (WORM) 2006.
10) Marium Jalal Chaudhry, Sadia Murawwat, Farhat Saleemi, Sadaf Tariq, Maria Saleemi, Fatima Jalal chuadgry, "Power optimized secure. Bluetooth communication", IEEE 2008.
11) W.Mickens and B.D.Noble, "Modeling Epidermic Spreading in Mobile Environment", Proc.Fourth ACM Workshop wireless security (Wise'05) sept, 2005.
12) Rajeev sharey, Brent A.Miller, "The Bluetooth Technology: Merits and Limitations", IEEE, 2000.
13) PC QUEST, A CYBERMEDIA publications, January 2008
14) PC QUEST, A CYBERMEDIA publications, , July 2008.
15) Vincenzo Auletta, Carlo Blundo & Emiliano DeCristofaro,A J2ME transparent middleware to support HTTP connections over Bluetooth.
16) http:// www.F-secureu.com./2009
17) http:// www.mobiletor.com
18) http://www.microsoft.com/technet/secutity/altert/info/malware.mspx.

A New Optimized Approach to Face Recognition Using EigenFaces

GJCST Computing Classification I.4.8 & I.5

Sheifali Gupta [1], O.P.Sahoo [2], Ajay Goel[3], Rupesh Gupta[4]
[1] Department of Electronics & Communication Engineering, Singhania University, Rajasthan,India
[2] Department of Electronics & Communication Engineering, N.I.T. Kurukshetra, India
[3] Department of Computer Science & Engineering, Singhania University, Rajasthan, India
[4] Department of Mechanical Engineering, Singhania University, Rajasthan, India
sheifali@yahoo.com , opsahu_reck@yahoo.co.in , goelajay1@gmail.com, rup_esh100@yahoo.co.in

Abstract- **Eigenface approach is one of the simplest and most efficient methods for face recognition. In eigenface approach chosing the threshold, value is a very important factor for performance of face recognition. In addition, the dimensional reduction of face space depends upon number of eigenfaces taken. In this paper, an optimized solution for face recognition is given by taking the optimized value of threshold value and number of eigenfaces. The experimental results show that if the threshold value is 0.8 times of maximum value of minimum Euclidian distances of each image from other images, then maximum recognition rate is achieved. Also only 15% of Eigenfaces with the largest eigen values are sufficient for the recognition of a person. Best optimized solution for face recognition is provided when both the factors are combined i.e. 15% of eigenfaces with largest eigen values are selected and threshold value is chosen 0.8 times maximum of minimum Euclidean distances of each image from all other images, it will greatly improve the recognition performance of a human face up to 97%.**

Keywords-: Eigenface, Face recognition, Euclidean distance.

I INTRODUCTION

The face is our primary focus of attention in social intercourse, playing a major role in conveying identity and emotion [1]. Hence Face recognition has become an important issue in many applications such as security systems, credit card verification and criminal identification [2]. Face Recognition is an emerging field of research with many challenges such as large set of images, improper illuminating conditions [3].

Much of the work in face recognition by computers has focused on detecting individual features such as the eyes, nose, mouth and head outline, and defining a face model by the position, size, and relationships among these features. Such approaches have proven to depend on the precise features [6].

Eigenface approach is one of the simplest and most efficient methods in developing a system for Face Recognition. In eigenface approach, after the dimensional reduction of the face space, the distance is measured between two images for recognition. If the distance is less than some threshold value, then it is considered as a known face, else it is an unknown face [4].

The approach transforms face images into a small set of characteristic feature images, called "eigenfaces", which are the principal components of initial training set of face images. Recognition is performed by projecting a new image into the subspace spanned by the eigenfaces and then classifying face by comparing its position in face space with the position of known individuals [4].

In this paper, section 2 gives the basic idea of Eigenface approach for human face recognition. Section 3 has brief description of steps involved in Eigenface algorithm. Section 4 deals with the threshold concept used in Eigenface approach Section 5 gives the experimental readings & results. Finally, section 6 gives the conclusion drawn from the research

II EIGENFACE APPROACH

Eigenspace-based approaches approximate the face vectors (face images) with lower dimensional feature vectors [8]. The main supposition behind this procedure is that the face space (given by the feature vectors) has a lower dimension than the image space (given by the number of pixels in the image), and that the recognition of the faces can be performed in this reduced space. This approach considers training, where the face database is created, and the projection matrix, the one that achieve the dimensional reduction, is obtained from all the database face images. Also mean face is calculated and the reduced representation of each database image with respect to mean face is calculated. These representations are the ones to be used in the recognition process [5]. An example of Eigenspace face recognition system is as given below. The basic steps involved in Face Recognition using Eigenfaces Approach [4] are as follows:

A. Initialization

i. Acquire initial set of face images known as Training Set.
ii. Calculate eigenfaces from training set keeping only M images that correspond to highest eigenvalues. These M images define the face-space.
iii. Calculate distribution in this M-dimensional space for each known person by projecting his or her face images onto this face-space.

B. *Recognizing New Face Images*

i. For given input image, calculate a set of weights based on M eigenfaces by projecting this new image onto each of eigenfaces.
ii. Determine whether the image is known face or not by checking whether its Euclidean distance with other images is less than some threshold value or not.

III EIGENFACE ALGORITHM

Step 1: Prepare the data
In this step, the faces constituting the training set (Γi) should be prepared for processing.
Step 2: Subtract the mean
The average matrix (Ψ) has to be calculated, then subtracted from the original faces (Γ_i) and the result stored in the variable Φ_i

$$\psi = \frac{1}{M}\sum_{n-1}^{M}\Gamma_n$$

$$\phi_i = \Gamma_i - \psi \qquad (1)$$

Step 3: Calculate the covariance matrix
In the next step the covariance matrix C is calculated according to

$$C = \frac{1}{M}\sum_{n-1}^{M}\phi_n\phi_n^T \qquad (2)$$

Now the eigenvectors (eigenfaces) u_i and the corresponding eigenvalues λ_i of vector C should be calculated.

Step 4: Calculate the eigenvectors and eigenvalues of the covariance matrix
The covariance matrix C in step 3 (see equation 2) has a dimensionality of $N^2 \times N^2$, so one would have N^2 eigenfaces and eigenvalues. For a 256 × 256 image that means that one must compute a 65, 536 × 65, 536 matrix and calculate 65, 536 eigenfaces. Computationally, this is not very efficient as most of those eigenfaces are not useful for our task. In general, PCA is used to describe a large dimensional space with a relative small set of vectors [4]. PCA tells us that since we have only M images, we have only M non-trivial eigenvectors. We can solve for these eigenvectors by taking the eigenvectors of a new M x M matrix:

$$L = A^T A$$

Because of the following math trick:

$$A^T A\, v_i = \mu_i v_i$$
$$A\, A^T A\, v_i = \mu_i\, Av_i$$

Where v_i is an eigenvector of L. From this simple proof we can see that Avi is an eigenvector of C. The M eigenvectors of L are finally used to form the M eigenvectors u_l of C that form our eigenface basis:

$$u_l = \sum_{k=1}^{M} v_{lk}\phi_k$$

Where u are the eigenfaces. Usually, we will use only a subset of M eigenfaces, the M/ eigenfaces with the largest eigenvalues. Eigenfaces with low eigenvalues can be omitted, as they explain only a small part of characteristic features of the faces.

Step 5: Recognizing the faces
The process of recognizing of a new (unknown) face Γ_{new} to one of the known faces proceeds in two steps. First, the new image is transformed into its eigenface components. The resulting weights form the weight vector Ω^T

$$W_k = \mu_k(\Gamma_{new} - \Psi)$$

$$\Omega^T = [w_1 w_2 \ldots\ldots w_{M'}$$

The Euclidean distance between two weight vectors d(i,j) provides a measure of similarity between the corresponding images i and j. If the minimum Euclidean distance between Γ_{new} and other faces exceeds - on average - some threshold value θ, one can assume that Γ_{new} is an unknown face, Else it is considered as a known face.

IV THRESHOLD DECISION

A. *Why Is The Threshold Important?*

Consider for simplicity we have only 10 images in the training set. And an image that is not in the training set comes up for the recognition task. The score for each of the 10 images will be found out with the incoming image. In addition, even if an image is not in the database, it will still say the image is recognized as the training image with which its score is the lowest. Clearly, this is a clash that we need to look at. It is for this purpose that we decide the threshold. The threshold is decided heuristically.

B. *How To Choose Threshold?*

Generally, threshold value is chosen arbitrarily. There is no formula for calculating the threshold value. Its value is chosen arbitrarily or taken as some factor of maximum value of minimum Euclidian distances of each image from other images. In this paper we have calculated what should be the value of threshold?

V EXPERIMENTAL RESULTS

To assess the effect of changing threshold value on the performance of human face recognition, we have performed several experiments on ORL databases using MATLAB. ORL database has images of 40 people, 10 images of each person. So in our database, there are 400 images total.
For the testing, 69 images have taken in test database. In the test database, some faces are from training database but having different face expression. Some faces are unknown faces which do not exist in training database. Some images are non-faces.

For result evaluation, recognition rate is taken as performance metric. Here recognition rate is defined as ratio of successfully recognized test images to the total number of test images. E.g. there are 100 test images out of which it recognizes 60 test images successfully i.e. whether they are known faces or unknown faces, and then recognition rate is 60%.

From the observations and graph, it is clear that if the threshold value is taken as 0.8 times of the maximum value of minimum Euclidian distances of each image from other images i.e. (θ=0.8*max(e)) , recognition rate is more in every observation as compared to the case when threshold is taken as maximum value of minimum Euclidian distances i.e. (θ = max(e)).

No.of Eigenfaces	θ = max(e) Recognition Rate (%)	θ = 0.8 * max(e) Recognition Rate (%)
10	84	94
20	90	96
30	90	97
40	91	97
50	94	97
60	94	97
70	94	97
80	94	97
90	93	97
100	93	97
110	93	97
150	93	97
180	93	97
190	91	96
200	90	96

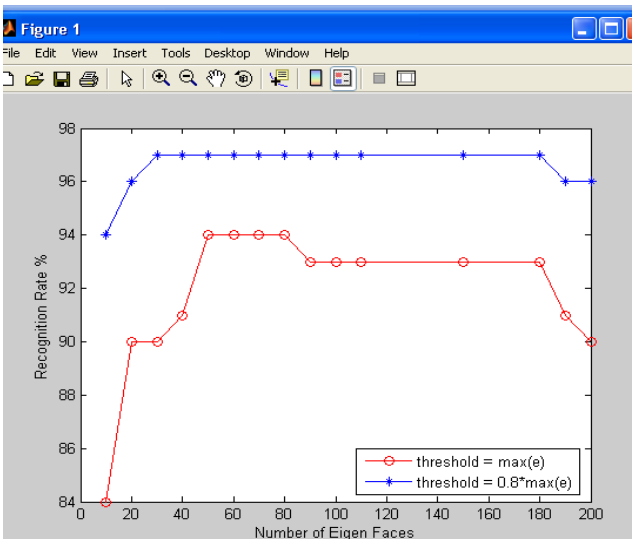


Figure-1

Also it is clear from the observations that as the number of eigen faces are increasing, recognition rate goes on increasing.

VI Conclusion

Highest recognition rate is achieved when 15% of eigenfaces are taken with threshold value equal to 0.8 times of the maximum value of minimum Euclidian distances. A recognition rate of 97% is achieved.

Face recognition has become an important issue in many applications such as security system, credit card verification and criminal identification. For example, the ability to model a particular face and distinguish it from a large set of stored face model would make it possible to vastly improve the criminal identification. Even the ability to merely detect faces, as apposed to recognizing them, can be important.

VII References

1) Zhujie, Y. L. Yu. "Face Recognition with eigenfaces", Proceedings of the IEEE International Conference on Industrial Technology, pp. 434-438., 1994,
2) Jolly D. Shah and S.H. Patil "Biometric Authentication based on Detection and Recognition of Multiple Faces in Image" UFL & JIITU, IC3–2008 , pp87-96
3) Lizama , E.; Waldoestl, D.; Nickolay, B., "An eigenfaces-based automatic face recognition system" IEEE International Conference Systems, Man, and Cybernetics, 1997. Volume 1, Issue 12-15 Oct 1997 Page(s):174 - 177
4) M.Turk and A.Pentland "Face Recognition Using Eigenfaces," Proceedings. IEEE Conference on Computer Vision and Pattern Recognition , pages 586-591, 1991.
5) Pablo Navarrete Javier Ruiz-del-Solar "Analysis and Comparison of Eigenspace-Based Face Recognition Approaches" IJPRAI 16(7): 817-830 (2002)
6) H.K. Ekenel, J. Stallkamp, H. Gao, M. Fischer, R. Stiefelhagen.; "Face Recognition for Smart Interactions", IEEE International Conference on Multimedia & Expo, Beijing, China, July 2007. Page(s):1007 - 1010
7) K.J. Karande, S.N. Talbar "Simplified and modified approach for face recognition using PCA" IET-UK International Conference on Information and Communication Technology in Electrical Sciences (ICTES 2007), Chennai, Tamilnadu, India, page(s): 523-526

Network Intrusion Detection by Applying Various Testing Techniques

Kurundkar G.D[1] Naik N.A.[2]
Dr. Khamitkar S.D.[3] Dr.Kalyankar N.V.[1]

GJCST Computing Classification
B.8, D.4.6 & K.6.5

Abstract- **Intrusion detection is the process of identifying unauthorized usage of a Network. It is important task to protect system from unauthorized users on the network and it is very difficult task to protect system from Intruder. For protecting such network-based system from unauthorized users, we have to implement some prevention measures. With the help of prevention measures, we can give protection to our network. It is necessary to test that prevention measures are working correctly or not, by applying various testing methods on Network. This paper describes methods of Software testing which are applicable for security measures and try to describe approach for on-line attack identification and uses similarity rules for generalization of attack signatures. We can immediately protect the system from many variants of previously known attacks. The architecture uses the comparison of outputs from diverse applications to provide a significant and novel intrusion detection capability. It also initiates attack diagnosis and blocking unauthorized user. The methodology consists of using scripts to generate both background traffic and intrusions with provisions for multiple, interleaved streams of activity.**

Keywords- Intruder; hacker; cracker; Intrusion detection; anomaly detection; verification; validation.

I INTRODUCTION

The ideal goal of any security expert in making a system secure enough to avoid any possible intrusion attempt is to provide the testing methods which are applicable to the known attacks and think about the future attacks which can be harmful to the network. Now a days the Network Protocols, Standalone systems or the single program is much more complex than the previous one. Time required for protecting the system from the intruder is much more than expected. It practically impossible to create a totally secure system. The use of firewall now a days is not enough for protecting your network from the outsiders. In fact, the defense layer itself can contain errors, or it can be bypass with the intelligent internal users.

It is not an intelligent selection to base the whole security process only on prevention techniques. A more reliable process must also take into account how to detect and identify failures. A security process that relies only on prevention without any detection mechanism can be very unstable, since it can create a false sense of protection.

A good prevention is still the first line of any security. Datasets used to test IDS can be described by two main characteristics: the type of intrusion detection technology used (signature-based or anomaly-based) and the location of the IDS (host-based or network-based). Network-based IDS are significantly different from those needed by an anomaly host-based IDS. [1] An attacker can explore a series of potential vulnerabilities until successful, and when successful, he can repeatedly use the exploit that provided success against the same system or other systems with the same vulnerability.[2] Malicious characteristics of computer and network attacks, additional innovative techniques are required. One area of research, which has been applied to this domain by others, is the study of Byzantine faults.[3] The methodology includes techniques and approaches that we adapted from the general field of software testing. Neither the methodology nor the software platform is specific to particular IDS, they can be used to test and compare several different Intrusion Detections. IDS developers can use the platform and methodology to supplement their own approaches to testing. [4] Classification of ID techniques are misuse detection, anomaly detection, and a specification-based approach.[5]. The testing technique is based on an automated mechanism to generate a large number of variations of an exploit by applying mutant operators to an exploit template. The mutant exploits are then run against a victim system where the vulnerable applications and/or operating systems are installed. The attacks are analyzed by a network- based intrusion detection system.[14]

II INTRUDERS

Broadly saying the intruder is an intelligent compuser who is good enough to break into others system or present system to do some unauthorized things. There are many ways of manipulating, illegally updating, or damaging IT Networks and of preparing an attack on IT Network
Intruders are classified into three groups. .[6]

A. Masqueraders

Masqueraders are the persons from the outside who have not right to log into system to use it but they are smart persons

Manuscript received "Thu, Jan 21, 2010 at 2:04 PM GMT"
Mr. G.D.Kurundkar is working as Lecturer with Dept. of Computer Science S.G.B. College Purna
E-Mail: gaju_k_2001@yahoo.com
Mr. N.A.Naik is working as lecturer with Dept. of Computer Science and IT , Yeshwant College Nanded
E-Mail: nitinnaik@live.com
Dr. S.D. Khamitkar is working as Director, School of Computational Sciences S.R.T.M.University , Nanded.
E-Mail: s_khamitkar@yahoo.comDr. N.V.Kalyankar is working as Principal, Yeshwant College Nanded.

who crack the password of the authorized users and then penetrates into system.

B. *Misfeasor*

These users are internal users from the network where they are blocked to gain access of the particular area but as they are also smart compusers they anyhow access the things they want from the system.

C. *Clandestine*

An individual who have supervisory power to control the system and have power to turn off audit control or to suppress audit collections.(It is likely to be outsider/insider.)[7]

III INTRUSION DETECTION MECHANISM

Intruders are unauthorized user and tries to access authorized system. They try to access protected information such as password. Typically, s system maintains a file that associated with a password with each authorized user.

A. *Statistical Anomaly Detection*

It involves the collection of data relating to the behavior of legitimate users over a period of time.

B. *Threshold Detection*

This approach involves defining thresholds independent of users, for the frequency of occurrence of various events.

C. *Profile Based*

A profile of the activity of each user is developed and used to detect changes in the behavior of individual accounts.

D. *Rule Based Detection*

It is a set of rules that can be used to decide that a given behavior is that of an intruder.

E. *Anomaly Detection*

Rules ate developed to detect deviations from previous usage patterns.

F. *Penetration Identification*

An expert system approach that searches for suspicious behavior

G. *Audit Record*

It is a fundamental tool for intrusion detection, some record of ongoing activity by users must be maintained, to an intrusion detection system.

H. *Learning and Detection*

Maintaining log based analysis detecting abnormal activities of unauthorized users. Experimenting with unsupervised learning algorithms and detection mechanisms is made easy. We are trying the K-Means algorithms, as well known clustering algorithm .Log analyzer can handle logs in other formats gathered from different running environments. The sources of audit data can be a keyboard input, command-based logs, application-based logs, and network traffic. According to the type of the audit data collected, we can classify the IDS into two categories [11]

IV TESTING

Testing involves the protection of system from known and unknown attacks. The conditions should include both normal and abnormal conditions. Testing should intentionally attempt to make things go wrong to determine if things happen when they shouldn't or things don't happen when they should. Testing is a process of quality control using a group of defined methods and evaluation criteria, together with guidelines for their use.[8].. An important defect is one that from the customer's perspective affects the usability or functionality of the application. The quality assurance aspect of testing is the degree to which the developers followed. Corporate standard processes or best practices are not responsibility of the testing team. The testing cannot improve quality but trying to find defect in developed application. They can only measure it, although it can be argued that doing things like designing tests before coding begins will improve quality because the coders can then use that information while thinking about their designs and during coding and debugging. Testing has three main purposes: verification, validation, and defect finding..The *verification* process confirms that the software meets its technical specifications..The *validation* process confirms that the software meets the requirements described by user and A *defect* is a variance between the expected and actual result. The defect's ultimate source may be traced to a fault introduced in the specification, design, or development (coding) phase. The difficulty in measuring the detection rate is that the success of IDS is largely dependent upon the set of attacks used during the test. In addition, the probability of detection varies with the false positive rate, and IDS can be configured or tuned to favor the ability either to detect attacks or to minimize false positives [9]
Testers must not only have good development skills—testing often requires a great deal of coding but also be knowledgeable in formal languages, graph theory, and algorithms. Indeed, creative testers have brought many related computing disciplines to bear on testing problems, often with impressive results.[13]

A. V-Model Of Application Testing

V-Model of testing incorporates testing into the entire software as well as application development life cycle. In a diagram of the V-Model, the V proceeds down and then up, from left to right depicting the basic sequence of development and testing activities. The model highlights the existence of different levels of testing and depicts the way each relates to a different development phase. Like any model, the V-Model has detractors and possibly has deficiencies and alternatives but it clearly illustrates that testing can and should start at the very beginning of the assignment. In the requirements gathering stage the business requirements can verify and validate the business case used to justify the assignment. The business requirements are also used to guide the user acceptance testing. The model illustrates how each subsequent phase should verify and validate work done in the previous phase, and how work done during development is used to guide the individual testing phases. This interconnectedness lets us identify important errors, omissions, and other problems before they can do serious harm. Application testing begins with Unit Testing, and in the section titled "Types of Tests" we will discuss each of these test phases in more detail. [10]

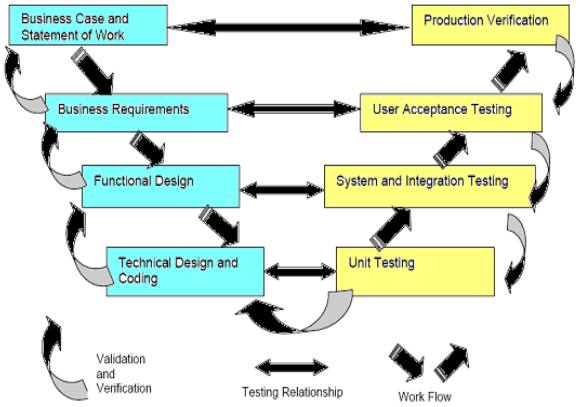


Fig.1.1 Show. The V-model of software/Application

Types of Software/Application Tests :

Phase	Guiding Document	Test Type
Development Phase	Technical Design	Unit Testing
System and Integration Phase	Function design	System testing and integration Testing
User Acceptance Phase	Business Requirements	User Acceptance Testing
Implementation phase	Business Case	Product Verification Testing
Regression Testing applies to all phases		

Table 1.1. The V-Model of testing identifies five software testing phases, each associated with type of test.

B. Various Testing Techniques

i. Unit Testing

A series of stand-alone tests are conducted during Unit Testing. Each test examines an individual component that is new or has been modified. A unit test is also called a module test because it tests the individual units of code that comprise the application. It based on the technical design documents Unit tests focus on functionality and reliability, and the entry and exit criteria can be the same for each module or specific to a particular module. If a defect is discovered during a unit test, the severity of the defect will dictate whether it will be fixed before the module is approved.

ii. System Testing

System testing tests all components and modules that are new, changed, affected by a change, or needed to form the complete application. The system test may require attachment of other systems & it should be minimized as much as possible to reduce the risk of externally induced problems. For example, the system test for a new web interface that collects user input for addition to a database. System testing requires many test runs because it entails feature-by-feature validation of behavior using a wide range of both normal and erroneous test inputs and data. The Test Plan is critical here because it contains descriptions of the test cases, the sequence in which the tests must be executed, and the documentation needed to be collected in each run. When an error or defect is discovered, previously executed system tests must be rerun after the repair is made to make sure that the modifications did not cause other problems.

iii. Integration Testing

Integration testing examines all the components and modules that are new, changed, affected by a change, or needed to form a complete system. Where system testing tries to minimize outside factors, integration testing requires involvement of other systems and interfaces with other applications, including those owned by an outside vendor, external partners, or the customer. Integration testing also differs from system testing in that when a defect is discovered, not all previously executed tests have to be rerun after the repair is made.
Integration testing has a number of sub-types of tests that may or may not be used, depending on the application being tested or expected usage patterns.

iv. Compatibility Testing

Compatibility tests insures that the application works with differently configured systems based on what the users have or may have. When testing a web interface, this means testing for compatibility with different browsers and connection speeds.

v. Performance Testing

Performance tests are used to estimate and understand the application's scalability when, for example, additional users are added or the volume of data increases. This is particularly important for identifying bottlenecks in high usage applications. For data retrieval application, reviewing the performance pattern may show that a change needs to be made in a stored SQL procedure or that an index should be added to the database design.

vi. Stress Testing

Stress Testing is performance testing at higher than normal simulated loads. Stressing runs the system or application beyond the limits of its specified requirements to determine the load under which it fails and how it fails. A continuing performance slow-down leading to a non-catastrophic system halt is the desired result, but if the system will suddenly crash and burn it's important to know the point where that will happen. This test is possibly the most important test for mission-critical systems.

vii. Load Testing

Load tests are the opposite of stress tests. They test the capability of the application to function properly under expected normal production conditions and measure the response times for critical transactions or processes to determine if they are within limits specified in the business requirements and design documents or that they meet Service Level Agreements. For database applications, load testing must be executed on a current production-size database.

viii. User Acceptance Testing

User Acceptance Testing is also called Beta testing, application testing, and end-user testing. Software vendors often make extensive use of Beta testing. Regardless of their best efforts, though, they probably don't find all the flaws in the application.

ix. Product Verification Testing

Production verification testing is a final opportunity to determine if the software is ready for release. Its purpose is to simulate the production cutover as closely as possible and for a period of time simulate real business activity. it should identify anomalies or unexpected changes to existing processes introduced by the new application. The application should be completely removed from the test environment and then completely reinstalled exactly as it will be in the production implementation. Unlike parallel testing in which the old and new systems are run side-by-side, mock processing may not provide accurate data handling results due to limitations of the testing database or the source data.

x. Regression Testing

Regression testing is also known as validation testing and provides a consistent, repeatable validation of each change to an application under development or being modified. Each time a defect is fixed, the potential exists to inadvertently introduce new errors, problems, and defects. Regression testing is the probably selective retesting of an application or system that has been modified to insure that no previously working components, functions, or features fail as a result of the repairs. Regression testing is conducted in parallel with other tests and can be viewed as a quality control tool to ensure that the newly modified code still complies with its specified requirements and that unmodified code has not been affected by the change. It is important to understand that regression testing does not test that a specific defect has been fixed. Regression testing tests that the rest of the application up to the point or repair was not adversely affected by the fix.

xi. Testing Using Disinfected Traffic Or Logs

In this approach, real background traffic is saved and then sanitized to remove any sensitive data. Then, intrusions are injected to the cleaned data. The main advantage of this approach is that tests can be publicly distributed and tests are repeatable. On the other hand, there are some difficulties facing this approach such as cleaning may end by removing needed data and cleaning could fail, thus leading to have some sensitive data disclosed to others. [12]

xii. Testing Using Simulated Traffic

Approach includes background traffic that is generated by complex traffic generators. The advantages of this approach are as follows:

• The data can be publicly distributed, since it does not contain any sensitive data.

• It could be repeatable, since the testing team can replay the same background data or using the simulator, testers can regenerate previous background traffic.

• It could be used to test both: possibility of detection and false-positives.

V CONCLUSION

The paper describes various software as well as application testing Techniques, which will adopt for purpose of testing IDS, We are trying to present the details of the methodology. Software testing is a critical element in the software development life cycle and has the potential to save time and money by identifying problems early and to improve customer satisfaction by delivering a more defect-free product We will perform results from searching various activities of unauthorized behavior patterns by log base experiments on the Network Security Monitor(NSM)for the testing of IDS The platform consist of windows, and in this research paper, we are trying to maintain unexpected

activities of unauthorized users, from regular users of system by maintain log file of person behavior. Without adequate testing, however, there is a greater risk that an application will inadequately deliver what was expected by the business users or that the final product will have problems such that users will eventually abandon it out of frustration. . More formal, rigorous testing will go far to reducing the risk that either of these scenarios occurs.

VI References

1) Frédéric Massicotte, François Gagnon & Yvan Labiche, Automatic Evaluation of Intrusion Detection Systems.
2) James C. Reynolds, James Just, Larry Clough, Ryan Maglich ,*On-Line Intrusion Detection and Attack Prevention Using Diversity, Generate-and-Test, and Generalization*, Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS'03) 0-7695-1874-5/03 2002 IEEE.
3) M. Castro and B. Liskov, "Practical Byzantine Fault Tolerance," Proceedings of the 3rd Symposium on Operating System Design and Implementation, New Orleans LA, February 1999.
4) Nicholas Puketza,Mandy Chung,Ronald A.Olsson & Biswanath Mukharji, *A Software Platform for Testing Intrusion Detection Systems* , May 1997. University of California.
5) Krzysztof M. Brzeziski *On Common Meta-Linguistic Aspects of Intrusion Detection and Testing.* Journal of Information Assurance and Security 2 (2007) 167{178) .
6) Giovanni Vigna, William Robertson & Davide Balzarotti. *Testing Network based Intrusion Detection Signatures Using Mutant Exploits, CCS'04,* October 25–29, 2003, Washington, DC, USA.
7) Williams Stalling *Computer System Security*,3rd Edition.
8) D Hannan, M Ward An alternative approach to software testing to enable SimShip for the localisation market, using Shadow™ BY K Arthur.
9) Peter Mell,Vincent Hu, Richard Lippmann, Josh Haines, Marc Zissman. Richard Lippmann, Josh Haines, Marc Zissman . *An Overview of Issues in Testing Intrusion Detection Systems.* National Institute of Standards and Technology.
10) John E. Bentley, Software Testing Fundamentals Concepts, Roles, and Terminology Wachovia Bank, Charlotte NC, Paper 141-30.
11) A. Mishra, K. Nadkarni, A. Patcha, and V. Tech, "Intrusion Detection in Wireless Ad Hoc Networks", IEEE Wireless Communication, IEEE press, 2004.
12) Hadi Otrok, Joey Paquet, Mourad Debbabi and Prabir Bhattacharya.*Testing Intrusion Detection Systems in MANET*: A Comprehensive Study Fifth Annual Conference on Communication Networks and Services Research(CNSR'07) 0-7695-2835-X/07 2007
13) James A. Whittaker**,** What Is Software Testing? And Why Is It So Hard? Florida Institute of Technology J a n u a r y / F e b r u a r y 2 0 0 0 IEEE SOFTWARE
14) Mr. Nitin A. Naik, Mr. Gajanan D. Kurundkar, Dr. Santosh D. Khamitkar, Dr. Namdeo V. Kalyankar,Penetration Testing: A Roadmap to Network Security, Journal Of Computing, Volume 1, Issue 1, December 2009

Study of Detection of IP Address Conflicts in MANETS

GJCST Computing Classification
C.2.5, C.2.m

S.Zahoor ul Huq[1],Dr. K.E.Sreenivasa Murthy[2],Dr. B.Sathya Narayana[3],D.Kavitha[4]
[1]Computer Science Engineering Department, G.Pulla Reddy Engineering College,
Kurnool , Andhra Pradesh,India
s_zahoor_2000@yahoo.com

***Abstract*- Much effort has been put into the development of routing protocols for route discovery and maintenance for the nodes in a MANET to communicate. The researchers in the routing area assume that all the nodes in the network are already configured to have unique IP Address in the network. Dynamic Address Allocation and Management is a very crucial and difficult problem in MANETs. Since there is no centralized administration and superior authority to guide the address allocation and distribution among the nodes, the IP address auto configuration is to be done by the individual nodes themselves. Research has been going on to avoid the address conflicts in MANETs using less complex methods that reduces communication overhead and best latency. In this paper we present a study of detection of Duplicate Address Conflicts that arise in an ad hoc network.**

Keywords- MANET, address conflicts, IP Addresses

I INTRODUCTION

A Mobile Ad hoc Network (MANET) [1] is formed by the wireless transmitting devices that communicate with each other through wireless channel and without the aid of any fixed or standard infrastructure. The nodes in an ad hoc network themselves acts as routers and cooperate among themselves to achieve communication between any two nodes of the network. The total network consists of simple nodes and the network does not need any centralized administration to guide them how to communicate. The ad hoc networks are used in some important and typical applications such as Military Field Activities, Disaster Situations, Local and Educational Requirements, Wireless Sensor Networks etc.

Routing is a major part of ad hoc communication. The communication between the nodes is done through a single path established and the establishment of the path is called Routing. Most of the researchers concentrated on routing protocols and of course is a major issue to be solved earlier. But the researchers in the routing area take into assumption that the nodes in the network can be uniquely identified using their IP Addresses. This means that the nodes in the network are assigned with a unique IP address each.

Hence most functionalities of the network are completely dependent on the IP addresses of the nodes. It is very much important to see if there are any two nodes in the network with the same IP address. Any nodes in the network with duplicate addresses may cause mal functioning of the network.

In a network, there arise two cases where a duplicate address is possible. Initially it is assumed that the network is initialized with n number of nodes and each node is assigned with a unique IP. Then duplicate addresses arise because of

Node Initialization: when in a network, a new node is initialized; it is to be detected if its self generated IP address matches with any other node in the network and in case, is to be assigned with new available IP, done by the nodes in the network itself.

Node mobility: when a node in a network or a partition of the network moves from its home network to a new network, it is to be monitored if its present IP address matches with that of any other node in the network. If the address already exists, the new node is to be assigned with the different and available IP.

Duplicate Address Detection (DAD) is the methodology introduced for monitoring the repetition of IP addresses by the individual nodes itself. This paper presents the importance of detection of IP address conflicts and different schemes introduced till date for detection.

The rest of the paper is organized as follows: Section II gives the related issues of the duplication of IPs, Section III briefs classification of DAD schemes. Section IV tells about existing methodologies and Section V presents a theoretical comparison of the existing methodologies and Section VI concludes the paper giving the scope for the researchers to concentrate in future.

II RELATED ISSUES

Unavailability of the centralized administration, a MANET requires a unique identifier for each host for reliable communication. Due to mobility, when a new node comes to join the network, it is necessary to see if there is already a node in the network with its IP.

In order to send or receive packets between two nodes, they should possess unique addresses in the network. IP address auto-configuration schemes have to be improved to remove the overhead of manual configuration. Node mobility can cause network partitions. In such partitioned networks, the nodes possess unique addresses independent of the other partitioned networks.

Duplicate addresses may occur in a network because of mobility of the nodes. The nodes under different networks or sub networks may have same IP addresses. This will not affect the functionality of the networks. When a node from

one network moves to another network, address conflicts may occur. Here two cases arise.

Case 1: when only one node moves from one network to another network. Here the mobile node breaks up its links with the older network and will be in contact with the new network only.

Case 2: when a group of nodes move from one network to another network. Here the nodes in the group are interconnected but they break up links with the older network.

Case 3: when an entire network move to merge into another network. This case is often referred to as network merging.

Care should be taken while designing a protocol for IP allocation and Duplicate IP detection for each case. The node resources and network resources are to be concentrated while developing a mechanism. For case 1, methods like simple broadcasting of IP and waiting for reply can be applied. But the same method results more overhead for case 2 and even more for case 3. For case 2, it could be suggested for the methods such as linear IP allocation. For case 3, the broadcasting may not yield better results. Mechanisms such as allocating new network id for the merged network may give better results.

It is also to be concentrated to develop methodologies those less use external equipment (such as GPS).

III CLASSIFICATION OF DAD SCHEMES

The duplicate addresses in the network can be detected using two different mechanisms

i. Leader Action: a leader is elected in the network based on different criteria. The leader is responsible for the detection of duplicates of the addresses in the network. The leader maintains a table of available and free IPs of the network.
ii. Individual Node Action: Here no leader exists. The individual nodes themselves monitor the network for the duplicates of the addresses. If any conflict is detected they themselves solve the conflict through exchange of messages.

In the first mechanism, the leader election process plays a vital role. When the leader fails, a new leader is to be elected. All these processes contribute to more overhead in the network.

In second mechanism, every node is a leader. The nodes monitor the network by exchanging different packets.

The Duplicate Address Detection can be classified based on the nature of the detection as

i. Proactive Duplicate Address Detection

In Proactive Duplicate Address Detection, frequent probing in the network is done for the detection of the duplicate addresses. For this purpose, some dedicated packets are employed to monitor the network.

The advantage of this methodology is that the duplicate addresses in the network can be completely removed. This methodology also got some disadvantage as the number of packet transmissions in the network are large and may lead to more overhead and bandwidth limitations. This methodology may use either the Leader Action mechanism or the Individual Node Action mechanism.

ii. Reactive Duplicate Address Detection

Here the duplicate addresses are detected only when some network action is performed. No separate packets are dedicated for the detection of the duplicate addresses. Routing is the basic functionality of MANETs. Using routing packets itself, any duplicates of the addresses are detected.

The main advantage of this methodology is that additional overhead is avoided for the detection of the duplicate addresses. For using this methodology, care should be taken as different cases may arise which cause false detection. The disadvantage of this methodology is that the duplicate addresses are detected only at the time of routing. This methodologies use only Individual Node Action mechanism.

The DAD schemes are again classified based on the accuracy of the detection, as

i. Strong Duplicate Address Detection Schemes (SDAD)

SDAD schemes use either Leader Action or Individual Node Action. These schemes use the methodology of Proactive Duplicate Address Detection. They probe the network for the duplicates of the addresses. These schemes maintain greater accuracy in detecting the duplicates of IP addresses. But large Overhead is observed in these schemes.

ii. Weak Duplicate Address Detection Schemes (WDAD

These schemes detect the duplicates less accurately. These schemes provide lesser overhead compared to SDAD schemes. These schemes utilize either Leader Action or Individual Node Action mechanisms. Both Proactive Duplicate Address Detection and Reactive Duplicate Address Detection methodologies can be employed in these schemes.

The DAD schemes are also classified based on the scope of detection as

i. Active Duplicate Address Detection Schemes (ADAD)

The ADAD schemes detect the winner and looser along with the detection of the duplicate addresses.

ii. Passive Duplicate Address Detection Schemes (PDAD)

The PDAD schemes detect only the duplicates in the network. The PDAD schemes are not concerned with the winner and loser.

IV EXISTING METHODOLOGIES

Strong Duplicate Address Detection Schemes were proposed in [2]. In this mechanism, the nodes generate their own IP and probe in the network for the repetition of the IP. If a reply is received, new IP is generated and the process is repeated.

Perkins et al [3] have proposed a simple Duplicate Address Detection Schemes where the nodes choose a random address and send a request to the address. When no reply is received the address is fixed as the permanent address. This method got some limitations as the probability of the number of repetitions of the process of generating the new address and probing, is not clear. When two networks merge, the process proposed could yield a high overhead and may malfunction because of bandwidth limitations.

Vaidya's proposal [4] was aimed at the packet delivery to the correct node even if two nodes are with same address. Strong Duplicate Address Detection is not possible in this scheme. The proposal requires the modification of existing routing protocols to implement this scheme.

In [5], the proposed scheme used a leader to identify the group, and the nodes joining the network are assigned with the sequential addresses, with the newest member taking over the charge as leader. Each node periodically sends an update beacon message to the nodes with the next and previous addresses so that the node looses can be detected. Any node that becomes inactive for a particular period should acquire new IP.

Prophet address allocation Scheme [6] uses a mechanism similar to that of [5]. The first node that initialized in the network acts as the prophet and it allocates IP address to the new nodes that join the network. The presence of Duplicate Addresses is detected by the prophet. But this mechanism requires a super node (called as Prophet) to monitor the network. This method limits that the super node got the additional responsibilities and may die out quickly because of battery depletion. Leader election process is to be followed for the newer prophet.

In [7], five different schemes were introduced which detect the duplicates of the addresses in the network using only the routing messages. The schemes use two types of information such as Location of the nodes and the Neighbor List of the nodes. The main advantage of these schemes is that they use no other probing messages for the detection of duplicates of the addresses in the network. In this paper, the authors didn't mention any mechanism through which winner and looser are detected and how new IP is assigned for the looser. Another limitation of the schemes is that the duplicate addresses are not detected proactively. This may lead to the presence of the duplicate addresses in the network. And more over, the schemes require additional facilities such as GPS.

In [8], schemes for duplicate address detection in on-demand routing protocols are presented. These schemes use no other control messages for the detection of the duplicate addresses. Hence a very low overhead is achieved. The routing messages such as RREQ and RREP are used for the detection of duplicates of the addresses in the network. However, the accuracy of detection is doubtful. These schemes may lead to false detections and which may result in the mal functioning of the nodes in the network.

V THEORETICAL ANALYSIS

A theoretical analysis of the Classifications of DAD Schemes is presented in Table 1.

The performance metrics of any DAD schemes are Accuracy, Detection Ratio, Overhead, and the load on any single node.

The Accuracy shows how accurately the duplicate addresses are detected avoiding the false detections. The Detection Ratio can be defined, as the ratio of total number of duplicate addresses detected in the network to the number of duplicate addresses actually exists.

Overhead is defined as the number of control packets needed in the network to the number of data packets that transmit in the network. Overhead represents both the node resources and network resources. If the overhead is more, the numbers of packets that transmit in the network are more and this leads to the quicker depletion of the battery of the nodes and limit the bandwidth.

The load on any single node is also should not be encouraged. This may lead to the failure of the node and for the election process, overhead may be incorporated. The failure of a node is disadvantageous and may even cause network partitioning.

TABLE I
Analysis of Classifications of DAD Schemes

Classification	Accuracy	Detection Ratio	Overhead	Load on any single node
Leader Action	--	--	--	high
Individual Node Action	--	--	--	Low
Proactive DAD	High	High	High	--
Reactive DAD	Low	low	Low	Low
SDAD	High	--	High	--
WDAD	Low	--	--	--

VI Conclusion

The major research efforts have been put to develop effective routing protocols. But the routing protocols function to the best levels if there are no address conflicts in the network. Pre-configuration is not always possible and has some drawbacks. More over a centralized control does not exist in ad hoc networks. So it is very much important to investigate mechanisms for address auto-configuration in MANETs. For monitoring the address assignments it is very much important to detect the address conflicts. A standardized mechanism has to be developed to overcome the address conflicts in the network.

To design and develop mechanism that detects the address conflicts in the network, different metrics are to be taken into consideration. In this paper we presented different DAD schemes to quickly and accurately detect address conflicts in an ad hoc network. The DAD schemes can be classified on different bases. The different classifications and the advantages and disadvantages of the classified schemes are also presented. Also the different areas that are to be concentrated while designing a DAD scheme are discussed.

VII References

1) Mobile Ad-hoc Networks (MANET), *"www.ietf.org/html.charters/manet-charter.html"*
2) C.E. Perkins, E. M. Belding-Royer, S. R. Das, "IP Address Auto-Configuration for Ad Hoc Networks", Mobile Ad Hoc Networking Working Group- Internet Draft, January 2001.
3) Perkins et al, "IP Address Autoconfiguration for Ad Hoc Networks" – Internet Draft(Nov. 2001)
4) Nitin Vaidya "Weak Duplicate Address Detection in Mobile Ad Hoc Networks", ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc), June 2002.
5) Nakjung Choi, C.K.Toh, Yongho Seok, Dongkyun Kim, Yanghee Choi, " Random and Linear Address allocation for Mobile Ad hoc Networks", IEEE Communications Society, WCNC 2005.
6) Hongbo Zhou, Lionel M. Ni, Matt W. Mutka, " Prophet address allocation for large scale MANETs", www.elsevier.com/locate/adhoc, 423-434, 2003.
7) Dongkyun Kim, Hong-Jong Jeong, C.K. Toh, Sutaek Oh, " Passive Duplicate Address Detection Schemes for On-demand Routing Protocols in mobile Ad hoc Networks", IEEE Transactions on Vehicular Technology, (To Appear 2009).
8) K. Weniger, "PACMAN: Passive Auto Configuration for Mobile Ad hoc Networks", IEEE Journal of Selected areas in communication , vol 23, No. 3, March 2005.

Evaluation Criteria for Routing In Mobile Ad Hoc Networks

GJCST Computing Classification
C.2.5, C.2.m

[1]P. Sai Kiran, [2]V. Krishna Reddy, [3]Dr. L.S.S Reddy
[1,2] Associate Professor, Department of Computer Science and Engineering,
Lakireddy Balireddy College of Engineering, Mylavaram, Andhra Pradesh
[3]Director, Lakireddy Balireddy College of Engineering, Mylavaram, Andhra Pradesh
[1]psaikiran@hotmail.com, [2]Krishna4474@yahoo.com, [3]director@lbrce.ac.in

Abstract- **Mobile Ad-hoc network typically have a dynamic topology, which will have profound effects on network characteristics. Network functions such as routing, address allocation, authentication, and authorization must be designed to cope with a dynamic and volatile network topology. Routing is a core problem in networks for delivering data from one node to another. Many routing algorithms have been proposed for MANET that belongs to different categories and with different criteria to improve the performance while reducing the overhead. In this paper we would like to exploit various characteristics and review those characteristics and their effect on performance of the proposed routing methods. In this paper, we have also tried to identify the issues that are to be considered while evaluating a routing algorithm for mobile ad hoc networks.**

I INTRODUCTION

According to the definition of IEEE 802.11: A network composed solely of stations within mutual communication range of each other via the wireless medium (WM). An ad hoc network is typically created in a spontaneous manner. The principal distinguishing characteristic of an ad hoc network is its limited temporal and spatial extent. These limitations allow the act of creating and dissolving the ad hoc network to be sufficiently straightforward and convenient to be achievable by non-technical users of the network facilities. No specialized "technical skills" are required and little or no investment of time or additional resources is required beyond the stations that are to participate in the ad hoc network. The term ad hoc is often used as slang to refer to an independent basic service set (IBSS)[13].

In this paper we would like to exploit the requirements for routing protocols and survey the different routing strategies for Mobile Ad Hoc Networks. Section II categorizes the various routing strategies for MANET while Section III describes the issues for evaluating the routing protocols proposed for the Mobile Ad Hoc networks.

II CHARACERISTICS REVIEW OF CURRENT ROUTING ALGORITHMS

We shall now review the main characteristics of proposed routing algorithms, in light of desired qualitative and quantitative properties, and few additional characteristics.

A. *Demand-Based Operation*

Routing algorithms can be classified as proactive or reactive. Proactive protocols maintain routing tables when nodes move, independently of traffic demand, and thus may Have unacceptable overhead when data traffic is considerably lower than mobility rate. The communication overhead involved in maintaining global information about the networks is not acceptable for networks whose bandwidth are battery power are severely limited.

Reactive algorithms-

Reactive algorithms are designing routes when they are needed, in order to minimize the communication overhead. They are adaptive to „sleep period" operation, since inactive nodes simply do not participate at the time the route is established. When requires the destination search will be initiated and the route will be computed for data transmission. The efficiency of destination search depends on the corresponding location update scheme. A quorum based, a home agent based, and a depth-first search based destination search and corresponding location update schemes are being developed. Other location update and destination search schemes may be used including an occasional flooding. In reactive routing, the communication overhead of routing algorithm is divided into the following components: location updates, destination searches (that are performed in accordance to location update scheme), and path creation (or reporting from destination back to source).

B. *Distributed Operation*

We shall divide all distributed routing algorithms into localized and non-localized. Localized algorithms are distributed algorithms that resemble greedy algorithms, where simple local behavior achieves a desired global objective. In a localized routing algorithm each node makes decision to which neighbor to forward the message based solely on the location of itself, its neighboring nodes, and destination. While neighboring nodes may update each other location whenever an edge is broken or created, the accuracy of destination location is a serious problem. Localized routing algorithms that guaranty delivery show that localized algorithms can nearly match the performance of shortest path algorithms.

All non-localized routing algorithms proposed in literature are variations of shortest weighted path algorithm. Zone based approaches, combining shortest paths within a zone and inter-zonal destination searches or routing tables.

In zone based routing algorithm, nodes are divided into non-overlapping zones. One way of forming Zone is to use the location information to form the zones and operate based on location information. Another way is not to use location information of nodes but selects one node in each grid or zone, and these nodes serve as backbone for routing tasks. Each node only knows node connectivity within its own zone, and routing within zone is performed directly while inter-zone would be performed by using backbone node.

C. *Location Information*

Most proposed routing algorithms do not use the location of nodes, that is their coordinates in two or three-dimensional space, in routing decisions. The distance between neighboring nodes can be estimated based on incoming signal strengths (if some control messages are sent using fixed power). Relative coordinates of neighboring nodes can be obtained by exchanging such information between neighbors.
Alternatively, the location of nodes may be available directly by communicating with a satellite, using GPS (Global Positioning System), if nodes are equipped with a small low power GPS receiver. We believe that the advantages of using location information outweigh the cost of additional hardware, if any. The distance information, for instance, allows nodes to adjust their transmission powers and reduce transmission power accordingly. This enables using power, cost, and power-cost metrics and corresponding routing algorithms in order to minimize energy required per routing task, and to maximize the number of routing tasks that a network can perform. Routing tables that are updated by mobile software agents modeled on ants. Ants collect and disseminate location information about nodes.

D. *Single-Path Vs. Multi-Path Strategies*

There exist several multi-path full message strategies, where each node on the path sends full message to several neighbors which are best choices for all possible destination positions. There is significant communication overhead, and lack of guaranteed delivery can make this approach inferior to even a simple flooding algorithm. Clever flooding algorithm may use about half of nodes only for retransmissions, which often matches the number of nodes participating in routing in this method. In addition, flooding guarantees delivery and requires no prior location updates for improved efficiency. Multi-path methods may be regarded as flooding that is restricted to the request zone, and as such can be used for geocasting (where a message is to be delivered to all nodes located within a region). Multi-path algorithm consisting of several single-paths is proposed.

E. *Loop-Freedom*

Interestingly, this basic criterion was neglected in many papers. GEDIR and MFR algorithms are inherently loop-free. The proofs are based on the observation that distances of nodes toward destination are decreasing.

F. *Memorization Of Past Traffic*

Most reported algorithms require some or all nodes to memorize past traffic, as part of current routing protocol, or to memorize previous best path for providing future path to the same destination. Solutions that require nodes to memorize route or particular information about past traffic are sensitive to node queue size, changes in node activity and node mobility while routing is ongoing. One form of such memorization are routing tables, which memorize last successful path to each destination.

III PERFORMANCE EVALUATION ISSUES

The important issues for evaluating the routing protocols for MANET are as follows.

A. *Delivery Rate*

Delivery rate is defined as the ratio of numbers of messages received by destination and sent by senders. The best methods by this metric are those that guarantee delivery, where message delivery is guaranteed assuming „reasonably" accurate destination and neighbor's location and no message collisions.

B. *End-To-End Data Delay*

This is also referred to as latency, and is the time needed to deliver the message. Data delay can be divided into queuing delay and propagation delay. If queuing delay is ignored, propagation delay can be replaced by hop count, because of proportionality. Retransmissions can be included if MAC (medium access control) layer is used in experiments. Several papers suggested that it is more important to minimize the power needed per message, or the number of routing tasks network can perform before partitioning.

C. *Communication Overhead*

It can be defined as the average number of control and data bits transmitted per data bits delivered. Control bits include the cost of location updates as preparation step, destination searches, and retransmission during routing process. However, this metric is rarely used in literature. In fact, most of proposed papers avoid measuring it altogether. The portion of ignored overhead may often be more significant than the measured one.

D. *Performance On Static Networks*

Although the algorithm is designed with moving nodes in mind, static nodes are important special case to be verified. Some networks, such as sensor networks, are static most of time, and sometimes destination and neighbors information is accurate.

E. MAC Layer Considerations

While initial experiments may ignore data link layer, for similar reasons (in the absence of message collisions, routing algorithms should have superb performance, e.g. to guaranty delivery), further experiments, even on static networks, should consider it. IEEE 802.11 is a standard for MAC specifications in wireless networks.

F. Comparison With The Shortest Path Algorithm

There is notable tendency in literature to compare the performance of proposed routing algorithms with the worst possible solution, flooding. Even such comparison is not properly done, since improper flooding algorithms are used. If flooding is taken for comparison then the proper version of it should be used. Although some existing algorithm can also be taken for comparison, especially if it belongs to the same class with classification criteria, the ideal shortest path algorithm is certainly the ultimate goal, and one should verify how far from that goal the proposed algorithm is. If the cost of location updates for both proposed and shortest path algorithms is ignored, flooding rate (the ratio of the number of message transmissions and the shortest possible hop count between two nodes) can be used for fair comparison, especially for multi-path methods. Each transmission in multiple routes is counted, and message can be sent to all neighbors with one transmission.

G. Generating Sparse And Dense Graphs

For experiments with static networks, random unit graphs should be generated. Each of nodes should select at random x and y coordinates. Sub-graphs can be used if obstacles are taken into account. The connectivity depends on the selected transmission radius. Since transmission radius for a given equipment is normally fixed or should be selected from few discrete values, most papers use fixed value of transmission radius and change the range of coordinates to evaluate graphs of different density. Ignoring graph density issue in performance evaluations is a single misleading point in the experimental design and interpretation of results.
Routing algorithms perform differently on sparse and dense graphs, thus it is the graph density that is a primary independent variable to be considered. The best measure of graph density is the average number of neighbors for each node. Generated graphs, which are disconnected, may or may not be eliminated.

H. Node Mobility

Some papers use random movements at each simulation step in four or eight possible direction. Random walks tend to keep all nodes close to their initial positions, and thus analysis using this model is largely misleading. One possible analogous design is as follows. Each node generates a random number wait in interval [0..maxwait]. The node does not move for wait seconds. This is called station time. When this time expires, node chooses to move with a probability p. It generates new wait period if it decides not to move. Otherwise, it generates a random number travel in interval [0,maxtravel], and a new random position within the same square, in the second case. Node then moves from old position to new position along the line segment joining them at equal speed for the duration of travel seconds. Upon arriving at new location, node again chooses waiting period etc. This movement patterns do not cover the case of nodes moving more or less in the same direction, which may often be the case in military and rescue operations. An additional component should be added in experiments, moving with same speed and in same direction by all nodes.

i. Simulator

Several wireless networks simulators are used in literature. Two most widely used are Glomosym [Glom] and ns-2[ns-2]. Although it is desirable to have some kind of benchmark testing facility, the problem with these simulators is a painful learning curve. Several researchers that used it confirmed that it takes about one month of full time work to learn how to use these simulators. Thus they are convenient choice for long term projects (and long term grant holders), but not for researchers with limited human resources. The other drawback of using these simulators is that experiments with static nodes and important parameters (e.g. graph density) are easily ignored. Preliminary experiments with static nodes and even moving nodes can be obtained by a simplified design using any programming language (e.g. C or Java) and valuable conclusions can be made. This shall be done even if simulator is used afterwards. We agree, of course, that real simulations are necessary for a complete performance evaluation, if resources for doing that are available.

IV CONCLUSION

Routing is a core problem in networks for delivering data from one node to another. Many routing algorithms have been proposed for MANET that belongs to different categories and with different criteria to improve the performance while reducing the overhead. In this paper we have exploited various characteristics that can be incorporated in routing algorithms and the way the characteristics effect the performance of routing in MANETs. Evaluation of routing protocols should not be limited to a particular issue like reduced overhead or increase in throughput as a particular routing protocol strategy would affect other performance factors but should be extended to include the mobility factors, MAC layer considerations etc. Thus in this paper we have tried to exploit all the issues to be evaluated for a Routing protocol.

V REFERENCES

1) David B. Johnson, "Routing in Ad Hoc Networks of Mobile Hosts", Proceedings of the IEEE workshop on Mobile Computing Systems and Applications, December 1994
2) M. Royer, and C-K. Toh, "A Review of Current Routing Protocols for Ad-Hoc Mobile Wireless Networks", IEEE Personal Communications, April 1999, pp. 46-55
3) Josh Broch, David A. Maltz, David B. Johnson, et al. "A Performance Comparison of Multi-Hop Wireless Ad Hod Network Routing Protocols", MOBICOM"98, October 1998
4) Zygmunt J. Haas, Marc R. Pearlman, and Prince Samar, "The Zone Routing Protocol (ZRP) for Ad Hoc Networks", draft-ietf-manet-zone-zrp-04.txt, July, 2002
5) Young-Bae Ko, and Nitin H. Vaidya, "Location-Aided Routing (LAR) in Mobile Ad Hoc Networks", MOBICOM"98, 1998
6) R. Jain., A. Puri, and R. Sengupta, "Geographical Routing Using Partial Information for Wireless Ad Hoc Networks", IEEE Personal Communication, February 2001, Vol 8, No 1, pp 48-57
7) Z. J. Haas and B. Liang, "Ad Hoc Mobility Management with Uniform Quorum Systems", IEEE/ACM Trans. Net., vol. 7, no. 2, Apr. 1999, pp. 228-40
8) Martin Mauve, Jörg Widmer, and Hannes Hartenstein, "A Survey on Position-Based Routing in Mobile Ad Hoc Networks", IEEE Network, Nov/Dec 2001, pp 30-39
9) S-J. Lee, M. Gerla, and C-C. Chiang, "On-Demand Multicast Routing Protocol", Proceedings of IEEE WCNC'99, New Orleans, LA, Sep. 1999
10) S-J. Lee, W. Su, J. Hsu, et al., "A Performance Comparison Study of Ad Hoc Wireless Multicast protocols", Proceedings of IEEE INFOCOM 2000, Tel Aviv, Israel, March 2000, pp. 565-574.
11) Rrank Kargl, Jūrgen Nagler, and Stefan Schlott, "Building a Framework for MANET Routing Protocols", July 2002
12) Xukai Zou, Byrav Ramamurthy and Spyros Magliveras. "Routing Techniques in Wireless Ad Hoc Networks Classification and Comparison", Proceedings of the Sixth World Multiconference on Systemics, Cybernetics and Informatics, SCI, July 2002.
13) IEEE802.11 working group, Wireless LAN Medium Access Control(MAC) and Physical Layer (PHY) specifications, 1999

A Framework for Creating Global Schema Using Global Views from Distributed Heterogeneous Relational Databases in Multidatabase System

GJCST Computing Classification
C.2.4, H.2.1 & H.2.8

Ali Mohammad Ghulam
Post Graduate Studies and Research
Indian Institute of Technology Kharagpur, West Bengal 721302, India
ali_iit@yahoo.com , ali@hijli.iitkgp.ernet.in
Murtuza Mohammad Ghulam
mgm_rsc@yahoo.co.uk

***Abstract*- Creating a global view, this is an alternative way of looking at the data in one or more tables from different geographical located heterogeneous local component databases with better security. A view is simply a stored query accessible as a virtual table that can be used similar to a table. In this paper, we propose a framework to integrate heterogeneous relational local component databases using Global View and construct a Global Schema with a set of Global Views in multidatabase system. We have also illustrated the method of integration with an example. Query submitted on Global View will be decomposed into a set of sub queries and will be executed at various remote sites and finally sub-queries will be composed and will get back result to concerned user. We also discuss some issues concerning distributed databases security in the Multidatabase System.**

Keywords- Multidatabase, Global Schema, Relational Database, Database View.

I INTRODUCTION

Keeping in mind the progress in communication and database technologies (concurrency, consistency, and reliability) has increased the data processing potential. Various protocols are proposed and implemented for network reliability, concurrency, atomicity, consistency, recovery and replication. The current demand is to access data from various existing databases distributed among sites in a network. If any organization has headquarter in any country and has many branches across the globe, wants efficient and quick retrieval of information for any kind of decision supports, the proposed framework will meet this requirement in this paper. This paper has addressed this issue nicely by implementing highly coupled federated multidatabase system.

The multidatabase system has gaining attention of many researchers that attempts to logically integrate several different independent distributed DBMSs while allowing the local DBMSs to maintain complete control of their operations. It means all existing databases are autonomous and evolve over times. In multidatabase system, it should be possible to address data in more than one database by a single query. On the other hand, it should be possible for different users to have different interpretation of the same data. Thus, the demand on a multidatabase is to provide an interpretation of data with same or similar meaning which have different representations. A multidatabase system is a database system that resides on top of existing component local database systems and presents a single database illustration to its users [1,2]. The Multidatabase System usually maintains a single global database schema, which is integration of all local component databases schemas and against which users issue queries and updates. Multidatabase System maintains only global schema and the local component database system actually maintains all user data. Creating and maintaining the global schema, which requires the use of database integration techniques, is a critical issue in the multidatabase system.

Variety of approaches to schema integration have been proposed e.g. [3,4,5,6,7,8,9,15]. We propose a framework to integrate multiple local component relational databases using database global view technique and query on global view will be translated to number of queries on physical remote distributed relational databases. However, effort for constructing the global view may be expensive with respect to the frequency of usage.

II DATABASE ARCHITECTURAL FRAMEWORK

The ANSI/X3/SPARC architecture [10,13,14] as shown in figure 1 is claimed to be based on the data organization in DBMS standardization. It recognizes three views of data:

A. Local internal schema/view

Local (internal) schema/view shows how the data is stored on each site. The format of the internal schema is dependent on the LDBMS of each site.

B. A global conceptual schema/view

A global (conceptual) schema describes the data throughout a network and shows what data is at what site. The global schema usually stored in a global directory.

C. A User external schema/view

A user (external) schema/view shows how user will view and manipulate the data.

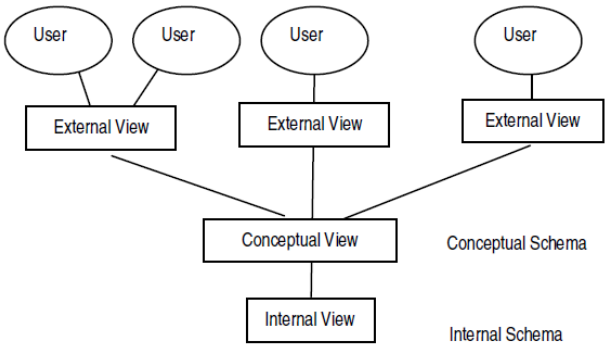


Figure 1. ANSI/SPARC Architecture

III PROPOSED FRAMEWORK

In our proposed framework, the Multidatabase system will control multiple gateways and will access to remote databases through these gateways. The proposed multidatabase system manages and retrieves data from multiple sites within a single graphical user interface based application or web-based application and resides on top of the Global Schema while providing complete autonomy to individual remote database systems.

The proposed framework is divided into four layers (as shown in Figure 2) based on a classical example (as shown in Figure 1) of a data-based architecture is the ANSI/SPARC model by Tsichritzis and Klug [10,13,14]. See the details work of the Multidatabase System as 4-tiered client-server model in the distributed databases [16].

Layer 1- Users are sitting on layer 1 that is top of the Multidatabase System and will submit queries and updates though a single multidatabase system application of the Global Multidatabase System to remote database servers. A single query can retrieve information from many remote sites table and can update many remote sites table through global views.

Layer 2- Layer 2 is an Application Server where a web based application or client-server application resides. If the application is a web-based then user will access application through specific URL and will submit queries and updates. If the application is a client-server, then the application will be launched through citrix metaframe under thin-client platform. Layer 2 contains business logic, API and access to Multidatabase system server.

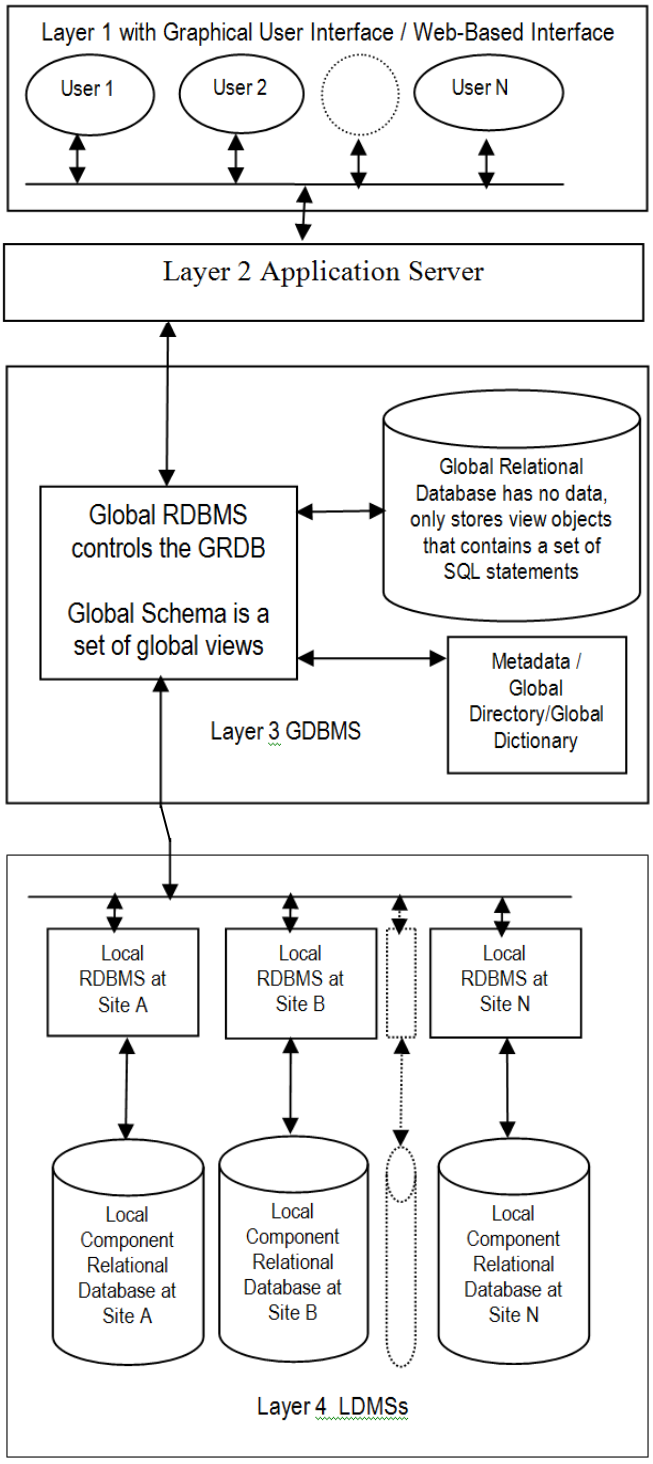


Figure 2 Proposed Framework

Layer 3- In layer 3, there is a Global Database Management System (GDBMS) where Global Schema is created with a set of global views and is stored in a Global Database (GDB). User will submit a query to Global Schema specific to any global view with the help of GUI or web-based interface, the GDBMS will scan, parse and validate query. During this process, the system will also use Global Directory/Global Dictionary/Metadata. The same query will be decomposed into a set of sub-queries and will go to respective remote local component relational database servers. All sub-queries will be executed locally and at that time sub-queries will use local Directory/Dictionary/Metadata information of the local database servers and all sub-queries will produce responses. In case of information retrieval, sub-results coming from individual local schemas are then to be composed and send back result to respective user.

Similarly, through view user can update remote sites with applying SQL query on global view. However, there will be some restrictions in updating remote site databases through view. We are not taking into consideration this issue in this paper. The global view in the Global Relational database will continuously grow based on the requirements of the organization and expansion of the remote databases over time.

In our proposed framework, when global views are created within the multidatabase system, at the same time metadata from the remote sites is stored within the multidatabase system tables as we named as Global Directory/Global Dictionary. This metadata can be queried locally to quickly obtain information about views. This information will include column attributes, index definitions and what data objects exist at which site(s) in the multidatabase system. Now global and local query optimization is another issue in the multidatabase system. We are not taking into consideration this issue in this paper.

Layer 4- in layer 4 all remote distributed local relational database servers are running and also maintaining metadata information of local server in local dictionary/local directory.

IV GLOBAL SERVER INTERFACE

We first define remote servers interface definition at Global Database Management System (GDMS) through which Global database server will be able to interact with the remote servers. The way of server interface definition is as follows:

Database Server Interface Driver Definition
Global Server Name
Global Server IP Address
Global Server Port
Communication Protocol

Example

Global Server Name: NewYork
Global Server IP Address: 144.16.192.211
Global Server Port: 5000
Communication Protocol: TCP

V REMOTE DATABASE GATEWAY

Database Gateways address the needs of data access in a distributed environment. Gateways make it possible to integrate with any number of remote database servers. Global view can only be created on the basis of using SQL

gateway otherwise system will not understand where to go during creating Global View. The way of Remote Database Gateway definition is as follows:
Global Server Name.Database Name.Database Owner Name.Object Name
Where object name is a name of relational table created at remote server by dbo.
Example
NewYork.StateDB.dbo.NameOfState

VI INTEGRATION METHODOLOGY AND CREATION OF GLOBAL VIEW IN MULTIDATABASE

Now we show how to create a global view using N numbers of remote database server's relational tables.

```
create view View1

 (CountryCode, StateCode, StateName)

as

select CountryCode, StateCode, StateName from
NewYork.StateDB.dbo.NameOfState

union

select CountryCode, StateCode, StateName from
London.StateDB.dbo.NameOfState

union
. . . . . .
union

select CountryCode, StateCode, StateName from
nCountry.StateDB.dbo.NameOfState
```

In the above example View1 is a name of Global view created in Global Schema. Attributes of the global view are CountryCode, StateCode and StateName. Attributes of remote relational tables are similarly CountryCode, StateCode and StateName and are mapped and that we are going to use in creating view. There may be more attributes in the remote relational tables. We may use N number of Remote gateway such as

NewYork,StateDB.dbo.NameOfState,

London,StateDB.dbo.NameOfState,

. . . .,

nCountry,StateDB.dbo.NameOfState

Where NewYork is the name of remote local database server located in New York, USA. StateDB is the name of database server which is created in the remote server, dbo is an owner of object who created object and NameOfState is a name of relational table created in the StateDB of the Remote Server NewYork. Similarly for London based remote server and for n numbers of countries.

VII DATABASE MODIFICATIONS PROPAGATION

If DBA at remote server changes the structure of the attributes such as type of the attribute, size of the attribute of the relation table will be automatically reflected in the Global View. Some user-defined functions may also be proposed to resolve schema conflicts. For example, attribute of one site‘s table is varchar and in other site it is char, then the user-defined function at the time of integration will solve this problem. I am not taking into consideration this issue in this paper. Many researchers have addressed these issues. Please see detailed work of [1,8,12,15].

VIII SUBMIT QUERY ON GLOBAL VIEW

If a user is submitting following query through their client application on global view of the multidatabase system:

Select * from View1

This will give a list of all states across n numbers of country using a single query. The federated query will be executed as follows:

Select CountryCode, StateCode, StateName from NewYork,StateDB.dbo.NameOfState
Union

Select CountryCode, StateCode, StateName from London,StateDB.dbo.NameOfState
Union

Select CountryCode, StateCode, StateName from Delhi,StateDB.dbo.NameOfState
Union
. . . .
Select CountryCode, StateCode, StateName from nCountry,StateDB.dbo.NameOfState

Similarly, we can modify data of all remote sites using a single SQL query on global view.

IX DATABASE SECURITY

The increased usage of databases to store large amounts of data has created new security problems. Typically a database contains data of various degrees of importance and levels of sensitivity. This data is shared among a wide variety of users with different responsibilities and privileges. It is therefore necessary to restrict users of the database to those portions of the total data that are necessary for their activities. Additionally, more control is needed over changes a user can make to data because of the many ways these changes can affect other users of the database [11]. The Network security expert at each remote site can better protect remote site database server by implementing a firewall between the Global Database Server and the local database server and will examine each incoming packets coming to local database server, will authenticate this and will decide whether this packet is to be denied, dropped or forwarded to local database server. Since the IP address, port and the type of network service that the Global Database Management System is using in communicating with the remote database server is known by the firewall policy rules, can easily forward, drop or deny incoming packets. The DBA at each local site will provide a better database server level and database object level security. The System Administrator at each local site will provide a better OS level security. How to exactly tackle all these issues, we are not explaining in details in this paper.

X CONCLUSION

The main objective of the work is to provide transparent access and manage distributed, heterogeneous, autonomous, and relational databases. This is a viable proposed framework in the integration of distributed, heterogeneous, autonomous, and relational databases in the Multidatabase System and easy to maintain the Global Schema. In future, we plan to address other issues in the multidatabase system.

XI REFERENCES

1) M. W. Bright, A. R Hurson, and S. H Pakzad, A Taxonomy and Current Issues in Multidatabase Systems. IEEE Computer, Vol. 25, No. 3, pp. 50-60, March 1992.
2) Pitoura, O. Bukhres, and A. Elmagarmid, Object-Oriented in Multidatabase Systems. ACM Computing Surveys, Vol. 27, No. 2, pp. 141-195, June 1995.
3) Czejdo, and M. Taylor, Integration of Database System Using an Object-Oriented Approach. Proceedings of 1st International Workshop on Interoperability in Multidatabase Systems, April 1991, pp. 30-37.
4) W. Gotthard, P. C. Lockemann, and A. Neufeld, System-Guided View-Integration for Object-Oriented Databases. IEEE Transactions on Knowledge and Data Engineering, Vol 4, No. 1, pp. 1-22, February 1992.
5) M. Kaul, K. Drosten, and E. J. Neuhold, ViewSystem: Integrating Heterogeneous

Information Bases by Object-Oriented Views. Proceedings of 6th International Conference on Data Engineering, 1990, pp. 2-10.

6) L. Koh, and A. L. P. Chen, Integration of Heterogeneous Object Schemas. in Entity-Relationship Approach, Springer-Verlag, 1994, pp. 297-314.
7) Chao Ching-Ming, Schema Integration between Object- Oriented Databases. Tomkang Journal of Science and Engineering, 2001.
8) Rehab M. Duwairi, A framework for Generating and Maintaining Global Schemes in Hetrogeneous Multidatabases Systems. *IEEE*, 2003.
9) Soon M. Chung, and Pyeong S. Mah, Schema Integration for Multidatabase Using the Unified Relational and Object-Oriented Model. ACM, 1995.
10) Umar, Distributed Database Management Systems Issues and Approach. The University of Michigan, Technical Report, July 1988.
11) Bertino, and L. M. Haas, Views and Security in Distributed Database Management Systems.
12) R. Motz, and P. Fankhauser, Propagation of semantic modifications to an integrated schema.
13) Tsichritzis, A. Klug, The ANSI/X3/SPARC DBMS Framework Report of the Study Group on Database Management Systems. Information Systems, 1:173- 191, 1978.
14) Tamer Ozsu, Patrick Valduriez, Principles of Distributed Database Systems.
15) M. G. Ali, Object-Oriented Approach for Integration of Heterogeneous Databases in a Multidatabase System and Local Schemas Modifications Propagation, IJCSIS, USA, 2009.
16) M. G. Ali, Multidatabase System as 4-Tiered Client- Server Distributed Heterogeneous Database System, IJCSIS, USA, 2009.

Shortest Path Algorithms in Transportation Networks

V.V.S.Chandra Mouli[1], S.Meena Kumari[2], N.Geethanjali[3]

GJCST Computing Classification C.2.1 & C.2.5

***Abstract*- Shortest Path problems are among the most studied network flow optimization problems with interesting applications in a wide range of fields. One such application is in the field of GPS routing systems. These systems need to quickly solve large shortest path problems but are typically embedded in devices with limited memory and external storage. Conventional techniques for solving shortest paths within large networks cannot be used as they are either too slow or require huge amounts of storage. In this project we have tried to reduce the runtime of conventional techniques by exploiting the physical structure of the road network and using network pre-processing techniques. Our algorithms may not guarantee optimal results but can offer significant savings in terms of memory requirements and processing speed. Our work uses heuristic estimates to bind the search and directs it towards a destination. We also associate a radius with each node that gives a measure of importance for roads in the network. The farther we get from either the origin or destination the more selective we become about the roads we travel with greater importance (i.e. roads with larger radii).By using these techniques we were able to dramatically reduce the runtime performance compared to conventional techniques while still maintaining an acceptable level of accuracy.**

Keywords- Routing, Shortest Path, Network, Radius.

I INTRODUCTION

We consider a long-studied generalization of the shortest path problem, in which not one but several short paths must be produced. The k shortest paths problem is to list the k paths connecting a given source-destination pair in the digraph with minimum total length. Our techniques also apply to the problem of listing all paths shorter than some given threshold length. Due to the nature of routing applications, we need flexible and efficient shortest path procedures, both from a processing time point of view and also in terms of the memory requirements. Unfortunately, prior research does not provide a clear direction for choosing an algorithm when one faces the problem of computing shortest paths on real road networks. Past research in testing different shortest path algorithms suggests that Dijkstra's implementation with double the best algorithm for networks with nonnegative arc lengths [1, 2]. However like most popular papers on Shortest Path algorithms, they have concentrated their focus on algorithms that guarantee optimality and have worked on tuning data structures used in implementing these algorithms. Since no "best" algorithm currently exists for every kind of transportation problem, research in this field has recently moved to the design and implementation of "heuristic" shortest path procedures, which are able to capture the peculiarities of the problem under consideration and improve the run time performance of a search, but at the cost of not guaranteeing optimality. As it is impossible to cover all search implementations, we use Dijkstra's algorithm as a building block to create an efficient search algorithm that implements an artificial intelligence approach to the routing problem that may not guarantee optimal results but gives significant savings in terms of memory requirements and processing speed. In the version of these problems studied here, cycles of repeated vertices are allowed. We first present a basic version of our algorithm, which is simple enough to be suitable for practical implementation while losing only a logarithmic factor in time complexity. We then show how to achieve optimal time (constant time per path once a shortest path tree has been computed) by applying Frederickson's algorithm for finding the minimum k elements in a heap-ordered tree.

Manuscript received "Feb 21, 2010 at 12:06 PM GMT"

1st V.V.S.Chandra Mouli, & 2nd S.Meena Kumari, Assistant Professor, G.Pulla Reddy Engineering College, Kurnool , A.P, India
(Email: chandu2527@gmail.com)
3rd N.Geethanjali, Associate Professor, Computer Science And Technology Department,Sri Krishna Devaraya University ,Anantapur,A.P,India
(Email: anjali.csd@yahoo.com)

II DIFFERENT SEARCH ALGORITHMS

In the following subsections we discuss about different searching techniques.

A. *Intelligent Transport System*

To fully appreciate the merits of a search technique it is important to understand the commercial environment in which these techniques are implemented. Many route finding systems are currently in development worldwide and the majority form part of much larger systems to our paper manage and operate the road network more efficiently. These management infrastructures are known as Intelligent Transport Systems (ITS) and vary in complexity and size. These systems fall into two main categories, centralized and decentralized systems [3]. Centralized systems are linked to an information centre which collates and processes traffic and network information. Typically a driver requests a particular route from onboard electronics. The route is then relayed to a central location that carries out all the processing of the route. Decentralized systems on the other hand offer information to the driver which is computed onboard using local information sources. Typically such

systems contain road network information on optical storage devices and electronics to feed a GPS.

B. Network Definitions

Before continuing let us introduce some notation and formally define the shortest path Problem. A network is a graph G = (N, A) consisting of a unique indexed set of nodes N With n = |N| and a spanning set of directed arcs A with m = |A|. Each arc a is represented as an ordered pair of nodes, in the form "from node i to j", denoted by a = (i,j). Each arc (i,j) has an associated numerical value l_{ij}, which represents the Distance, time or cost incurred by traversing the arc. Each node i has a set of successors S(i) (i.e. the set of all nodes j: (i,j)∈A) and predecessors P(i) (i.e. set of all nodes j: (j,i) ∈A).

C. Search Algorithms

One possible approach to solving shortest path problems would be to pre-calculate and store the shortest path from every node to every possible other node, which would allow us to answer a shortest path query in constant time. Unfortunately the required storage size and computation time grows with the square of the number of nodes. With realistic road networks in mind this processing would take years if not decades and be impossible to store. Hence to overcome this problem we require real time search techniques. From previous studies [1, 2, 4] we know that the implementation of labeling algorithms are the fastest for one-to-one searches.

Two aspects are particularly important to the shortest path algorithms discussed in this project:

i. The strategies used to select the next node to be visited during a search, and
ii. The data structures utilized to maintain the set of previously visited nodes.

A number of data structures can be used to manipulate the set of nodes in order to support search strategies. These data structures include arrays, singly and doubly linked lists, stacks, heaps, buckets and queues. Detailed definitions and operations related to these data structures are standard knowledge and are well documented. Past research has concentrated mainly on the issue of data structures, which can be manipulated and bounded to form clever techniques in creating priority queues for selecting nodes to be scanned. A good example of this is the Dijkstra implementation with double buckets [1]. In a labeling algorithm, the number of visited nodes during a search is a good indication of the size of the search space. This means that a search strategy which visits fewer nodes during a search is generally more efficient in terms of processing speed. The number of nodes visited depends on the depth d (i.e. the number of arcs on the optimal path) of the destination from the origin, and the branching factor b. For a 'best first search' the number of nodes explored during a search is of the order O(bd) [3]. This exponential growth in the number of explored nodes is known as "combinatorial explosion" and is the main obstacle in computing shortest paths in large networks. (Note that even though Dijkstra's algorithm is polynomial in the number of nodes n in the graph, this bound is no restriction on how the number of nodes visited varies with d). For general search this exponential growth with depth makes many problems unsolvable on current hardware, as memory is soon exhausted and a solution may take an unreasonable time to compute. These effects can be lessened by using artificial intelligence (heuristic type) techniques which will be discussed later. However let us first define and implement Dijkstra's labeling algorithm.

D. Dijkstra's Naive Implementation

Your Dijkstra's labeling method is a central procedure in shortest path algorithms. The output of the labeling method is an out-tree from a source node s, to a set of nodes L. An out-tree is a tree originating from the source node to other nodes to which the shortest distance from the source node is known. This out-tree is constructed iteratively, and the shortest path from s to any destination node t in the tree is obtained upon termination of the method.

Three pieces of information are required for each node i in the labeling method while constructing the shortest path tree:

i. The distance label, d (i),
ii. The parent-node/predecessor p (i),
iii. The set of permanently labeled nodes L.

The distance label d(i) stores an upper bound on the shortest path distance from s to i, while p(i) records the node that immediately precedes node i in the out-tree. If a node has not yet been added to the out-tree, it is considered 'unreached'. Normally the distance label of an unreached node is set to infinity. When we know that the shortest path from node s to node i is also the absolute shortest path, then node i is called permanently labeled. When further improvement is expected to be made on the distance from the origin to node i, then node i is considered only temporarily labeled. It follows that d(i) is an upper bound on the shortest path distance to node i if node i is temporarily labeled, and d(i) represents the final optimal shortest path distance to node i if the node is permanently labeled [1,2]. By iteratively adding a temporarily labeled node with the smallest distance label d(i) to the set of permanently labeled nodes L, Dijkstra's algorithm guarantees optimality. One advantage with Dijkstra's labeling algorithm is that the algorithm can be terminated when the destination node is permanently labeled. Most other algorithms guarantee optimal shortest paths only upon termination when the entire shortest path tree has been explored.

E. Symmetrical Dijkstra Algorithm

Pohl adapted Dijkstra's shortest path algorithm to decrease the size of the search space [1]. Pohl's algorithm was the first to use a bi-directional search method. This algorithm consists of a forward search from an origin node to the destination node and a backwards search from the destination node to the origin node. This was done in an attempt to reduce the search complexity to $O\ (b^{d/2})$

compared to $O(b^d)$ as with Dijkstra's algorithm. This search method assumes that the two searches grow symmetrically and will meet in some middle area. Sometimes this might not be the case, and as a worst-case scenario, this might instead become two $O(b^d)$ searches. The symmetrical or Bi-directional Dijkstra's algorithm by Pohl grows two search trees, one from the origin, giving a tree spanning a set of nodes *LF* for which the minimum distance/time from the origin is known, and a second from the destination that gives a tree spanning a set of nodes *LB* for which the minimum distance/time to the destination is known. We iteratively add one node to either *LF* or *LB* until there exists an arc crossing from *LF* to *LB*. Like Dijkstra's algorithm Pohl's bi-directional search chooses the node with the smallest cost label to label permanently. By selecting the new permanently labeled node from either the forward or backward phases we maintain the Dijkstra criterion required for optimality.

F. A Search*

So far we have examined search techniques that can be generalized for any network (as long as it does not contain negative length cycles). However the physical nature of real road networks motivates investigation into the possible use of heuristic solutions that exploit the near-Euclidean network structure to reduce solution times while hopefully obtaining near optimal paths. For most of these heuristics the goal is to bias a more focused search towards the destination. As we shall see, incorporating heuristic knowledge into a search can dramatically reduce solution times. When the underlying network is Euclidean or approximately Euclidean as is the case of road networks, then it is possible to improve the average case run time of the Dijkstra and Symmetrical Dijkstra algorithms. This is usually at the expense of optimality; solutions are now not guaranteed to be the best. Typically when solving problems on directly based or variations on Dijkstra's labeling algorithm

.The A* algorithm by Hart and Nilsson [2] formalized the concept of integrating a heuristic into a search procedure. Instead of choosing the next node to label permanently as that with the least cost (as measured from the start node), the choice of node is based on the cost from the start node plus an estimate of proximity to the destination (a heuristic estimate) [4]. To build a shortest path from the origin s to the destination t, we use the original distance from s accumulated along the edges (as in Dijkstra's algorithm) Plus an estimate of the distance to t. Thus we use global information about our network to guide the search for the shortest path from s to t. This algorithm places more importance on paths leading towards t than paths moving away from t. In essence the A* algorithm combines two pieces of information:

i. The current knowledge available about the upper bounds (given by the distance labels d (i)), and
ii. An estimate of the distance from a leaf node of the search tree to the destination.

There are several ways to estimate the lower bound from a leaf node in the search tree to the destination node. These estimations are carried out by so called "evaluation" functions [3]. The closer this estimate is to a tight lower bound on the distance to the estimation, the better the quality of the A* Search. Hence the merits of an A* search depends highly on the evaluation function h(i,j). There are two main evaluation functions used in the A* search. A true lower bound between two points is the length of a straight line between those two points (i.e. the Euclidean distance):

$$H_E(i,t) = \text{sqrt}[(x(i)-x(t))^2 + (y(i)-y(t))^2]$$

where x(i), y(i) and x(t), y(t) are the coordinates for node i and the destination node t respectively. The other commonly used evaluation function is the Manhattan distance hM. In this case the estimated lower bound distance is the sum of distance in the x and y coordinates.

$$H_M(i,t) = |x(i) - x(t)| + |y(i) - y(t)|$$

The Manhattan distance is not the true lower bound between two points and hence will typically yield non-optimal results. By using time as a measure of cost, the network becomes near-Euclidean. This is because of the varying speeds of roads in the network. Roads of similar lengths might have different times associated with using those roads. If the network is not strictly

Euclidean but near-Euclidean then our selection criteria for the next node to label permanently will not yield optimal results. By using the A* search, the shortest path tree should now grow towards t (unlike Dijkstra's algorithm where the tree grows approximately radially). As before, the search for the shortest path is terminated as soon as t is added to the shortest path tree. Earlier we discussed the problem of combinatorial explosion with a blind search time complexity in the order of $O(b^d)$. With A* search this is reduced to $O(b_e^d)$ where b_e is the effective branching factor. The A* search reduces the search space by reducing the number of node expansions. Although A* is still susceptible to the problem of combinatorial explosion, it decreases the effect by reducing the size of the base in the complexity term.

G. Weighted A Search*

By choosing an appropriate multiplicative factor we can increase the contribution of the estimated component in calculating the label of a vertex (i.e. increase the contribution of the evaluation function) [4]. From an intuitive standpoint this corresponds to further biasing the forward search towards the destination and the backward search towards the origin. The heuristic is parameterized by the multiplicative factor termed the "overdo" parameter used to weight the evaluation function. This modification will generally not yield optimal paths, but we would expect it to further reduce the search space. The aim is to find an "optimal" multiplicative or over do factor for which the running time is significantly improved while the solution quality is still acceptable. Thus there will be an empirical

time/performance trade-off as a function of the overdo parameter.

H. Radius Search

To eliminate or minimize the effects of combinatorial explosion we need to adopt a search technique similar to the way humans approach navigation problems. So far we have not implemented any intelligence once within a search which can filter out roads that are less likely to be traveled on. This type of intelligence requires some form of historical knowledge about the network. Since the road network does not change very often it is possible to calculate auxiliary information in a pre-processing step. Perhaps the most obvious way to classify the roads in the network is to identify the class of each road (i.e. motorways, highways, local roads etc), and then to exploit these classes in the search. This is similar to the way humans approach routing problems and is known as Hierarchical Search [3,5]. Hierarchical methods offer the prospect of greatly reducing the size of any search by simplifying the search through a series of simplified levels, where each of these levels is an abstraction of the previous level. These abstractions reduce the overall size of the search space that an algorithm addresses and thus the complexity of any search is reduced. For route finding, hierarchical levels are constructed in which higher speed roads are placed higher up in the hierarchy. However by introducing these arbitrary hierarchies the path optimality is often lost [3].

The hierarchical algorithm uses a discrete number of hierarchy levels. A Radius search is a hierarchical search with a continuous range of hierarchy levels. A Radius search takes advantage of the fact that the fastest path between two junctions is more likely to use a highway than a local road, especially if the two junctions are far apart. In this method each node i has an associated radius r(i). Before we consider how r(i) is calculated, we first examine how radii can be used to restrict a search. When looking for a shortest path from s to t, a node i is considered as a possible node to include in the search only if s or t lies inside a circle of radius r(i) cantered at node i. If both distances are greater than the node radius, the node is simply ignored [5].For any given origin and destination node, we can immediately simplify the network by removing all the nodes (and associated arcs) whose radii do not encircle the origin or destination nodes. The radius search is not a search algorithm by itself, but an independent mechanism of reducing search complexity. Hence the radius concept can be used in conjunction with any search algorithm.

The optimal radius for a node i is the smallest radius r(i) for which the radius centred at node i encircles either the origin or destination node for all optimal paths that include node i. If the radii are calculated as a maximum over all such shortest paths, then it is guaranteed that the radius search algorithm is exact (i.e. guaranteed optimality). The radii are also minimal since with any smaller radius at least one optimal shortest path will not be found. One possible difficulty is that the calculation of the radii by examining all paths over a particular node takes much too long since every possible shortest path in the network has to be calculated at least once. Instead we implemented a heuristic approach to calculate these radii [5]. In the first phase of this heuristic approach we divide the network into overlapping grids of approximately 2000 nodes and initialize all node radii to be 0. We then select a random starting node s from all possible nodes N and a random destination node t within the same grid as s. Using the Symmetric Dijkstra algorithm we solve for the shortest path R from s to t. We continue this process of selecting random starting and destination nodes and updating the radii of nodes in the shortest path as many times as possible.

If we do not generate enough random paths in the first phase then the radii of some nodes will never have been updated and hence will still be 0. However if a node is a 'closed node' (i.e. the node is only used in a shortest path if it is either the origin or destination of that shortest path) then it will never be part of a shortest path unless we start or finish at that node. Hence the radii of closed nodes will always be 0. In the second phase of this modified algorithm we go through all nodes in the network and examine their radii. If a node is not closed and has 0 radius, then we conduct shortest path searches in the vicinity of the nodes that generate a reasonable lower bound on its radius. We do this in the second phase by creating a sub graph of 200 of the closest nodes and associated arcs GSUB2 to the node with 0 radius and solve all-to-all shortest paths on GSUB2. This should force some shortest paths R through this node and give it a better radius lower bound than 0. So far in the first two phases we have calculated shortest paths within grids. Hence the radii are no larger than the grids they are created in. As a result, after the first two phases we have a fairly good coverage of local radii only (i.e. these radii only restrict a search for shortest paths within grids). If we were to use these radii to restrict a search over a large distance (i.e. over several grids) then we would not be able to find a path because no nodes exist which have radii greater than the size of a single grid. To travel over large distances we need to calculate radii of roads such as highways and motorways

III CONCLUSION

By exploiting the physical structure of road networks, the A* algorithm is able to bias its search towards a goal and reduce the search space. By using the concept of radii as a measure of importance of nodes, we are able to incorporate pre-processing within our shortest path algorithm to further restrict the search space. This dramatically reduces the search complexity in terms of the run time performance while still maintaining an acceptable level of inaccuracy. For a one to one shortest path or the shortest paths from one to some, it may be worthwhile to consider one of the Dijkstra 's implementations. But Dijkstra implementations depend on the maximum size of the network arc lengths Dijkstra approximate buckets implementation (DIKBA) is recommended for less arc length. For problems with a maximum arc length greater than 1500, the Dijkstra double buckets (DIKBD) implementation should also be considered since it appears to be less sensitive to problems in data set 1

with very large arc lengths. The Bellman Ford Moore implementations with parent checking (BFP) have serious difficulties on large networks. So this algorithm is not recommended for road network and for being coded in a GIS package. This system can efficiently generate less similar paths and provide users more wide choices than other system. Because of the simplicity of the topological structure and the k-shortest path algorithm, the developer can also easily develop a rich featured user interface for displaying and setting.

IV References

1) Coello, C. A. C.,(1999). An Updated Survey of Evolutionary Multiobjective Optimization Techniques: State of the Artand Future Trends,(1999). Proceedings of the Congress on Evolutionary Computation, 1, 3-13, IEEE Press,6-9
2) Cormen T.H, Introduction to Algorithms. MIT Press, Massachusetts, USA. Cherkassy B V, Goldberg A V and Radzik T. (1993) Shortest Paths Algorithms:
3) Theory and Experimental Evaluation. Research project, Department of Computer Science, Cornell and Stanford Universities and Krasikova Institute for Economics and Mathematics.
4) Hart P E and Nilson N J. (1968) A formal basis of the heuristic determination of minimum cost paths. IEEE Transactions of Systems Science and Cybernetics 4(2), 100-107. 45 Pearsons J. (1998) Heuristic Search in Route Finding. Master's Thesis, University of Auckland.
5) Sedgewick R and Vitter J S. (1986) Shortest Paths in Euclidean Graphs Algorithmica 1, 31-48.
6) Ertl G.(1996) Optimierung und Kontrolle, Shortest Path Calculations in Large Road Networks. Project in Discrete Optimisation, Karl-Franzens-Universitat Graz .
7) Aggarwal, B. Schieber, and T. Tokuyama (1993). Finding a minimum weight K-link path in graphs with Monge property and applications. Proc. 9th Symp. Computational Geometry, pp. 189–197. Assoc. for Computing Machinery.
8) R. K. Ahuja, K. Mehlhorn, J. B. Orlin, and R. E. Tarjan. (1990) Faster algorithms for the shortest path problem. J. Assoc. Comput. Mach. 37:213–223. Assoc. for Computing Machinery.
9) Azevedo, M. E. O. Santos Costa, J. J. E. R. Silvestre Madeira, and E. Q. V. Martins (1993). An algorithm for the ranking of shortest paths. Eur. J. Operational Research 69:97–106.

Lightweight Multipriority Framework Classes for Bandwidth Differentiation Service

Mr. R. SENTHIL KUMAR., M.E.,(Ph.D),
ASSISTANT PROFESSOR,
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING,
V.M.K.V.ENGINEERING COLLEGE,
SALEM, TAMILNADU, INDIA.
E-mail: trsk_kavi2042@yahoo.co.in

GJCST Computing Classification
C.2.1, C.2.2 & C.2.m

***Abstract*- Internet deploy new applications, such as voice over IP, video conference, video streaming, audio streaming, mission-critical financial data, employ the user datagram protocol (UDP) and have different requirements on the bandwidth from those of existing applications using transmission control protocol (TCP).**

TCP-friendliness and multimedia-friendliness "contradict" with each other. The possible approaches to address these problems are enhance the Internet with resource reservation, admission control and adopt source adaptation schemes such that the sending rate of an application is adjusted

According to the current network condition. The source adaptation scheme, better preferred has two drawbacks i.e., it is more moderately responsive to transient changes in congestion, including a slower response to an abrupt change in the available bandwidth and the requirement of multimedia applications cannot be satisfied by equation-based source adaptation when a large number of flows share a link.

The proposed system, develop a framework which extend the concepts of fairness, TCP-friendliness and TCP-compatibility, by designing TCP and UDP friendly source adaptation schemes and develop a less

Conservative framework to provide bandwidth differentiation service without any change to the router of the existing Internet.The proposed scheme is extended towards the handling of multiple bottleneck links available in the real time networks. In addition the scheme combines with TCP slow-start, timeout to provide a relatively differentiated bandwidth service in the existing Internet, without any change to the router. The newly developed source adaptation scheme is improved to work with multiple priority classes.

I INTRODUCTION

Many new applications are being widely deployed in the Internet, such as voice over IP, video conference, video streaming, audio streaming, mission-critical financial data, and so on. Many of these applications employ the user datagram protocol (UDP) and have different requirements on the bandwidth from those of existing applications using transmission control protocol (TCP). For example, multimedia applications have the following two distinguished requirements:

i. The sending rate should be adjusted smoothly
ii. The sending rate should be greater than a threshold.

TCP-friendly source adaptation schemes do not work well for multimedia applications. It seems that TCP-friendliness and multimedia-friendliness "contradict" with each other. Thus, these new applications present new problems for the Internet.

There are two possible approaches to address these problems. One is to enhance the Internet with resource reservation, admission control and so on. The other is to adopt source adaptation schemes such that the sending rate of an application is adjusted according to the current network condition. Compared to the first approach, the second one has the advantage of better utilization of the available time-varying network resources.

II LITERATURE REVIEW

Recently, an interesting framework, called Choose Your Response Function (CYRF), was proposed for memory less window-based source adaptation protocols by using the fairness index.

Many existing congestion control schemes, like additive-increase multiplicative-decrease (AIMD), generalized AIMD (GAIMD)[4], binomial congestion control (BCC)[7], are special cases of the CYRF. The CYRF has many nice features, such as: it converges to fairness and efficiency with a simple source adaptation scheme, it can be easily adapted to suit different applications and network requirements and it can be made TCP-friendly, and so on. However, the CYRF has two disadvantages.[2, 3]

One is that it is difficult to use the CYRF to design multimedia friendly source adaptation schemes. The other is that the requirements of CYRF are still a little conservative. For example, a protocol named LOG is not CYRF. Thus the concept of 1-CYRF is to cover the case, which is a protocol with a smooth increase policy whose and are monotonically non-decreasing (one of them must be strictly increasing). [6]

Another important type of source adaptation is equation based source adaptation. The advantage of this type of source adaptation, as compared to the CYRF, is a smoother sending rate. It can also be used to maintain a relatively steady sending rate. However, it relies heavily on the accurate measurement of round trip time, steady-state loss event rate and the TCP retransmit time out value, especially the steady-state loss event rate that is very difficult to be measured.[5]

Furthermore, it has another two disadvantages:

i. It is more moderately responsive to transient changes in congestion, including a slower response to an abrupt change in the available bandwidth. As a result, it cannot aggressively find and use available bandwidth
ii. The second requirement of multimedia applications cannot be satisfied by equation-based source adaptation when a large number of flows share a link.

The proposed method, extend the concepts of fairness, TCP friendliness and TCP-compatibility, such that the following two objectives can be achieved.

i. A platform can be set up to design TCP and UDP friendly source adaptation schemes.
ii. Bandwidth differentiation service can be provided by only using source adaptation with signal fed back from the destination.

This will increase the incentives for applications to use end-to-end source adaptations, thus contributing to the overall stability and evolution of the Internet. The proposed method, develop a less conservative framework to provide bandwidth differentiation service without any change to the router of the existing Internet.

III SYSTEM MODEL

The system model of the thesis consists of the following components

A. *Monotonic Response Function*

In this module a general framework for window-based memory less source adaptation protocols that is convergent to the desired fairness is developed. In this framework a monotonically non-increasing function and a monotonically non-decreasing function, and that the sending rates of flows are adjusted by increase policy and decrease policy.

B. *Smoothness and Efficiency*

The smoothness of traffic adjustment is an important property when video, audio or speech is transmitted over the network. A window increase (decrease) policy is said to be smooth if the window size increase (decrease) from a single application is at least an order of magnitude smaller than the current window size.

Meanwhile, the policies are required to move the total bottleneck link utilization closer to the link capacity for efficient use of the bandwidth. This can be achieved by the principle of negative feedback, i.e., each flow increases its window size when the bottle link is under-utilized and decreases its window size when it is overloaded.

C. *TCP Fairness Convergence*

The system set four preconditions on the network.

i. Synchronization Assumption

It is assumed that all the flows in the network get the same feedback and get the feedback simultaneously.

ii. Feedback Signal

The feedback is binary and limited to a single bit indicating whether the network is overloaded, i.e. a '1', or if the bandwidth is not fully utilized, i.e. a '0'.

iii. Response Functions

Assume that all sources use the same functions. When an ACK is received and the feedback is 1, the next window size is computed by the policy. If the feedback is 0, the policy is used to calculate the next window size.

iv. Number of Bottleneck Links

It is suppose that the number of bottleneck links is satisfied after each application of policy or and at least one of the two policies must ensure a strict inequality.

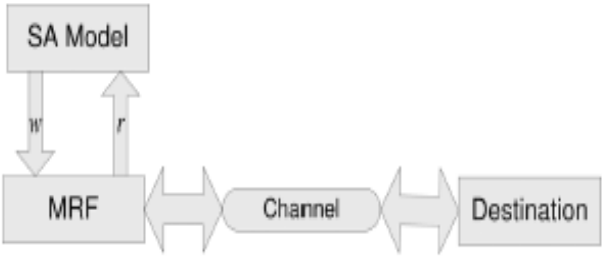


Fig: Structure of source adaption model

D. *TCP-Friendly MRF*

To guarantee the display quality of video at the receiver side, smooth source adaptation schemes are desirable. This type of source adaptation schemes are also required to be TCP-friendly w.r.t. predefined weighting factors. The system shall derive necessary and sufficient condition for our MRF with a smooth increase policy to be TCP-friendly w.r.t. the predefined weighting factor.

It employed a new source adaptation agent instead of using the conventional TCP congestion avoidance mechanism to reduce the disturbance from other mechanisms such as slow-start or time-out etc. It has a single bottle neck link with a default bandwidth of 1 MB and a delay of 30 ms. The algorithm of RED is used at each router to drop packets in case of congestion. Several TCP-Reno flows (with different RTT) were introduced as the back ground traffic. The packet size is fixed at 500 bytes and all flows start at random times.

IV EXPERIMENTATION

The experimentation employed a new source adaptation agent instead of using the conventional TCP congestion avoidance mechanism to reduce the disturbance from other

mechanisms such as slow-start or time-out etc. It has a single bottle neck link with a default bandwidth of 1 MB and a delay of 30 ms. The algorithm of RED is used at each router to drop packets in case of congestion. Several TCP-Reno flows (with different RTT) were introduced as the back ground traffic. The packet size is fixed at 500 bytes and all flows start at random times.

V Conclusion

In the proposed method, differentiation bandwidth service can be provided by adopting the monotonic responsive function (MRF) as the source adaptation scheme with necessary feedback signal from the destination. With the MRF as the source adaptation scheme, a flow can not only find and use the available bandwidth but also adjust its sending rate smoothly. More applications over the Internet can then be supported.

The proposed scheme is easier to use the Taylor's expansion to study our MRF with a smooth increase policy. The MRF can be easily used to analyze existing network protocols and construct new network protocols. The MRF does not need to measure round trip time, steady-state loss event rate and the TCP retransmit time out value as required by equation-based source adaptation. The MRF is a good platform for source adaptation of TCP and UDP. It can also be used to design TCP-friendly and multimedia-friendly source adaptation schemes. This is very helpful for multimedia applications over the Internet.

The MRF framework with a new fairness index analyze memory less window based congestion control protocols. A necessary and sufficient condition is also derived for stepwise convergence of our MRF to the fairness. The condition is used to construct increase–decrease policies. The MRF framework is used for constructing existing protocols, like additive increase multiplicative-decrease (AIMD), generalized AIMD, and binomial congestion control.

The proposed system handles multiple bottleneck links which normally present between the source and destination. It is thus of practical and theoretical interests to derive equivalent results for the case of multiple bottleneck links. In addition the proposed source adaptation works well for multiple priority classes. With our framework, design a new pricing system for the Internet by fully utilizing the feature of Internet service.

VI References

1) Cheng Chen, Zheng Guo Li, and Yeng Chai Soh, "MRF: A Framework for Source and Destination Based Bandwidth Differentiation Service", IEEE/ACM Transactions on Networking, vol.15, no.3, June 2007
2) Z. G. Li, C. Zhu, N. Ling, X. K.Yang, G. N. Feng, S.Wu, and F. Pan, "A unified architecture for real time video coding systems," IEEE Trans. Circuits Syst. Video Technol., vol. 13, no. 6, pp. 472–487, 2003.
3) X. Wang and H. Schulxrinne, "Incentive-compatible adaptation of internet real-time multimedia," IEEE J. Sel. Areas Commun., vol. 23, no. 2, pp. 417–436, Feb. 2005.
4) R. Jain, D.-M. Chiu, and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems," Digital Equipment Corp., Tech. Rep. TR-301, Sep. 1984.
5) V. Jacobson and M. Karels, "Congestion avoidance and control," ACM Comput. Commun. Rev., vol. 18, no. 4, pp. 314–329, 1990.
6) Y. R. Yang and S. S. Lam, "General AIMD congestion control," in IEEE Int. Conf. Network Protocols, Osaka, Japan, Nov. 2000, pp. 187–198.
7) Bansal and H. Balakrishnan, "Binomial congestion control algorithms," in Proc. IEEE INFOCOM 2001, Anchorage, AK, Apr. 2001, pp. 631–640.
8) S. Floyd, M. Handley, and J. Padhye, A comparison of equation-based and AIMD congestion control Mar. 2000 [Online]. Available: http:// www.aciri.org/trfc/

Mask-Co: Multi-Agent System Based Knowledge Management System to Facilitate Knowledge Sharing in Construction Organization Environment

Amir Mohamed Talib
Faculty of Computer Science & IT, University Putra Malaysia
43400 UPM, Serdang, Selangor, Malaysia
Email:ganawa53@yahoo.com
Rodziah Atan, Rusli Abdullah and Masrah Azrifah Azmi Murad
Faculty of Computer Science & IT, Information System Department,
University Putra Malaysia
43400 UPM, Serdang, Selangor, Malaysia
Email: rodziah@fsktm.upm.edu.my
Email: rusli@fsktm.upm.edu.my
Email: masrah@fsktm.upm.edu.my

GJCST Computing Classification
H.2.8, K.6.4, H.2.m

***Abstract*- Multi-agent system based knowledge Management system construction organization environment (MASK-CO) is a set of agents, software and technology designed to focus and enhance the communication, deliberations, and decision-making of groups. MASK-CO is successful in improving the efficiency, reliability and quality of the group decision-making process. Knowledge management (KM) has received considerable attention in recent years. Some consider knowledge the most strategically important resource, and learning the most strategically important capability for business organizations. Major construction organizations have recognized the benefits that KM can offer and have thus invested in KM. This paper reports on a survey of these companies. The purpose of the survey was: (1) to examine the importance of KM to construction organizations; (2) to investigate the resources used to implement KM strategies. The survey found that the main reasons for implementing a KM strategy was the need to share the tacit knowledge of key employees and to disseminate best practice. In addition, significant resources in terms of staff time and money were being invested in KM. In order to resolve the problems in construction organization and promote the performance of construction organization, a MASK-CO is designed based on Prometheus Design Tool (PDT).**

Keywords- construction organizations, knowledge management, knowledge management system, multi-agent system, and Prometheus Design Tool.

I INTRODUCTION

CO (MAS architecture to facilitate knowledge sharing in construction organization environment) is Architecture aims to provide facilitating knowledge sharing, supporting the system users to successful access to the system resources.

This techniques was inspired by the "there is lack of Architecture of MAS-Based KMS in order to product the sharing of knowledge in construction organization (MASK-COArchitecture)". As well as "lack of decision-making process, users do not know what they need, lack of trust, lack of change management, lack of risk assessment and

failure to learn from successful projects" and also" there is inconsistency of MAS using in test of its functionality", the MASK-CO model has been developed to solve this problem. The main goal of this paper is to design, develop and applying MAS techniques-based KMS in a collaborative environment of lotus notes to facilitate knowledge sharing of CO among the users.

The paper, therefore, investigates the approach adopted by construction organizations in three areas: (1) the importance of KM to the organization; (2) the resources used to implement a KM strategy; and (3) the barriers to managing knowledge within individual organizations. The first area (importance of KM) investigates how widespread proactive KM is within the construction sector. The second area (resources allocated to implementing the KM strategy) provides an insight into the type of infrastructure used to support the KM strategy. The third area (barriers to KM within organizations) identifies problem areas that need to be addressed for KM activities to bring about tangible behavioral and performance improvement.

The construction industry is facing many of the same problems as the software industry. The problems are that projects often run late, cost increases and, in many cases, results in failure. Construction projects are among the most complicated of the human enterprises. There is a high level of skill and knowledge required to translate a client's version or list of requirements into plans and specifications and then into a real building that functions well for the people who will live or work there.

Besides technical skills required in construction projects, it is also important to have people skills in order to coordinate the diverse efforts of the many people involved. There are inevitable problems encountered in the course of a construction project.

The emergence of the knowledge economy means that organizations' know how is becoming more important than traditional sources of economic power (Scarborough and Swan, 1999). Moreover, knowledge is now considered the most strategically important resource, and learning the most strategically important capability for business organizations (Zack, 1999). Thus, knowledge assets must be managed deliberately, systematically and with expertise to ensure corporate survival.

Two types of knowledge are widely accepted: tacit knowledge and explicit knowledge (Nonaka, 1991; Nonaka & Takeuchi, 1995). Tacit knowledge is highly personal, developed from experience, and hard to formalize; therefore, it is difficult to communicate. Explicit knowledge, on the other hand, is formal and systematic. It is, therefore, easy to communicate and share, for example, in product specifications or codes of practice. Drew (1999) described four types of knowledge: 1) what we know, we know; 2) what we know, we don't know; 3) what we don't know, we know; and 4) what we don't know, we don't know. He emphasized that most KM programs were concerned with processes for sharing and distributing existing knowledge that is "what we know, we know." However, he recognized that the increasing use of intelligence gathering based on knowledge networks and intranets contributed towards "what we know, we do not know."

Quintas et al. (1997) defined knowledge management as "the process of continually managing knowledge of all kinds to meet existing and emerging needs, to identify and exploit existing and acquired knowledge assets and to develop new opportunity." Webb (1998) defined knowledge management as "the identification, optimisation [sic] and active management of intellectual assets to create value, increase productivity and gain and sustain competitive advantage." It is important to recognize that KM involves the sharing of knowledge, as well as other processes. Several authors have identified these different processes. For example, Ruggles (1997) categorized these processes as generate, codify, and transfer. Siemieniuch & Sinclair (1999) identified five processes: generate, propagate, transfer, locate and access, and maintain and modify. Tiwana (2002) identified five categories as find, create new, package and assemble, apply, and reuse and revalidate knowledge. Laudon and Laudon (2000) recognized that these processes can be cyclical and iterative, and that they all have different process requirements.

Knowledge is increasingly recognised as the most important resource in organisations and a key differentiating factor in business today. It is being increasingly acknowledged that KM can bring about the most needed innovation and improved business performance in the construction industry (Egbu, Sturgesand and Gates, 1999). Knowledge is defined as a dynamic human process of justifying personal belief towards the truth (Nonaka and Takeuchi, 1995). It can also be defined as ‘know-why’, ‘know-how’ and ‘know-who’, or an intangible economic resource from which future resources will be derived (Rennie, 1999). Knowledge is built from data, which is first processed into information (i.e. relevant associations and patterns). Information becomes knowledge when it enters the system and when it is validated (collectively or individually) as a relevant and useful piece of knowledge to implement in the system (Carrillo, Anumba and Kanara, 2000). Besides the meaning of knowledge, it is the identification of the kind of knowledge that has to be managed. There are various kinds of classification of knowledge: formal (explicit) and tacit (expertise) knowledge; foreground and background knowledge; knowledge of business environment or knowledge for control activities (Carrillo, Anumba and Kanara, 2000).

According to (Brelade and Harman, 2001), KM is obtaining and using resources to create an environment in which individuals have an access to information and in which individuals obtain, share and use this information to raise the level of their knowledge. In addition to this, individuals are encouraged and enabled to obtain new information for the organisation. KM is referred to as the process of creating, codifying and disseminating knowledge for a wide range of knowledge intensive tasks. (Harris et al., 1998). These tasks can be decision support, computer-assisted learning, research (e.g. hypothesis testing) or research support.

The construction industry delivers large, expensive, custom-built facilities at the end of a construction process. It is a strong, knowledge-based industiy that relies heavily on the knowledge input by different participants in a project team. Some aspects of KM have been around for awhile, such as the attempt to capture tacit knowledge in Expert Systems and Knowledge-Based Systems during the 1980s. However, these had limited success in much defined areas, such as diagnosing the cause of dampness in buildings (Allwood, 1989). This approach of trying to capture personal experiences in information technology (IT) systems was not very successful. Technology has advanced and there is now a common understanding that IT is simply a facilitator and not the KM system.

The changeable character of the CO requires that the information generated be controlled, stored, and shared. We proposed in order to manage the knowledge generated a MAS formed of three agents are under the client agents implementation. One agent, called the send and receive mail agent, is in charge of organizing the information sent and received from the group. The other two agents are general agent (Interface Agent and Personal Agent).

The rest of the agents are also communicated, thus enabling them to interchange information. The roles of these agents are summarized as follows

i. Comparing new information with that which has already been stored in order to detect inconsistencies between old and new information. If an inconsistency is detected the agent must inform the rest of the agents in order to discover why the inconsistency has occurred.
ii. Informing other agents about changes produced.
iii. Advising certain employee to do a specific job. The system has information about each employee's skills, their performance metrics, and the projects they have worked on. Agents may process this

information to suggest which person is most suitable to carry out a task.

iv. Estimating the cost of future interventions. Information available may be used to make statistical analyses that help predict effort and costs.

II Related Works

Knowledge about agent concept alone is not sufficient to build a good agent system. There are some fundamental issues needed to drive the design of an agent (Bigus, J. P., Bigus, J., 2001). The first is to view the agents as adding value to a single standalone application, or as a freestanding community of agents that interact with each other and other applications. The first type views the agent from the perspective of application-centric, where the agents are helpers to the application, while the second is more agent-centric, where the agents monitor and drive the application.

In recent years, Multi-Agent System (MAS) has been an active research topic. Due to the difficulties in solving process planning and production scheduling problems using traditional centralized problem solving methodology, MAS approach – a distributed problem-solving paradigm is used as another attempt to solve the planning and scheduling problems. As a distributed problem-solving paradigm, MAS breaks complex problems into small and manageable sub-problems to be solved by individual agents co-operatively (Vermeulen, S. Bohte, D. Somefun & Poutré J. L, 2006).

Agent paradigm lets users think in term of agents rather than objects / functions. The agent exhibits presents high dependencies compared with an object-oriented approach. Such a software application needs an appropriate software development method. An analysis and design methodology is intended to assist first in gaining understanding of a particular system, and secondly in designing it (Wooldridge, M, 2004). There are few choices of agent-oriented methodologies to help software engineers to specify, design and build agents to achieve the system's goals.

(Dignum, V., 2006) proposed Operation per Organizations (OperA), a model for agent's organization, society and interaction model. The Organizational Model implements the desired organizational structure of an agent society, the description of an agent population that will enact the roles described in the structure is detailed in the Social Model, and the specification of agent interactions to achieve the desired society global objectives is described in the Interaction Model. However, this model needs other agent oriented methodology to help designing the system.

(Park, S., Sugumaran, V., (2005) introduced a framework of multi-agent system (MAS) development that considers both functional (services to solve complex problems in distributed environments) and non-functional service (capability to reuse, easy to extend, adapt and process uncertain data) of the system. They also suggested that, in order to develop MAS in a systematic way, system should be analyzed in terms of its ultimate goals and the system should be designed both in the abstract as well as concrete by mapping the goals and the sub-goals to software agents.

(Elst,L. V., Dignum V., & Abecker A., 2004) asserted a three-dimension overview on agent-mediated knowledge management which includes (i) understanding the stage in a system's development process where agents are used (analysis, conceptual design, or implementation) (ii) analyzing the architecture / topology of the agent system, and (iii) identifying KM functionality / application focused on.

MAS developed for job shop scheduling problems in which standard operating procedures are combined with a look-ahead coordination mechanism that should prevent 'decision myopia' on part of the agents. Using their approach, system performance is said to improve in tightly-coupled, real-time job-shop scheduling environments. However, their coordination mechanism is not appropriate for competitive, self-interested agents, which makes it an undesirable choice for coordination in a de-icing setting (Liu & Sycara, K. P, 1996).

According to (Wetherill et al., 2002), knowledge in construction can be classified into three categories: domain knowledge, organizational knowledge and project knowledge.

Domain knowledge forms the overall information content. It includes administrative information, (e.g. zoning regulations, planning permission), standards, technical rules, product databases, etc. This information is available to all companies, and is partly stored in electronic databases.

Organizational knowledge is company-specific, and is the intellectual capital of the firm. It resides both formally in company records and informally through the skilled processes of the firm. It comprises knowledge about the personal skills, project experience of the employees and cross-organizational knowledge. The latter knowledge covers the knowledge involved in business relationships with other partners, including clients, architects, engineering companies and contractors.

Project knowledge is the potential for usable knowledge and is at the source of much of the knowledge identified earlier. It is both the knowledge that each company has about the project and the knowledge that is created by the interaction between firms. It is not held in a form that promotes reuse. Companies and partnerships are often unable to capitalize on this potential for creating knowledge. It includes both project records and the recorded and unrecorded memory of the processes, problems and solutions. This paper is mainly concerned with project knowledge.

Our KM system consists of four main components comprising:

i. Interface for input and updating of captured knowledge into the knowledge web-based portal
ii. The web-based portal that stores the knowledge base which allows the users of the KM system to access the hosted knowledge
iii. A search engine that provides some searching mechanism to allow the users to search
iv. For the desired knowledge and provide a set of alternative solutions if the user is
v. Looking for solutions to a particular problem

vi. An open forum to allow all users to contribute and show their solutions on a particular problem, share their knowledge and information and get the updates of specific projects.

III METHODOLOGY

Our methodology composed of three main phases as followed:

Phase 1 – MASK-CO design by Prometheus Design Tool (PDT)

The Prometheus methodology consists of three phases (Padgham. L, & Winikoff. M, 2002):

i. System Specification: where the system is specified using goals and scenarios; the system's interface to its environment is described in terms of actions, percepts and external data; and functionalities are defined.
ii. Architectural Design: where agent types are identified; the system's overall structure is captured in a system overview diagram; and scenarios are developed into interaction protocols.
iii. Detailed Design: where the details of each agent's internals are developed and defined in terms of capabilities, data, events and plans; process diagrams are used as a stepping stone between interaction protocols and plans.

Each of these phases includes models that focus on the dynamics of the system, (graphical) models that focus on the structure of the system or its components, and textual descriptor forms that provide the details for individual entities.

Phase 1.1 Systems Specifications

Phase 1.1.1 Goals

As shown in Figure 1 below, there is one main goal for the agents, and how they are achieved, are described as follows:

A. Send and Receive Mail

i. Send mail to the destination user.
ii. Receives mail from the source user.

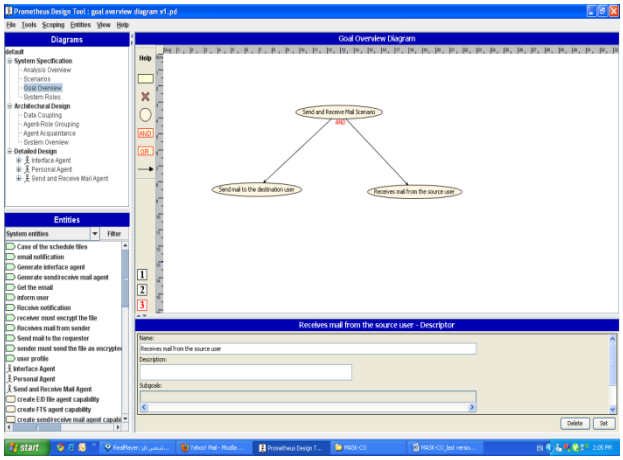


Figure 1: Goal Overview Diagram

Phase 1.1.2 System Roles

Based on the different functionality/scenarios, different roles may be extrapolated as above as shown in Figure 2 below.

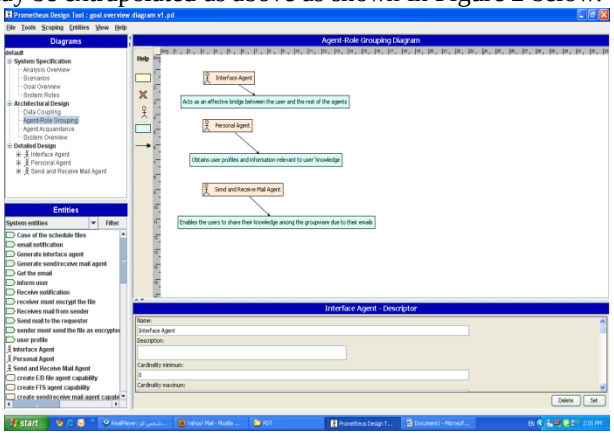


Figure 2: System Role Diagram

Phase 1.2 Architectural Design

Phase 1.2.1 System Overview Diagram

To explain in detail the functionality of each agent, the System Overview Diagram shall be used as shown in Figure 3.The identifies the Scenarios, the Agents, the Data, the Actions and the messages that are used by all Agents.

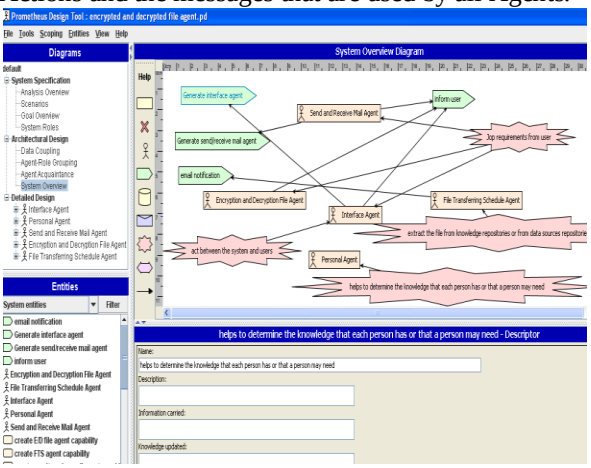


Figure 3: System Overview Diagram

Phase 1.3 Detailed Agent Design

Phase 1.3.1 Interface Agent

Interface Agent acts as an effective bridge between the user and the rest of the agents. Such agents actively assist a user in operating an interactive interface as shown in Figure 4.

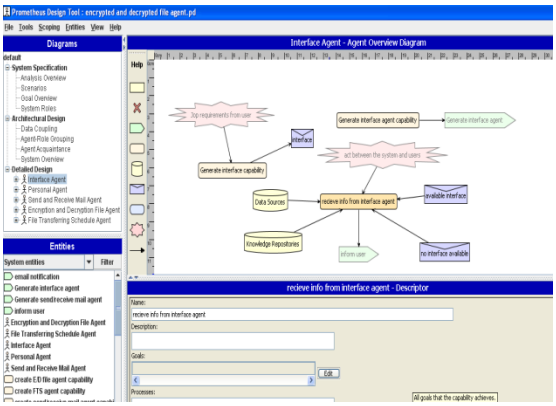


Figure 4: Interface Agent

Phase 1.3.2 Personal Agent: which obtains user profiles and information relevant to user' knowledge that helps to determine the knowledge that each person has or that a person may need as shown in Figure 5.

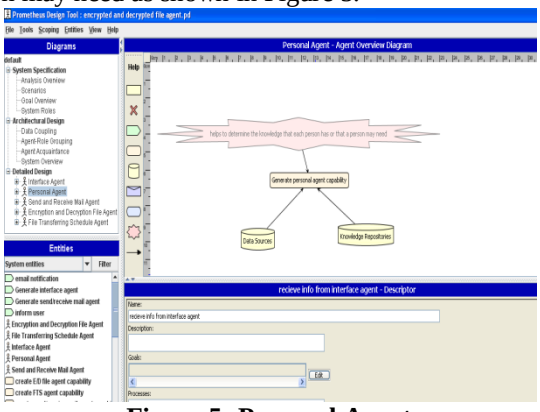


Figure 5: Personal Agent

Phase 1.3.3 Send and Receive Mail Agent: is enables the users to share their knowledge among the groupware due to their emails. This process is provided by this agent. It's also learns about interactions of a user and E-mail application to perform the tasks on E-mail according to the user preferences as shown in Figure 6.

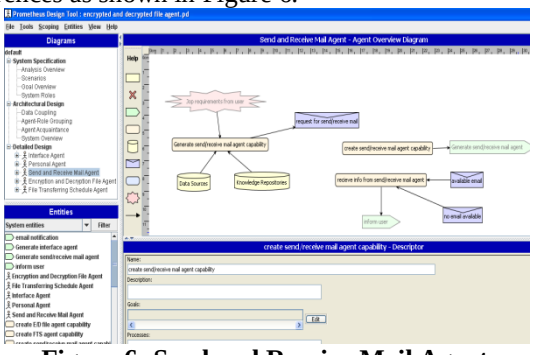


Figure 6: Send and Receive Mail Agent

Phase 2 – MASK-CO development

By using the groupware of Lotus Notes (Lotus company, 2007), the best agent technology capability that could be developed is used Java Script programming that comes along with this package.

Figure 7(a), 7(b) and 7(c) describes the communication between the agent and the whole system among the users' mails and also demonstrates the agents into the system.

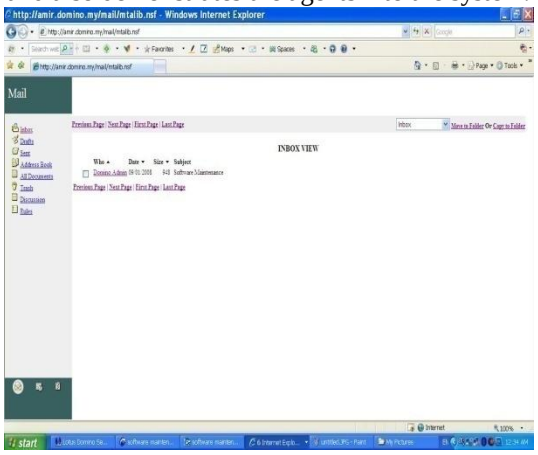


Figure 7(a): Mhd talib mail in lotus notes COE

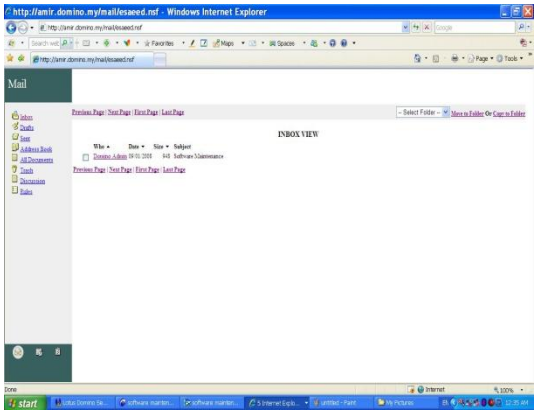


Figure 7(b): egbal saeed mail in lotus notes COE

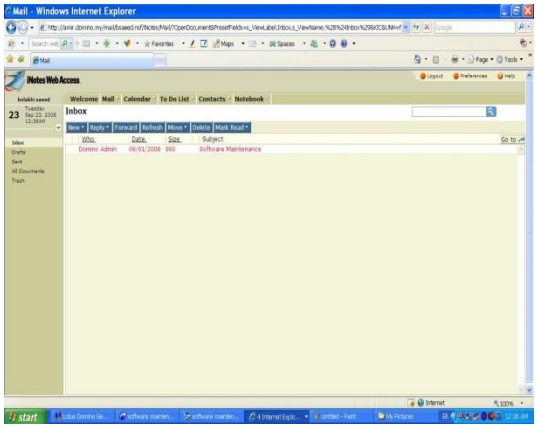


Figure 7(c): bolakhi saeed mail in lotus notes COE

Phase 3– MASK-CO evaluation

Phase 3.1 Participants

The respondents including System Analyst, System Developer, Software Engineer, and User and will be chosen to fill the questionnaire of this study. The respondents should be applying the system before solving the questionnaire to be situated.

Phase 3.2 Procedures

In the beginning, the respondents will receive a short, scripted verbal orientation. Then they will be asked to complete a short background questionnaire to collect their demographic characteristics. The respondents will be asked to perform a set of information about how to share knowledge as a kind of multi-agent technology for COE. The tasks were written on a sheet of paper that included a space where respondents will be asked to indicate their answers. Once the tasks are completed, respondents will be asked to complete a short participant satisfaction questionnaire to collect and test their own perceptions towards CO.

Phase 3.3 Tasks

Respondents will complete three tasks:

i. They will complete a background/experience questionnaire that including name, gender, age, education level, Major/Department, and years of experience.
ii. They will perform tasks using the questionnaire's sheet.
iii. There is also a post-survey questionnaire that specifically examines MAS techniques. After completing a task, the respondents will ask to rank satisfaction and to write down comments.

Phase 3.4 Data collection

This evaluation model considers both quantifying elements of performance (experience and experiment) as well as subjective empirical. If the answer is wrong, or he/she not familiar with this question then skip to the second question until all the question will be solved. We will, however, record whether respondents are able to complete tasks successfully. The criteria for successful task completion are:

i. Participant is able to give a correct answer based on his own information about the system. Any guessed or assumed answers, whether correct or not, are not record as successfully completed tasks.
ii. Participant is able to give a definite answer to the question. Where respondents indicated they are unsure about the answer or would seek clarification, the task will record as not successfully completed.

Phase 3.5 Survey

The purpose of the survey is to prove:

i. Handle the interpretation of the term KM and the company's key objective in CO.
ii. Handle the aspects that come into play in KM, such as the existence of a strategy, the processes of quality control of data, the content that is being managed, and the functioning of communities of practice.
iii. Identify the Multi-Agent technique of willingness of cooperation for research work.
iv. Identify the Multi-Agent technique for helping the user according to his needs.

IV Conclusion

Knowledge Management is of increasing interest to the construction sector. Many of the individuals tasked with implementing KM strategies have an engineering background and no previous experience in managing knowledge on a corporate or business-unit scale. They are, therefore, eager to learn from others who may be further advanced than themselves.

KM is being given a high profile throughout the construction industry. It is seen as a mechanism to support the continuous improvement being sought and complements a number of other business-improvement measures. The findings documented herein should help to provide organizations with insight into the KM activities currently being undertaken by other organizations.

This process takes a lot of time and effort. Besides, it generates a huge amount of different kinds of knowledge that must be suitably managed. MAS in charge of managing this knowledge might improve the construction organization since agents would help developers find information and solutions to problems and to make decisions, thus increasing organization's competitiveness. KMS is a good place where people could share their knowledge between the CoP. In this case, agent's technology is a tool that could be used in order to act on behalf of CoP of CO to do something repetitively and time based system. The agent techniques describe send and receive agent use to enable the user to share their knowledge among their emails. We have briefly presented the Prometheus methodology for designing our MAS. The methodology provides detailed guidance in terms of processes as well as notations. It is not intended to be prescriptive, but is rather an approach which has evolved out of experience, and which the authors expect to be further adapted, refined and developed to suit the needs of agent software developers.

V References

1) Bigus, J. P., Bigus, J. (2001) "Constructing Intelligent Agents Using Java". Second Edition 2001. Wiley Computer Publishing, John Wiley & Sons, Inc.
2) Brelade, S. and Harman, C. (2001) 'How human resources can influence knowledge management', Strategic HR Review (Melcrum Publishing, London), Vol. 1, pp.30–33.
3) Carrillo, P.M., Anumba, C.J. and Kanara, J.K. (2000) 'Knowledge management for construction: key IT and contextual issues', In G. Gudson (Ed.), Paper presented at the International Conference on Construction Information Technology (CIT 2000). Reykjavik, Iceland, June 28–30, Vol. 1, pp.155–165. In proceedings.

4) Dignum, V. (2006) "A Model for Organizational Interaction: Based on Agents, Founded in Logic". PHD Thesis, Utrecht University
5) Egbu, C., Sturgesand, J. and Bates, B. (1999) 'Learning from knowledge management and transorganisational innovations in project management environments', In W.P. Hughes (Ed.), Paper presented at the 15th Annual Conference of the Association of Researchers in Construction Management (ARCOM) John Moore's University, Liverpool, UK, September 15–17, pp.95–103. In proceedings
6) Elst,L. V., Dignum V., & Abecker A., (2004) "Towards Agent-Mediated Knowledge Management". Achmea & University Utrecht
7) Harris, K., Fleming, M., Hunter, R., Rosser, B. and Cushman, A. (1998) 'The knowledge management scenario: trends and directions for 1998–2003', Technical Report, Gartner Group.
8) Liu & Sycara. K. P. (1996). "Multiagent coordination in tightly coupled task scheduling". Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pages 164-171.
9) Lotus Company. (Access on August 2009) available at:www.lotus.com
10) Nonaka, I. and Takeuchi, H. (1995) The Knowledge Creating Company. Oxford: Oxford University Press.
11) Padgham, L., & Winikoff. M., (2002), "Prometheus: A Pragmatic Methodology For Engineering Intelligent Agents". RMIT University.
12) Park, S., & Sugumaran, V. (2005). "Designing Multi-agent Systems: A Framework and Application". Expert System with Applications 28 259-271. Elsevier Ltd.
13) Rennie, M. (1999) 'Accounting for knowledge assets: do we need a new financial statement?', Int. J. of Technology Management, Vol. 18, pp.648–659.
14) Vermeulen, S. Bohte, D., Somefun, & Poutré J. L. (2006). "Improving patient activity schedules by multi-agent Pareto appointment exchanging".
15) Wetherill, M., Rezgui, Y., Lima, C. and Zarli, A. (2002) 'Knowledge management for the construction industry: the e-cognos project', Electronic Journal of Information Technology in Construction, Vol. 7, pp.183–196. Available at :http://www.itcon.org/2002/12
16) Wooldridge, M. (2004). "An Introduction to Multi-Agent Systems" John Wiley & Sons Ltd. (Chapter 2)

Appendix A
QUESTIONNAIRE

Appendix A.1 Pre-Survey Questionnaire

Thank you very much for agreeing to participate in this experiment. All of your personal data that we collect will be entirely confidential. I would like to gather a bit of background information about u.

Participant
Name_____
Gender: _____Male _____Female
Date_____
How old are you? 20-29 30-39 40-49 50-59 60 or above
Level of education:
_____Certification Bachelor _____ Certification Diploma
_____Degree Postgraduate
Race: _____Malaysian (Local)
_____International
Years of Experience
_____

Appendix A.2 Testing Questions

The goal of this Survey to evaluate the KMS and prove the KMS is a useful support system.

I will ask you a series of questions and would like you to think out loud while you look for the answer. Please remember that we are testing the effectiveness of the KM and this is not a test of you. The whole test should take less than one hour. Thank you

Description for How to Answer the Question:

Evaluation of the matrix: Assign yourself the following points for each

NA = 0, where 0 is doing nothing at all = NONE and

1 = Don't Know, Not Sure or Can't Say = NO

2 = Not Important or as Not been Addressed = MINIMALLY

3 = Partially Beneficial or somewhat Effective or Less Scope for Overall Improvement = PARTIALLY

4 = Important or May not be effective but other associated necessary actions being taken =SUBSTANTIALLY

5 = Critical or already in place and effective = FULLY

Also, the scale can generally be summarized as follows for majority situations

'NA 1 2 3 4 5' is calibrated as in

'5 (Always) 4 (Often) 3 (Sometimes) 2 (Occasionally) 1 (Never)'

NA (Not Applicable), (Note: "NA" and "1" scale values are equivalent.)

QUESTIONNAIRE - Part One (Quantitative Analysis)

1.Is recording and sharing knowledge a routine and like any other daily habits for the employees?

NA 1 2 3 4 5

2. Are the employees co-operative and helpful when asked for some information or advice?

NA 1 2 3 4 5

3. Is Knowledge sharing seen as strength and knowledge hoarding as a weakness?

NA 1 2 3 4 5

4. Is good knowledge management behavior like sharing, reusing knowledge actively promoted on a day-to-day basis?

NA 1 2 3 4 5
5. Are people in the organization aware of the need to proactively manage knowledge assets?
NA 1 2 3 4 5
6. Do people at all levels in the organization participate in some kind of a community or communities of practice?
NA 1 2 3 4 5
7. Is there top management representation for KM?
NA 1 2 3 4 5
8. Is knowledge management a formal function area in the organization?
NA 1 2 3 4 5
9. Are the teams in the organization effective? Are self managed teams composed of individuals capable of learning from each other?
NA 1 2 3 4 5

QUESTIONNAIRE- Part Two (Qualitative Analysis)

Do the employees share their knowledge?
Yes No
2. Is the intranet used to share knowledge in an informal manner (non-routine, personal and unstructured way)?
Yes No
3. Do workplace settings and format of meetings encourage informal knowledge exchange?
Yes No
4. Are there incentives given for knowledge contribution, exchange or on knowledge sharing in your firm?
Yes No
5. Is the support from executive management to KM (Knowledge Management)\ knowledge sharing VISIBLE?
Yes No
6. Are there specific knowledge roles identified and assigned?
Yes No
7. Are all senior managers and professionals trained in knowledge management techniques?
Yes No
8. Is knowledge validated through peer or superior review or, is there some kinds of librarians or information management staff that coordinate knowledge repositories.
Yes No
9. Is knowledge sharing across departmental boundaries actively encouraged? (Not similar to ‘’incentives’’)
Yes No

Appendix A.3 Post-Survey Questionnaire

Thanks again for participating in this experiment. This questionnaire gives you an opportunity to tell us your reactions to the system you used. Please circle a number on the scale to indicate your reactions. Thank you
The goal of this part to evaluate the MAS that applying into the Lotus Notes Domino and to prove the MAS will help the users according to their needs.

QUESTIONNAIRE - Part One (Quantitative Analysis)
1.Is it possible to change the send and receive agent schedule.
NA 1 2 3 4 5
2. We can run the send and receive agent "After new mail arrives" and "Before new mail arrives".
NA 1 2 3 4 5
3. Send and receive agent option will appear in the current mail file.
NA 1 2 3 4 5
4. One of our users left the office without enabling the send and receive agent. We can enable it for him or her.
NA 1 2 3 4 5
5. I sent to someone multiple e-mails while that person is out of the office. So I will receive only one e-mail notification.
NA 1 2 3 4 5
6. To customize the "Welcome Back" message, the "Disable Reminder" message, or the default wording of the e-mail notifications sent to all senders of e-mail.
NA 1 2 3 4 5
7. In order to notice the Domino Designer 5 client has new agent properties, such as "Allow user activation" and "Run on behalf of." The both of these we need to set in the mail template (on the server) or in the individuals' mail files for the send and receive agent to work properly.
NA 1 2 3 4 5
8. The send and receive agent work in a clustered environment.
NA 1 2 3 4 5
9. We can enable the scheduler agent for leaving "Today" instead of the recommended "Tomorrow" or another date in the future.
NA 1 2 3 4 5

QUESTIONNAIRE- Part Two (Qualitative Analysis)

1. We can set the scheduler agent for an absence period of a half day or a few hours.
Yes No
2. Whenever we receive a warning in Designer while attempting to save an agent
"You do not have execution access privileges for this agent on server ". This indicates one of two things: either the agent signer does not have the rights on the scheduled server, or that server is not reachable to check the signer rights. Running agent "test" in the Designer will give you a better indication.
Yes No
3. "Do you know why I get 'Object variable not set'?" This is a result of a logic error in the code. The problem should become clear if you single step through the code in debugger (File - Tools –Lotus Script debugging). Server might be configured to delay execution of your agents.
Yes No
4. If these tips don't help you figure it out on your own, when you post in the forum please include in your post screen shot of server log output with agent manager debug flags set to '*' (best) and/or diagnostic output of "agent test"

(a good second choice when you don't have access to the server log).
Yes No
5. It is possible to pass parameters between agents.
Yes No
6. It is easy to sign an agent with a server.id For Lotus Notes 5.
Yes No
7. It is easy to console commands from send and receive and scheduler agent.
Yes No
8. Agents runs but mail is not being sent. If our agent runs to completion (i.e. no run time errors that stop the agent before it gets to the send logic) this symptom usually means that it is configuration issue, not an agent problem.
Yes No
9. Does the agents that applied will help the users of the system?
Yes No
Comments about the system:

Identification of Methodology for Analysis of the Risk Factors in Software Development Environment

GJCST Computing Classification
D.2.6, D.2.11 & D.2.1

Abdullah S. Al-Mudimigh
Basit Shahzad, , Zahid Ullah
College of Computer & Information Science, King Saud University, Saudi Arabia
mudimigh@ksu.edu.sa basit.shahzad@gmail.com, , zahid@ksu.edu.sa

***Abstract*- Software engineering has attracted the recent focus of academia and researchers by providing them means of effective software development. The effective risk management has also played a vital role is making the software development practices more reliable and organized. Ample consideration is being given to the software risk analysis and that has enabled the more reliable software management. With emergence of the need for managing the risks in software, it is essential that suitable methodology be identified for identifying the risks. This paper, this way, works as a reverse engineering approach for the identification appropriate methodology for identifying the risks in software risks.**

Keywords- Risk identification, Risk priority, Risk emergence, software failures

I Introduction

The management of risks in software is not as old as the software development itself is. The software industry, after the orientation of the need for the documentation for any software developed, came to know that certain risks are faced during the software development lifecycle. Coper Jhons [25] in his survey in 1996, provided a detailed information about the failure of the projects and concluded that the projects or either delayed or fail due to the poor risk management techniques in practice. The delayed or a failed project does not only mean a failure of that project specifically but it also means the revenues and the reputation of the development firm also comes under question, and the development team has to strive hard to re-gain its market position: both in terms of revenues and reputation. In the recent past, software risk management has gain the attention specially and much contribution has become from the academia to help in the identification [15], management [16] and prioritization [17] of the software risk factors.

Many research factors have been investigated and identified in recent past and the race of identification is still on and certainly with more passion. However, the time demands a smart work perhaps not the hard work. It is therefore can be considered that instead of being a part of the identification race; why not provide help to the world with a way to more easily identify the risk factors. For this purpose, in this paper, we will discuss already identified risk factors, and will see that to which strategy they match perfectly for the purpose of identification and this process may be repeated for many risk factors available till now to see the most appropriate technique for their identification. Concluding the paper, a comparison will be made to show the most effective technique and suggestions will be made to use that technique to gain the maximum benefits.

II Framework of Methodologies

In this section, we discuss the working range of each methodology, in order to establish the effectiveness of each methodology. All three methodologies are discussed for this purpose.

i. Questionnaire (Q): Questionnaires are used when an opinion is to be gathered from the public or a group of people from different localities. Questionnaires are generally not descriptive and just provide possible options to choose from.
ii. Direct Communication (DC): DC has different meanings in different circumstances, e.g. DC means communication between the risk manager and development team while at some other point DC would reflect the meanings of communication between the risk/project manager and the customer, and sometimes it may be from risk/project manger to the management of the organization itself.
iii. Experience & Knowledge (EK): Many risks can only be identified by recalling the successes and failures in the previous projects. The intuition can work as a magic and can only not help in identification of risk factors but also in the effective management of the identified risks by incurring least resources.

III Identification Techniques & Methodology

Identification of software risks is such a dynamic and comprehensive activity that it cannot be performed by either an individual or a specific department. Risk identification, infact is governed by the mesh of activities taking place during the software development life cycle in the entire organization. To be precise, the identification of risks in an ongoing project is not limited to the current activities only, but a huge amount of other continuing

factors contribute in making the risk factors present and evident.

In order to propose some methodologies for the identification of risk factors a detailed literature survey was conducted. [22] Has considered that the management of risk registers is suitable for the identification of software risk, while [23] believes in categorizing the risks for the purpose of identification. But the categorization itself is not possible without the identification, so this approach somehow produces a deadlock in identification and categorization. Joe Hennessy at ISD-NASA (2004), in his report on ISD software risk identification has discussed the process of risk identification and has urged that before the start of risk identification process the risk management team must be armed with the list of risk already identified in the domain including technical, budget and management risks etc. Joe focuses on the discussion of each identified risk with the development team and decides an applicability of that risk in the project under consideration. Further the development team may check that either some generic risk factor that apply to the organization apply to the specific project or not? Thus contributing to develop a complete list of risk factors that is relevant to the project. Joe has precisely emphasized the need for communication between the manager, customer, and development team. The process of identification and handing of software risk factors proposed by Joe is consistent with NPR 8000.4: Risk management procedural requirements, GPG 7120.4: Risk management, GPG 8700.5: In-house development and maintenance of software projects and IEEE std 1540-2001: Standard for software life cycle processes-risk management. [17]

In another report on "taxonomy-based risk identification" by Software Engineering Institute: the method described in the report consists of developing the taxonomy-based questionnaires for identification of software risk. Taxonomy is a scheme that partitions the body of knowledge and defines the relationship among the pieces [18]. This report emphasis the need of questionnaires for identification of risk factors.

In anther report by Software Engineering Institute (2008), the author Ray C. Williams has argued that the biggest need in managing the risks is risk identification. Ray has referenced a situation in which 40 field tests were conducted with a broad range of software developers to identify that who have good communication skills and techniques to help in the process of risk identification with their own experience and by interviewing others. Ray also argues in the favor of using the "inter-organization-communication" to report any risk that is observed at any level. He suggests that the higher management must be open with the middle management and workers by sharing the risks and inviting them to share theirs.

In yet another paper on checklist of risk identification by Mark Li (2007), different milestones of risk identification with the identification methodology has been described. Boehm (1991) identified 10 risk factors by survey of the experienced risk managers [19]. Barki et al. (1993) identified 23 risk factors by just doing the systematic literature review [20]. Heemstra and Kusters (1996) identified 36 risk items by doing the literature survey combined with experiences [21]. Moynihan (1997) identified 21 risk factors by Interviewing with 14 application developers [22]. Ropponen and Lyytinen (2000) identified six risk items by doing a survey of 83 project managers covering nearly 1100 projects[23]. Han and Huang (2007) identified six dimensions of 27 risks by an analysis of 115 software projects [8].The literature survey proposes that the risks, whatever they are, can be identified by doing the questionnaire and enhancing the inter-organization-communication, once the initial list of relevant risk factors has been identified. Although some other factors like intuition, relevance, etc can also be used for risk identification yet they are neither essential nor universal enough to be learnt as established risk identification methodologies. Therefore we consider it essential to restrict this work to three most relevant and most referenced identification methodologies, namely: Questionnaire (Q), Direct Communication (DC) and Experience & Knowledge (EK) as they have been observed to be necessary and sufficient for identification of any risk factor also these three approaches have been in wide use for the identification of risk factors in the leading academic and commercial environments throughout the world.

The table 1 describes different risks identified from the literature study. In the proceeding section each risk factor is considered separately, and all three methodologies (Q, DC, EK) are applied and observed that which technique can best generate /identify this risk factor. A Suitability Index (SI) is determined in this regard. The SI value of 1 shows that some risk factor e.g. x can only be generated by some specific methodology e.g. EK, which also mean that other two techniques have no contribution in the identification of that risk factor at all. The SI value 0.05 means that the identification of a risk factor e.g. y is just minutely dependent on one methodology i.e DC, and the rest 0.95 is covered by either EK or Q. likewise the SI value of 0.0 for any methodology means that this methodology has got absolutely no role to play in the identification of the risk factor under consideration. In the proceeding section we use a scale of values ranging from 0 to 1 with a multiple of 0.05. The higher the value against a specific methodology the greater the role it will possess in identification of a specific risk factor.

It is also worth mentioning that the values provided against each methodology have been derived from the survey conducted by the author and also by the author's continues experience in the domain of software risk identification and management, the author has a vast experience and contribution in this domain of knowledge [10, 11, 12, 13, 14, 15, and 16].

Having repeated this exercise for all risk factors, the total SI count is calculated for each methodology among all risk factors, and the one with highest count is declared as most suitable risk identification methodology. In order to identify the appropriate strategies for identification of risk factors, a survey was conducted on 100 individuals

working in academia, management and software development across different countries in the world including Saudi Arabia, Pakistan, Sweden, Denmark and Malaysia. The survey form is shown in Table 1. For the convenience of responders, the survey has been designed in a way to require minimum effort from the respondendent.hey are just required to tick the choice(s) that they feel appropriate.

IV SURVEY DESIGN AND CONDUCTING METHODOLOGY

The risk factors from 1-26, in the table 1, have been identified by the comprehensive study of literature [12, 13, 14, 19, 20]. In most cases the source of identification was not known. However, it has been observed that most of the risk factors can be identified by using experience, direct communication between stake holders of the software or by using the questionnaire for information acquisition. The 26 most frequently used risks in software environment have been used to perform the reverse engineering to identify the methodology by which they can be identified. For this purpose, a questionnaire consisting of all 26 risk factors, with a choice of three possible answers was designed. The respondents were requested to tick the most suitable methodology that, in their opinion, can best identify that risk factor. The respondents were also free to choose both or all three methodologies to demonstrate that some specific risk factor can't be identified by using only one or two methodologies. The results of the survey are discussed in section 6 individually and in section 8 as a summary.

Risk Identification Survey

Please check as appropriate, where EK=Experience and knowledge, Q=Questionnaire, DC=Direct Communication. This is to establish how the risk factors can be identified more adequately.

Risk ID	Risk Title	EK	Q	DC	Justification (if any)
1	Inadequate requirement description				
2	Project size estimation				
3	Funding loss/uncertainty				
4	Staff inexperience				
5	Staff Turnover				
6	Management changes circumstances				
7	Loss of actual documents and data				
8	Low estimation of time				
9	Low estimation of cost				
10	Lack of intuition				
11	Developer's lack of commitment				
12	Customer's dis-satisfaction				
13	Change in the hardware defaults				
14	Develop mate team's continues work				
15	Requirement's postponement				
16	Immature coding practices				
17	Presence of bugs and errors				
18	Over-acceptability of product and insufficient data handling				
19	Hackers, viruses and Trojan horse etc				
20	Delayed implementation suffering				
21	Market denial				
22	Over estimation about workers skills				
23	Lack of technical feedback				
24	Save prestige not money				
25	Economic distortion				
26	Building loss/fire				

- We encourage you to identify any other risk factor that is not already included in the list.

Designation:_____, Organization:_____

Table 1: Survey for choosing the identification methodology

V WEIGHTAGE OF THE METHODOLOGY

It has been observed that all three methodologies used for identification of risk factors in a software development environment have got certain pros and cons which are built-in with the methodology itself, e.g. the methodology 'EK' has the build-in advantage of saving time and resources and as only experienced people practice this methodology, the probability of its appropriate use is very high, and therefore proper results can be derived by using this methodology.

The 'DC' methodology is a bit slower is suggesting the response, as the project manager has to discuss the details about the project circumstances before a decision can be made, therefore the response becomes slower and some resources are to be invested to get the response. It also to be noted that the respondents of a 'DC' take this time out of their normal schedule in which they are supposed to perform some other duties as well.

The 'Q' technique suffers from the built-in disadvantage of being least explanatory and most time expensive, also the response gained from the survey is normally delayed. A survey can only be recommended if it is performed in adequate time limits required by the software development deadlines. In light of above discussion, a waitage index is suggested keeping in view the build-in pros and cons of each methodology

Method	Person hour	Schedule disturbance	waitage
EK	5	N	2.00
DC	10	Y	1.00
Q	20	Y	0.50

Table2: Weightage Index of risk identification methods

In table 2, we use a situation in which the project manager, by using his experience and knowledge suggests the solution of a problem, this single-handed effort is supposed to take 5 hours and yet not disturbing any other duty of the manager. Therefore this emerges to be the most appropriate solution and hence given the highest weightage of 2.0. If the manager determines to use services of 5 other individuals to identify the expected risk factors in software development environment, the solution will become costly and hence time consuming as well. Therefore, if the team can complete the problem in 2 hours, the organization will cost 10 person hours and the normal duties of the development team will be disturbed. As this is second most appropriate solution, the weightage index suggested for this methodology is 1.0, while weightage index for questionnaire methodology is calculated to be 0.5 as it is slowest and most resource consuming in terms of identification of any expected risk factor.

VI PROBABILISTIC IDENTIFICATION OF EACH RISK FACTOR

The approach focuses on the identification of each risk factor and it is argued that which technique, out of available three, can be most beneficial in identification of the risk

factor. The approximate weightage is being given to each methodology based on its suitability, which is supported by the survey conducted in this regard and the author's intuition.

A. *Inadequate Requirement Description*

It is often known to the project manager that the customers can hardly describe the adequate amount of information about their requirements. Although the manager can identify with its experience that the requirements are incomplete and tend to change in future, yet the overall measure to which the requirements are missing and can change can only be known through by using the DC. Questionnaire, in this regard can be of a very little help, hence, the weightage of each methodology, as per survey, will be as follows:

Method	Suitability Index	WI	Suitability Weighted Index
EK	0.3	2	0.6
DC	0.6	1	0.6
Q	0.1	0.5	0.05

Table 3: Suitability Weighted Index for risk factor No.1

B. *Project Size Estimation*

Calculating the actual size of the project under consideration has been a serious question in the software cost estimation domain. It has been observed that the questionnaire can be of very little help in this regard not only because of the general irresponsiveness but also because of the natural disability of less descriptive. DC with the customer and within the development teams helps a lot in identifying the actual project size. The 'experience' plays a vital role in identification of the actual project size, and without experience other two methodologies tend to fail badly, hence, the weightage of each methodology's SI will be as follows:

Method	Suitability Index	WI	Suitability Weighted Index
EK	0.5	2	1.0
DC	0.4	1	0.4
Q	0.1	0.5	.05

Table 4: Suitability Weighted Index for risk factor No.2

C. *Project Funding Loss*

Due to the inadequate handling of the project in the beginning, or for any other reason the project may not meet the milestones and consequently delivery deadlines cannot be met. An effective project manager can, by using his experience promptly, very effectively predict about the development delays and can propose extra measures in achieving the milestones. Questionnaire oriented information gathering regarding this risk factor has been extremely un-helpful. Such risk can only be identified either by experience or mainly because of the DC; hence, the weightage of each methodology in SI will be as follows:

Method	Suitability Index	WI	Suitability Weighted Index
EK	0.35	2	0.70
DC	0.6	1	0.6
Q	0.05	0.5	.025

Table 5: Suitability Weighted Index for risk factor No.3

D. *Staff Inexperience*

As expertise and experience of the individuals working in an organization are known to the management, therefore, the most effective way of finding the expertise of individuals is through the DC. Questionnaires have been found to be of least usefulness because of their descriptive nature and immaturity. The developers also may not like to provide the written proof about their deficiencies, etc. Experience also plays a vital role in the identification of any such risk factor;, the weightage of each methodology in SI will be as follows:

Method	Suitability Index	WI	Suitability Weighted Index
EK	0.35	2	0.70
DC	0.50	1	0.50
Q	0.15	0.5	.075

Table 6: Suitability Weighted Index for risk factor No.4

E. *Staff Turnover*

Staff turnover is one of the most dynamically faced challenge not only in the software development organizations but also in general as well. The change of job can hardly be evaluated by the questionnaire. DC can be of help only when the employee has shown its intentions in advance to leave the job. The experienced managers, however, can estimate and expect some staff turnover during the lifetime of the project. In estimating the staff turnover, nothing has been found more appropriate than the experience which allows the managers to plan ahead and train and attach some extra workforce with the project. , The weightage of each methodology in SI will be as follows:

Method	Suitability Index	WI	Suitability Weighted Index
EK	0.7	2	1.4
DC	0.25	1	0.25
Q	0.05	0.5	.025

Table 7: Suitability Weighted Index for risk factor No.5

F. *Management Changes Circumstances*

Change of circumstances to meet the deadlines is considered normal when the requirements are deficient in exploration at the beginning of the project. The change in requirements directly effects the time and budget allocated for the project, in order to cope with this the manager needs to change circumstances to accommodate the changes. This

risk factor is quite evident and its existence can be established either by the questionnaires or by communicating to the manager and customers directly. Experience and intuition still play a vital role in the identification of any such risk factor. The weightage of each methodology in SI will be as follows:

Method	Suitability Index	WI	Suitability Weighted Index
EK	0.5	2	1.0
DC	0.25	1	0.25
Q	0.25	0.5	.125

Table 8: Suitability Weighted Index for risk factor No.6

G. Loss Of Actual Documents And Data

The loss of documents and data is not a common risk factor in the software development life cycle. Loss of data and document can be for any reason, including the theft, fire, loss etc. An experienced manager can have ample wisdom about the risk and maintains the data on multiple sites and servers and duplicate copies of documents are also maintained. DC has also a major role to play in the identification of this risk factor. The questionnaire methodology has been observed to be of least significance among the three; hence, the weightage of each methodology in SI will be as follows:

Method	Suitability Index	WI	Suitability Weighted Index
EK	0.60	2	1.20
DC	0.15	1	0.15
Q	0.25	0.5	.125

Table 9: Suitability Weighted Index for risk factor No.7

H. Low Estimation Of Time

Due to the inbuilt and perhaps genuine problem of the requirement statement by the customer, the development team remains in continuous loop for a fairly longer period to time to finalize the requirements and based on that the time and budget for the accomplishment of the project are also calculated. Questionnaire's approach may be of slight help but takes a huge amount of time and hence stands less adequate in the race of being the fittest. DC allows the manager to communicate with the development team and the customer to manage this risk. It has been observed that an experienced manager will already know that this risk can come and he can identify such risks with the experience, hence, the weightage of each methodology in SI will be as follows:

Method	Suitability Index	WI	Suitability Weighted Index
EK	0.55	2	1.1
DC	0.35	1	0.35
Q	0.1	0.5	.05

Table 11: Suitability Weighted Index for risk factor No.9

I. Lack Of Intuition

Use of intuition plays a major role in smelling the out of box problems and can suggest the possible solutions. The lack of intuition may mean that a development team works more and yields less. In order to make the team productive, it is necessary that they are advised to learn from experience, use the re-usable code, be coherent with the circumstances and also keep their efforts synchronized. The lack of intuition must be identified by the higher management and when identified should be immediately in place. Although the identification of this risk factor can be done well by the experience and direct communication methodologies, yet the questionnaire methodology has been of adequate importance in the identification of this risk factor, the weightage of each methodology in SI will be as follows:

Method	Suitability Index	WI	Suitability Weighted Index
EK	0.45	2	0.90
DC	0.30	1	0.30
Q	0.25	0.5	0.125

Table 12: Suitability Weighted Index for risk factor 10

J. Developer's Lack Of Commitment

The project starts with a positive node assuming that the work force deployed on the project is loyal, motivated, and committed but sometimes the situation may be otherwise. It is of utmost importance that the roles of each individual are discussed before they are assigned. This can help in keeping the developers committed to their work. As this risk factor is evident only after the start of the project, a good manager can identify this kind of risk before the start of the project. Such risk factors can be identified either by using the experience or by direct communication but not by questionnaire's methodology by any means as an un-committed developer will not like admitting about its lack of commitment, hence, the weightage of each methodology in SI will be as follows:

Method	Suitability Index	WI	Suitability Weighted Index
EK	0.7	2	1.40
DC	0.25	1	0.25
Q	0.05	0.5	0.025

Table 13: Suitability Weighted Index for risk factor 11

K. Customer's Dis-Satisfaction

With the emergence of agile computing and prototype models of software development the customers role as an active entity have increased gradually, over the time. Customers now have the liberty to show the consent about the development under consideration. Keeping in view, that a dis-satisfied customer may cause the funding uncertainty the manager can use the questionnaire or direct communication to get the feedback of the customer and can elaborate on that by using his experience, hence,

Method	Suitability Index	WI	Suitability Weighted Index
EK	0.35	2	0.70
DC	0.4	1	0.4
Q	0.25	0.5	0.125

Table 14: Suitability Weighted Index for risk factor 12

L. Change In Hard-Ware Defaults

This risk factor is more common for the products that are not developed for some specific customer but for the public. The development team must ensure that they do not waste (take) a huge amount of time to develop the product, so that the hardware defaults do not change when the product becomes available to use. This has to be done with immense speed, as the hardware defaults are changing dynamically, presently. The manager of the development team may not have the liberty to use the direct communication with the firms developing the hardware. Thus, generally, the manager has to rely either on his own experience or on the questionnaire that may contain the probabilistic questions to forecast about the future development in the hardware defaults. Even if such changes in the hardware are known, the development team may not easily adopt them with the same pace. Only the questionnaire and experience methodologies can be used hence, the weightage of each methodology in SI will be as follows:

Method	Suitability Index	WI	Suitability Weighted Index
EK	0.75	2	1.50
DC	0	1	0
Q	0.25	0.5	0.125

Table 15: Suitability Weighted Index for risk factor 13

M. Development team continuous work

This risk factor originates either because of the change in the requirements or because of the continuous business of the organization. It has been observed that the software developers have to work more than their fixed timings in order to meet the deadlines. It is therefore important that sufficient manpower is placed to e ensure that each employee works for not more than 4o hours a week, in normal circumstances. As the future business of the organization can hardly be forecasted, therefore the direct communication and questionnaire can help in identifying this risk factor hence, the weightage of each method in SI will be :

Method	Suitability Index	WI	Suitability Weighted Index
EK	0.05	2	0.1
DC	0.6	1	0.6
Q	0.35	0.5	0.175

Table 16: Suitability Weighted Index for risk factor 14

N. Requirement Postponement

Requirement gathering is difficult because of the fact that the customer may not explicitly mention what he needs. In such case, where the requirements are hard to find it is properly inadequate to postpone the gathered requirements. The idea of postponement comes only when the development team tries to make the customer happy by showing him something instead of the complete working product. Being up-to-date with the project scope and milestones the project manager can directly communicate with the development team and customer to see if some requirements can be scrubbed. An experienced manager can also prepare a questionnaire for the customer to see which requirements can be postponed, if any. Therefore, the weightage of each methodology in SI will be as follows:

Method	Suitability Index	WI	Suitability Weighted Index
EK	0.3	2	0.60
DC	0.35	1	0.35
Q	0.35	0.5	0.175

Table 17: Suitability Weighted Index for risk factor 15

O. Immature Coding Practices

The implementation being the core of the project requires more attention as compared to any other phase in software development. In order to ensure that the development team is doing the coding accurately, purposefully and error free, it is necessary that suitable coding practices are introduced in the organization. The employees may be trained and test for having the adequate standards of software development. If the coding standards are inadequate, the manager has to know this much earlier otherwise the failure or at least delay of the project is guaranteed. Direct communication, Questionnaire, and experience are all important to identify this risk factor respectively. Hence, the weightage of each methodology in SI will be as follows:

Method	Suitability Index	WI	Suitability Weighted Index
EK	0.35	2	0.70
DC	0.40	1	0.40
Q	0.25	0.5	0.125

Table 18: Suitability Weighted Index for risk factor 16

P. Presence Of Bugs And Errors

It is adequately important that the developer unit test each module and piece of code that they have developed, in order to reduce the chances of errors at later stage. More lately an error is identified, the cost to rectify will be higher. Direct communication and questionnaires can help in the identification of these risk factors while the experience can help in rectifying the identified errors. Hence, the weightage of each methodology in SI will be as follows:

Method	Suitability Index	WI	Suitability Weighted Index
EK	0.1	2	0.20
DC	0.65	1	0.65
Q	0.25	0.5	0.125

Table 19: Suitability Weighted Index for risk factor 17

Q. Over-Acceptability Of Product And Insufficient Data Handling

Innovations are generally appreciated in any domain. Sometimes, the product developed by a firm is overwhelmingly welcomed in the market and hence the stress on application increases both: in terms of access rate and in terms of data storage. If any such application has been publicized with a limited storage and inadequate response time the likelihood for the application crash will increase. The manager, while developing the product must also idealize about the overwhelming success of the project. The manager must try to provide as much functional facilitations as possible by not disturbing the efficiency and reliability of the system. As such situations are not common, experience of manager may not be of adequate help. Therefore this risk factor can better be identified by using the direct communication and questionnaires. The weightage of each methodology in SI will be as follows:

Method	Suitability Index	WI	Suitability Weighted Index
EK	0.05	2	0.10
DC	0.6	1	0.60
Q	0.35	0.5	0.175

Table 20: Suitability Weighted Index for risk factor 18

R. Hackers, Viruses And Trojan Horse Etc

The testing team must ensure that the system implemented is error free and must ensure that a mechanism is in place to restrict any friendly or unfriendly program to access the system without permission. The manager may also contribute to provide the updated versions of anti-viruses to ensure the maximum safety against any such event. Although this risk can more easily be identified with experience, yet the orientation of this risk factor is also possible through direct communication with development team and also by providing the questionnaires to reply accordingly. The weightage of each methodology in SI will be as follows:

Method	Suitability Index	WI	Suitability Weighted Index
EK	0.5	2	1.0
DC	0.35	1	0.35
Q	0.15	0.5	0.075

Table 21: Suitability Weighted Index for risk factor 19

S. Delayed Implementation Suffering

Software requirements are not easy to determine and determined requirements must be implemented without any delay. A delayed requirement implementation makes the job of the development team difficult and consumes extra resources. Every experienced manger is aware of the problems that can be faced because of the delayed implantation. The weightage of each methodology in SI will be as follows:

Method	Suitability Index	WI	Suitability Weighted Index
EK	0.8	2	1.60
DC	0.15	1	0.15
Q	0.05	0.5	0.025

Table 22: Suitability Weighted Index for risk factor 20

T. Market Denial

Market denial has been a serious issue in the product development. The company's management must have done adequate study about the acceptability of the product in the market before the actual work on the product starts. The complete or partial market denial after the competition of a product can suffer the business and market reputation of an organization. The organization can be in direct communication with the market or can put a survey to identify the acceptability of a specific product. The experienced manager can also use its intuition in this regard. The weightage of each methodology in SI will be as follows:

Method	Suitability Index	WI	Suitability Weighted Index
EK	0.25	2	0.50
DC	0.60	1	0.60
Q	0.15	0.5	0.075

Table 23: Suitability Weighted Index for risk factor.21

U. Over Estimation About Workers Skills

The cost to be over-optimistic is very high is software cost/time estimation. While calculating the cost and time required completing the project the analysts and managers sometime over estimate the skill of their workers and under estimate the scope of the project. This leads to a huge failure as the movement for developing the software starts and immediately the management knows about the risk of miss-calculation and realizes about the over estimation about the workers skill. Worker's skills are generally known and can be further tested either through questionnaire or by communicating directly. The weightage of each methodology in SI will be as follows:

Method	Suitability Index	WI	Suitability Weighted Index
EK	0.3	2	0.60
DC	0.45	1	0.45
Q	0.25	0.5	0.125

Table 24: Suitability Weighted Index for risk factor 22

V. Lack Of Technical Feedback

The requirement gathering process requires a thorough consideration and effective communication at the level of team leader/analyst and technical people at the customer side. The head of organization must not sign a contract without consulting his technical team to minimize the chance of reduction in profit. The development team must try to cover all requirements in the first iteration and not to leave any requirements unaddressed. It has been observed that more requirements identified in the beginning leads to less changes in the future. By expecting only a few changes, in the future it can be expected that the project can lead to a success. The experienced manger can identify this risk either by direct communication with customer or by putting a questionnaire. The weightage of each methodology in SI will be as follows:

Method	Suitability Index	WI	Suitability Weighted Index
EK	0.5	2	1.0
DC	0.25	1	0.25
Q	0.25	0.5	0.125

Table 25: Suitability Weighted Index for risk factor 23

W. Save Prestige Not Money

It has been observed that the customer opts to get its software developed from the reputed software development firms only. The reputation of firms is decided not only based on revenues but also on the basis of the goodwill and cordial relationships that they have with other groups. A failed project not only harms the revenues of the firm but also disturbs the reputation as well. Therefore, the firms try their hard not to let a project fail and even at the cost of financial losses, they would like to save their name to maintain the reputation and goodwill of the market.

It is imperative to state that a risk should always be identified before it actually starts harming the system. Once the risk has shown his presence, it doesn't remain in isolation and invites other risk factors to make a mesh and insure the project to delay if not fail at all. So in order to continue gaining business in the future, the firm may like to develop the project successfully even by going in financial deficit. The experienced manger can identify such situation by communicating with management and customers: directly or by sending questionnaires. The weightage of each methodology in SI will be as follows:

Method	Suitability Index	WI	Suitability Weighted Index
EK	0.6	2	1.20
DC	0.25	1	0.25
Q	0.15	0.5	0.075

Table 26: Suitability Weighted Index for risk factor 24

X. Economic Distortion

The management of software development firm must try to commit advance payment from the customer if the economic situation of the country/market is not stable. In the economic crisis, the firm must try maximizing its profit and should try to provide benefits to the employees to enable them to face the poor economic situation. Although economic distortions may be difficult to identify well in time, yet the experienced managers can have adequate vision to predict such events. The manager can communicate with top management and customers to identify the economic distortion. The weightage of each methodology in SI will be as follows:

Method	Suitability Index	WI	Suitability Weighted Index
EK	0.6	2	1.20
DC	0.25	1	0.25
Q	0.15	0.5	0.075

Table 27: Suitability Weighted Index for risk factor 25

Y. Building Loss/Fire

The firm must ensure that the working environment across the organization conducive and safe for the employees. Proper smoke detectors and fire alarms must be installed in the building to detect the fire and the emergency exit should be provided .The organization must also ensure that the building codes have been followed and the structure is according to the prescribed standards. The management must keenly observe the building structure and the life of building must also be known by communicating with architects and management. The weightage of each methodology in SI will be as follows:

Method	Suitability Index	WI	Suitability Weighted Index
EK	0.7	2	1.40
DC	0.25	1	0.25
Q	0.05	0.5	0.025

Table 28: Suitability Weighted Index for risk factor 26

VII RESULTS

The values of Suitability Weighted Index (SWI) presented in the section 6 of this paper are summarized here. It can be observed that the cumulative value of EK methodology is 23.50, which is highest as compared to other two values. This means that the usage of EK methodology for risk identification is most suitable in terms of identifying the risks and is cost and resource effective. As an outcome of the experimental evidence presented in the section 5-6 of this paper it is highly recommended that EK be used for the identification of risks that are probable to be present in the software development environment. The DC methodology owes the value of 9.8, which is the second best among the three. The use of questionnaire, although, may be helpful in some situations, yet it is not encouraged as a methodology for identification of all risks factors. Table 29 contains the details:

Risk No.	EK	DC	Q
1	0.6	0.6	0.05
2	1.0	0.4	0.05
3	0.7	0.6	0.025
4	0.7	0.5	0.075
5	1.4	0.25	0.025
6	1.0	0.25	0.125
7	1.2	0.15	0.125
8	1.10	0.35	0.05
9	1.10	0.35	0.05
10	0.90	0.30	0.125
11	1.4	0.25	0.025
12	0.7	0.4	0.125
13	1.50	0.0	0.125
14	0.1	0.6	0.175
15	0.6	0.35	0.175
16	0.7	0.4	0.125
17	0.20	0.65	0.125
18	0.10	0.6	0.175
19	1.0	0.35	0.075
20	1.6	0.15	0.025
21	0.5	0.60	0.075
22	0.6	0.45	0.125
23	1.0	0.25	0.125
24	1.20	0.25	.075
25	1.20	0.25	0.075
26	1.40	0.25	0.075
Total	23.50	9.55	2.40

Table 29: SWI for EK, DC and Q methodology

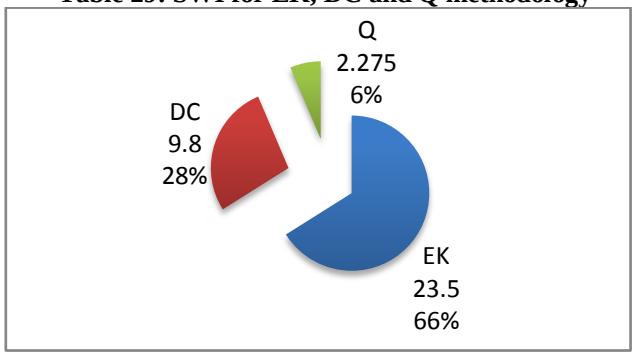


Fig. 2. SWI for EK, DC and Q methodology

From table 29, it can be observed that out of the 26 risk factors discussed in this paper, the SWI for EK methodology for 21 risk factors is highest as compared to that of the other two methodologies, while SWI for four factors is highest for Dc mythology and for one risk factor SWI for DC and EK are equal. Q methodology in none case could become the methodology of first choice as observed from table 29, although it plays a supportive role for identification of some risk factors partially.

It can therefore be concluded that the choice of methodology has a direct relation with the resources and time incurred as the consequence of that methodology. One of the reasons for the overwhelming acceptability of EK is being less expensive and highly responsive. The cost of Identification (COI) for each methodology can be identified with the formula:

$$COI=1/(\sum_{i=1}^{n} M) \quad (1)$$

Where n=26: the no. of risk factors
M=Methodology

$$COI_{EK}=1/(\sum_{i=1}^{n} EK) \quad (2)$$
$$=1/23.5$$
$$=0.0426$$

$$COI_{DC}=1/(\sum_{i=1}^{n} DC) \quad (3)$$
$$=1/9.55$$
$$=0.105$$

$$COI_{Q}=1/(\sum_{i=1}^{n} Q) \quad (4)$$
$$=1/2.40$$
$$=0.416$$

From equation 1, 2, 3 and 4 it can be observed that COI for Q methodology is highest while the COI for EK methodology is the lowest, which provides the justification for usage of EK for the purpose of risk identification.

VIII Conclusion

As per the survey results and the results of the equation 2,3 and 4, propose that the 'EK' is the most appropriate strategy for the identification of software risks. From Table 29, and equation 2, 3, 4 it can be concluded that the performance of EK methodology is more than twice better than 'DC' and even 10 times better than 'Q' methodology in terms of usage of resources, budgeting, time and effectiveness. It can therefore be concluded that the usage of 'EK' methodology for the purpose of identification of software risk factors is highly appropriate. It is worthwhile to note that only 'EK' can help in identifying the entire risk factors single handedly without the support of any other methodology. The 'DC' methodology has performed better than 'EK' in only a few identifications. It is also recommended that the reader establish sound knowledge about the risks and the methodologies before taking a final decision. But in most cases 'EK' is the default solution. So, even a random choice of 'EK" has more probability of success as compared to any other methodology.

IX Future Work

This work can be expanded to more generalized version of the risk identification model by suggesting other techniques for the identification of the risk factors and also by introducing ore adequate methodologies to identify these risk, having introduced the methodologies, the SI will become more precise and hence will be able to produce more realistic and appropriate feedback about the choice of methodology for the identification of risk factors.

X REFERENCES

1) Xiangnan Lu. , Yali Ge (2004), "Risk analysis in software development project with Owners and contractors" ", Engineering Management Conference, 2004 IEEE International Volume 2, 18-21 Oct. 2004 Page(s):789 – 793.
2) Miller J , Gorski J (2004) "Risk identification Patterns for software projects " Foundations of Computing and Decision Sciences Vol. 29, No. 1-2, pp. 115-131.
3) Schmidt, R., et al., Identifying software project risks (2001): An international Delphi study. Journal of Management Information Systems, 2001, 17(4): pp.5-36.
4) Powell, P.L., Klein, J.H. 1996. Risk Management for Information Systems Development. Journal of Information Technology, Vol 11, 309-319.
5) Linda Wallace and Mark Keil,"Software project risks and their affects on Outcomes", April 2004/vol 47 No.4 Communication of ACM
6) Risk Register http://www.brighthub.com/office/project-management/articles/3247.aspx
7) Risk categorization http://www.idrisk.com/risk_consulting.asp
8) http://www.project-management-basics.com/risk-management-4.shtml risk management approach
9) http://www.businessdictionary.com/definition/risk-identification.html risk definition
10) Kamran Khan, Zaeem ahmad, Basit Shahzad, "Adressing problems and risks in software development life cycle", 7th CIIT Workshop on Research in Computing, held at Lahore June 23, 2008, pp 123-128.
11) Basit Shahzad, Sara Safvi, "Risk mitigation and management scheme based on risk's priority", accepted in 7th WSEAS conference held in Hangzhou-China, April 6-8, 2008. pp 86-91
12) Basit Shahzad, Javed Iqbal, ""Software Risk Management – Prioritization of frequently occurring Risk in Software Development Phases. Using Relative Impact Risk Model", 2nd International Conference on Information and Communication Technology (ICICT2007), December 16-17, 2007, IBA Karchi.
13) Javed Iqbal, Basit Shahzad, "Prioritization and Handling of Commonly Occurring Risks– An Analytical Approach", 5th conference on Intelligent Systems & Networks, February 22-24, 2008(ISN-2008),Haryana, India
14) Basit Shahzad, Javed Iqbal, Zeeshan ul Haq, Saleem Raza and Salman Aslam "Distributed risk analysis using relative impact technique" , 3rd Asian Conference on Intelligent Systems and Networks, February 24-25, 2006, Haryana-India, pp 433-439.
15) Basit Shahzad, Najam us saqib, "Escape from delays and failures for software solutions", 6th Jordanian International Electronic and Electrical Engineering Conference in Amman-Jordan, March 14-16, 2006
16) Basit Shahzad, Tanvir Afzal, "Enhanced risk analysis and relative impact factorization", 1st IEEE international conference on information and communication technology, IBA Karachi, August 27-28, 2005 ,pp 290-295.
17) http://standards.ieee.org/catalog/olis/se.html
18) IEEE software engineering standards collection, IEEE-STD-610.12, 1990.
19) Boehm, B.W., 1991. Software risk management: principles and practices.IEEE Software 8 (1), 32–41.
20) Barki, H., Rivard, S., Talbot, J., 1993. Toward an assessment of software development risk. Journal of Management Information Systems 10 (2),203–225.
21) Heemstra, F.J., Kusters, R.J., 1996. Dealing with risks: a practical approach. Journal of Information Technology 11 (4), 333–346.
22) Moynihan, T., 1997. How experienced project managers assess risk. Software, IEEE 14 (3), 35–41.
23) Ropponen, J., Lyytinen, K., 1997. Can software risk management improve system development: an exploratory study. European Journal of Information Systems 6 (1), 41–50.
24) Han, W.-M., Huang, S.-J., 2007. An empirical analysis of risk components and performance on software projects. journal of Systems and Software 80 (1), 42–50.
25) Source: Patterns of software failures and successes, Capers Jones, 1996

Global Journals Guidelines Handbook 2010

www.GlobalJournals.org

FELLOW OF INTERNATIONAL CONGRESS OF COMPUTER SCIENCE AND TECHNOLOGY (FICCT)

- FICCT' title will be awarded to the person after approval of Editor-in-Chief and Editorial Board. The title 'FICCT" can be added to name in the following manner e.g. **Dr. Andrew Knoll, Ph.D., FICCT, Er. Pettor Jone, M.E., FICCT**
- FICCT can submit two papers every year for publication without any charges. The paper will be sent to two peer reviewers. The paper will be published after the acceptance of peer reviewers and Editorial Board.
- Free unlimited Web-space will be allotted to 'FICCT 'along with subDomain to contribute and partake in our activities.
- A professional email address will be allotted free with unlimited email space.
- FICCT will be authorized to receive e-Journals - GJCST for the Lifetime.
- FICCT will be exempted from the registration fees of Seminar/Symposium/Conference/Workshop conducted internationally of GJCST (FREE of Charge).
- FICCT will be an Honorable Guest of any gathering hold.

ASSOCIATE OF INTERNATIONAL CONGRESS OF COMPUTER SCIENCE AND TECHNOLOGY (AICCT)

- AICCT title will be awarded to the person/institution after approval of Editor-in-Chef and Editorial Board. The title 'AICCTcan be added to name in the following manner:
  eg. **Dr. Thomas Herry, Ph.D., AICCT**
- AICCT can submit one paper every year for publication without any charges. The paper will be sent to two peer reviewers. The paper will be published after the acceptance of peer reviewers and Editorial Board.
- Free 2GB Web-space will be allotted to 'FICCT' along with subDomain to contribute and participate in our activities.
- A professional email address will be allotted with free 1GB email space.
- AICCT will be authorized to receive e-Journal GJCST for lifetime.
- A professional email address will be allotted with free 1GB email space.
- AICHSS will be authorized to receive e-Journal GJHSS for lifetime.

ANNUAL MEMBER

- Annual Member will be authorized to receive e-Journal GJCST for one year (subscription for one year).
- The member will be allotted free 1 GB Web-space along with subDomain to contribute and participate in our activities.
- A professional email address will be allotted free 500 MB email space.

PAPER PUBLICATION

- The members can publish paper once. The paper will be sent to two-peer reviewer. The paper will be published after the acceptance of peer reviewers and Editorial Board.

Process of submission of Research Paper

The Area or field of specialization may or may not be of any category as mentioned in 'Scope of Journal' menu of the GlobalJournals.org website. There are 37 Research Journal categorized with Six parental Journals GJCST, GJMR, GJRE, GJMBR, GJSFR, GJHSS. For Authors should prefer the mentioned categories. There are three widely used systems UDC, DDC and LCC. The details are available as 'Knowledge Abstract' at Home page. The major advantage of this coding is that, the research work will be exposed to and shared with all over the world as we are being abstracted and indexed worldwide.

The paper should be in proper format. The format can be downloaded from first page of 'Author Guideline' Menu. The Author is expected to follow the general rules as mentioned in this menu. The paper should be written in MS-Word Format (*.DOC,*.DOCX).

The Author can submit the paper either online or offline. The authors should prefer online submission.

Online Submission: There are three ways to submit your paper:

(A) (I) Register yourself using top right corner of Home page then Login from same place twice. If you are already registered, then login using your username and password.

(II) Choose corresponding Journal from "Research Journals" Menu.

(III) Click 'Submit Manuscript'. Fill required information and Upload the paper.

(B) If you are using Internet Explorer (Although Mozilla Firefox is preferred), then Direct Submission through Homepage is also available.

(C) If these two are not convenient, and then email the paper directly to dean@globaljournals.org as an attachment.

Offline Submission: Author can send the typed form of paper by Post. However, online submission should be preferred.

MANUSCRIPT STYLE INSTRUCTION (Must be strictly followed)

Page Size: 8.27" X 11'"

- Left Margin: 0.65
- Right Margin: 0.65
- Top Margin: 0.75
- Bottom Margin: 0.75
- Font type of all text should be Times New Roman.
- Paper Title should be of Font Size 24 with one Column section.
- Author Name in Font Size of 11 with one column as of Title.
- Abstract Font size of 9 Bold, "Abstract" word in Italic Bold.
- Main Text: Font size 10 with justified two columns section
- Two Column with Equal Column with of 3.38 and Gaping of .2
- First Character must be two lines Drop capped.
- Paragraph before Spacing of 1 pt and After of 0 pt.
- Line Spacing of 1 pt
- Large Images must be in One Column
- Numbering of First Main Headings (Heading 1) must be in Roman Letters, Capital Letter, and Font Size of 10.
- Numbering of Second Main Headings (Heading 2) must be in Alphabets, Italic, and Font Size of 10.

You can use your own standard format also.

Author Guidelines:

1. General,

2. Ethical Guidelines,

3. Submission of Manuscripts,

4. Manuscript's Category,

5. Structure and Format of Manuscript,

6. After Acceptance.

1. GENERAL

Before submitting your research paper, one is advised to go through the details as mentioned in following heads. It will be beneficial, while peer reviewer justify your paper for publication.

Scope

The Global Journals welcome the submission of original paper, review paper, survey article relevant to the all the streams of Philosophy and knowledge. The Global Journals is parental platform for Global Journal of Computer Science and Technology, Researches in Engineering, Medical Research, Science Frontier Research, Human Social Science, Management, and Business organization. The choice of

specific field can be done otherwise as following in Abstracting and Indexing Page on this Website. As the all Global Journals are being abstracted and indexed (in process) by most of the reputed organizations. Topics of only narrow interest will not be accepted unless they have wider potential or consequences.

2. ETHICAL GUIDELINES

Authors should follow the ethical guidelines as mentioned below for publication of research paper and research activities.

Papers are accepted on strict understanding that the material in whole or in part has not been, nor is being, considered for publication elsewhere. If the paper once accepted by Global Journals and Editorial Board, will become the *copyright of the Global Journals.*

Authorship: The authors and coauthors should have active contribution to conception design, analysis and interpretation of findings. They should critically review the contents and drafting of the paper. All should approve the final version of the paper before submission

The Global Journals follows the definition of authorship set up by the Global Academy of Research and Development. According to the Global Academy of R&D authorship, criteria must be based on:

1) Substantial contributions to conception and acquisition of data, analysis and interpretation of the findings.

2) Drafting the paper and revising it critically regarding important academic content.

3) Final approval of the version of the paper to be published.

All authors should have been credited according to their appropriate contribution in research activity and preparing paper. Contributors who do not match the criteria as authors may be mentioned under Acknowledgement.

Acknowledgements: Contributors to the research other than authors credited should be mentioned under acknowledgement. The specifications of the source of funding for the research if appropriate can be included. Suppliers of resources may be mentioned along with address.

Appeal of Decision: The Editorial Board's decision on publication of the paper is final and cannot be appealed elsewhere.

Permissions: It is the author's responsibility to have prior permission if all or parts of earlier published illustrations are used in this paper.

Please mention proper reference and appropriate acknowledgements wherever expected.

If all or parts of previously published illustrations are used, permission must be taken from the copyright holder concerned. It is the author's responsibility to take these in writing.

Approval for reproduction/modification of any information (including figures and tables) published elsewhere must be obtained by the authors/copyright holders before submission of the manuscript. Contributors (Authors) are responsible for any copyright fee involved.

3. SUBMISSION OF MANUSCRIPTS

Manuscripts should be uploaded via this online submission page. The online submission is most efficient method for submission of papers, as it enables rapid distribution of manuscripts and consequently speeds up the review procedure. It also enables authors to know the status of their own manuscripts by emailing us. Complete instructions for submitting a paper is available below.

Manuscript submission is a systematic procedure and little preparation is required beyond having all parts of your manuscript in a given format and a computer with an Internet connection and a Web browser. Full help and instructions are provided on-screen. As an author, you will be prompted for login and manuscript details as Field of Paper and then to upload your manuscript file(s) according to the instructions.

To avoid postal delays, all transaction is preferred by e-mail. A finished manuscript submission is confirmed by e-mail immediately and your paper enters the editorial process with no postal delays. When a conclusion is made about the publication of your paper by our Editorial Board, revisions can be submitted online with the same procedure, with an occasion to view and respond to all comments.

Complete support for both authors and co-author is provided.

4. MANUSCRIPT'S CATEGORY

Based on potential and nature, the manuscript can be categorized under the following heads: Original research paper: Such papers are reports of high-level significant original research work.

Review papers: These are concise, significant but helpful and decisive topics for young researchers.

Research articles: These are handled with small investigation and applications

Research letters: The letters are small and concise comments on previously published matters.

5. STRUCTURE AND FORMAT OF MANUSCRIPT

The recommended size of original research paper is less than seven thousand words, review papers fewer than seven thousands words also. Preparation of research paper or how to write research paper, are major hurdle, while writing manuscript. The research articles and research letters should be fewer than three thousand words, the structure original research paper; sometime review paper should be as follows:

Papers*:* These are reports of significant research (typically less than 7000 words equivalent, including tables, figures, references), and comprise:

(a)*Title* should be relevant and commensurate with the theme of the paper.

(b) A brief Summary, *"Abstract"* (less than 150 words) containing the major results and conclusions.

(c) Up to *ten keywords*, that precisely identifies the paper's subject, purpose, and focus.

(d) An *Introduction*, giving necessary background excluding subheadings; objectives must be clearly declared.

(e) Resources and techniques with sufficient complete experimental details (wherever possible by reference) to permit repetition; sources of information must be given and numerical methods must be specified by reference, unless non-standard.

(f) Results should be presented concisely, by well-designed tables and/or figures; the same data may not be used in both; suitable statistical data should be given. All data must be obtained with attention to numerical detail in the planning stage. As reproduced design has been recognized to be important to experiments for a considerable time, the Editor has decided that any paper that appears not to have adequate numerical treatments of the data will be returned un-refereed;

(g) Discussion should cover the implications and consequences, not just recapitulating the results; *conclusions* should be summarizing.

(h) Brief Acknowledgements.

(i) References in the proper form.

Authors should very cautiously consider the preparation of papers to ensure that they communicate efficiently. Papers are much more likely to be accepted, if they are cautiously designed and laid out, contain few or no errors, are summarizing, and be conventional to the approach and instructions. They will in addition, be published with much less delays than those that require much technical and editorial correction.

The Editorial Board reserves the right to make literary corrections and to make suggestions to improve briefness.

It is vital, that authors take care in submitting a manuscript that is written in simple language and adheres to published guidelines.

Format

Language: The language of publication is UK English. Authors, for whom English is a second language, must have their manuscript efficiently edited by an English-speaking person before submission to make sure that, the English is of high excellence. It is preferable, that manuscripts should be professionally edited.

Standard Usage, Abbreviations, and Units: Spelling and hyphenation should be conventional to The Concise Oxford English Dictionary. Statistics and measurements should at all times be given in figures, e.g. 16 min, except for when the number begins a sentence. When the number does not refer to a unit of measurement it should be spelt in full unless, it is 160 or greater.

Abbreviations supposed to be used carefully. The abbreviated name or expression is supposed to be cited in full at first usage, followed by the conventional abbreviation in parentheses.

Metric SI units are supposed to generally be used excluding where they conflict with current practice or are confusing. For illustration, 1.4 l rather than 1.4 × 10-3 m3, or 4 mm somewhat than 4 × 10-3 m. Chemical formula and solutions must identify the form used, e.g. anhydrous or hydrated, and the concentration must be in clearly defined units. Common species names should be followed by underlines at the first mention. For following use the generic name should be constricted to a single letter, if it is clear.

Structure

All manuscripts submitted to Global Journals, ought to include:
Title: The title page must carry an instructive title that reflects the content, a running title (less than 45 characters together with spaces), names of the authors and co-authors, and the place(s) wherever the work was carried out. The full postal address in addition with the e-mail address of related author must be given. Up to eleven keywords or very brief phrases have to be given to help data retrieval, mining and indexing.
Abstract, used in Original Papers and Reviews:
Optimizing Abstract for Search Engines
Many researchers searching for information online will use search engines such as Google, Yahoo or similar. By optimizing your paper for search engines, you will amplify the chance of someone finding it. This in turn will make it more likely to be viewed and/or cited in a further work. Global Journals have compiled these guidelines to facilitate you to maximize the web-friendliness of the most public part of your paper.

Key Words
A major linchpin in research work for the writing research paper is the keyword search, which one will employ to find both library and Internet resources.
One must be persistent and creative in using keywords. An effective keyword search requires a strategy and planning a list of possible keywords and phrases to try.

Search engines for most searches, use Boolean searching, which is somewhat different from Internet searches. The Boolean search uses "operators," words (and, or, not, and near) that enable you to expand or narrow your affords. Tips for research paper while preparing research paper are very helpful guideline of research paper.

Choice of key words is first tool of tips to write research paper. Research paper writing is an art.A few tips for deciding as strategically as possible about keyword search:

- One should start brainstorming lists of possible keywords before even begin searching. Think about the most important concepts related to research work. Ask, "What words would a source have to include to be truly valuable in research paper?" Then consider synonyms for the important words.
- It may take the discovery of only one relevant paper to let steer in the right keyword direction because in most databases, the keywords under which a research paper is abstracted are listed with the paper.
- One should avoid outdated words.

Keywords are the key that opens a door to research work sources. Keyword searching is an art in which researcher's skills are bound to improve with experience and time.

Numerical Methods: Numerical methods used should be clear and, where appropriate, supported by references.

Acknowledgements: Please make these as concise as possible.

References

References follow the *Harvard scheme* of referencing. References in the text should cite the authors' names followed by the time of their publication, unless there are three or more authors when simply the first author's name is quoted followed by et al. unpublished work has to only be cited where necessary, and only in the text. Copies of references in press in other journals have to be supplied with submitted typescripts. It is necessary that all citations and references be carefully checked before submission, as mistakes or omissions will cause delays.

References to information on the World Wide Web can be given, but only if the information is available without charge to readers on an official site. Wikipedia and Similar websites are not allowed where anyone can change the information. Authors will be asked to make available electronic copies of the cited information for inclusion on the Global Journals homepage at the judgment of the Editorial Board.

The Editorial Board and Global Journals recommend that, citation of online-published papers and other material should be done via a DOI (digital object identifier). If an author cites anything, which does not have a DOI, they run the risk of the cited material not being noticeable.

The Editorial Board and Global Journals recommend the use of a tool such as Reference Manager for reference management and formatting.

Tables, Figures and Figure Legends

Tables: Tables should be few in number, cautiously designed, uncrowned, and include only essential data. Each must have an Arabic number, e.g. Table 4, a self-explanatory caption and be on a separate sheet. Vertical lines should not be used.

Figures: Figures are supposed to be submitted as separate files. Always take in a citation in the text for each figure using Arabic numbers, e.g. Fig. 4. Artwork must be submitted online in electronic form by e-mailing them.

Preparation of Electronic Figures for Publication

Even though low quality images are sufficient for review purposes, print publication requires high quality images to prevent the final product being blurred or fuzzy. Submit (or e-mail) EPS (line art) or TIFF (halftone/photographs) files only. MS PowerPoint and Word Graphics are unsuitable for printed pictures. Do not use pixel-oriented software. Scans (TIFF only) should have a resolution of at least 350 dpi (halftone) or 700 to 1100 dpi (line drawings) in relation to the imitation size. Please give the data for figures in black and white or submit a Color Work Agreement Form. EPS files must be saved with fonts embedded (and with a TIFF preview, if possible).

For scanned images, the scanning resolution (at final image size) ought to be as follows to ensure good reproduction: line art: >650 dpi; halftones (including gel photographs) : >350 dpi; figures containing both halftone and line images: >650 dpi.

Color Charges: It is the rule of the Global Journals for authors to pay the full cost for the reproduction of their color artwork. Hence, please note that, if there is color artwork in your manuscript when it is accepted for publication, we would require you to complete and return a color work agreement form before your paper can be published.

Figure Legends: Self-explanatory legends of all figures should be incorporated separately under the heading 'Legends to Figures'. In the full-text online edition of the journal, figure legends may possibly be truncated in abbreviated links to the full screen version. Therefore, the first 100 characters of any legend should notify the reader, about the key aspects of the figure.

6. AFTER ACCEPTANCE

Upon approval of a paper for publication, the manuscript will be forwarded to the dean, who is responsible for the publication of the Global Journals.

6.1 Proof Corrections

The corresponding author will receive an e-mail alert containing a link to a website or will be attached. A working e-mail address must therefore be provided for the related author.

Acrobat Reader will be required in order to read this file. This software can be downloaded

(Free of charge) from the following website:

www.adobe.com/products/acrobat/readstep2.html. This will facilitate the file to be opened, read on screen, and printed out in order for any corrections to be added. Further instructions will be sent with the proof.

Proofs must be returned to the dean at dean@globaljournals.org within three days of receipt.

As changes to proofs are costly, we inquire that you only correct typesetting errors. All illustrations are retained by the publisher. Please note that the authors are responsible for all statements made in their work, including changes made by the copy editor.

6.2 Early View of Global Journals (Publication Prior to Print)

The Global Journals are enclosed by our publishing's Early View service. Early View articles are complete full-text articles sent in advance of their publication. Early View articles are absolute and final. They have been completely reviewed, revised and edited for publication, and the authors' final corrections have been incorporated. Because they are in final form, no changes can be made after sending them. The nature of Early View articles means that they do not yet have volume, issue or page numbers, so Early View articles cannot be cited in the conventional way.

6.3 Author Services

Online production tracking is available for your article through Author Services. Author Services enables authors to track their article - once it has been accepted - through the production process to publication online and in print. Authors can check the status of their articles online and choose to receive automated e-mails at key stages of production. The authors will receive an e-mail with a unique link that enables them to register and have their article automatically added to the system. Please ensure that a complete e-mail address is provided when submitting the manuscript.

6.4 Author Material Archive Policy

Please note that if not specifically requested, publisher will dispose off hardcopy & electronic information submitted, after the two months of publication. If you require the return of any information submitted, please inform the Editorial Board or dean as soon as possible.

6.5 Offprint and Extra Copies

A PDF offprint of the online-published article will be provided free of charge to the related author, and may be distributed according to the Publisher's terms and conditions. Additional paper offprint may be ordered by emailing us at: editor@globaljournals.org .

Informal Tips for writing a Computer Science Research Paper to increase readability and citation

Before start writing a good quality Computer Science Research Paper, let us first understand what is Computer Science Research Paper? So, Computer Science Research Paper is the paper which is written by professionals or scientists who are associated to Computer Science and Information Technology, or doing research study in these areas. If you are novel to this field then you can consult about this field from your supervisor or guide.

Techniques for writing a good quality Computer Science Research Paper:

1. Choosing the topic- In most cases, the topic is searched by the interest of author but it can be also suggested by the guides. You can have several topics and then you can judge that in which topic or subject you are finding yourself most comfortable. This can be done by asking several questions to yourself, like Will I be able to carry our search in this area? Will I find all necessary recourses to accomplish the search? Will I be able to find all information in this field area? If the answer of these types of questions will be "Yes" then you can choose that topic. In most of the cases, you may have to conduct the surveys and have to visit several places because this field is related

to Computer Science and Information Technology. Also, you may have to do a lot of work to find all rise and falls regarding the various data of that subject. Sometimes, detailed information plays a vital role, instead of short information.

2. Evaluators are human: First thing to remember that evaluators are also human being. They are not only meant for rejecting a paper. They are here to evaluate your paper. So, present your Best.

3. Think Like Evaluators: If you are in a confusion or getting demotivated that your paper will be accepted by evaluators or not, then think and try to evaluate your paper like an Evaluator. Try to understand that what an evaluator wants in your research paper and automatically you will have your answer.

4. Make blueprints of paper: The outline is the plan or framework that will help you to arrange your thoughts. It will make your paper logical. But remember that all points of your outline must be related to the topic you have chosen.

5. Ask your Guides: If you are having any difficulty in your research, then do not hesitate to share your difficulty to your guide (if you have any). They will surely help you out and resolve your doubts. If you can't clarify what exactly you require for your work then ask the supervisor to help you with the alternative. He might also provide you the list of essential readings.

6. Use of computer is recommended: As you are doing research in the field of Computer Science, then this point is quite obvious.

7. Use right software: Always use good quality software packages. If you are not capable to judge good software then you can lose quality of your paper unknowingly. There are various software programs available to help you, which you can get through Internet.

8. Use the Internet for help: An excellent start for your paper can be by using the Google. It is an excellent search engine, where you can have your doubts resolved. You may also read some answers for the frequent question how to write my research paper or find model research paper. From the internet library you can download books. If you have all required books make important reading selecting and analyzing the specified information. Then put together research paper sketch out.

9. Use and get big pictures: Always use encyclopedias, Wikipedia to get pictures so that you can go into the depth.

10. Bookmarks are useful: When you read any book or magazine, you generally use bookmarks, right! It is a good habit, which helps to not to lose your continuity. You should always use bookmarks while searching on Internet also, which will make your search easier.

11. Revise what you wrote: When you write anything, always read it, summarize it and then finalize it.

12. Make all efforts: Make all efforts to mention what you are going to write in your paper. That means always have a good start. Try to mention everything in introduction, that what is the need of a particular research paper. Polish your work by good skill of writing and always give an evaluator, what he wants.

13. Have backups: When you are going to do any important thing like making research paper, you should always have backup copies of it either in your computer or in paper. This will help you to not to lose any of your important.

14. Produce good diagrams of your own: Always try to include good charts or diagrams in your paper to improve quality. Using several and unnecessary diagrams will degrade the quality of your paper by creating "hotchpotch." So always, try to make and include those diagrams, which are made by your own to improve readability and understandability of your paper.

15. Use of direct quotes: When you do research relevant to literature, history or current affairs then use of quotes become essential but if study is relevant to science then use of quotes is not preferable.

16. Use proper verb tense: Use proper verb tenses in your paper. Use past tense, to present those events that happened. Use present tense to indicate events that are going on. Use future tense to indicate future happening events. Use of improper and wrong tenses will confuse the evaluator. Avoid the sentences that are incomplete.

17. Never use online paper: If you are getting any paper on Internet, then never use it as your research paper because it might be possible that evaluator has already seen it or maybe it is outdated version.

18. **Pick a good study spot:** To do your research studies always try to pick a spot, which is quiet. Every spot is not for studies. Spot that suits you choose it and proceed further.

19. Know what you know: Always try to know, what you know by making objectives. Else, you will be confused and cannot achieve your target.

20. Use good quality grammar: Always use a good quality grammar and use words that will throw positive impact on evaluator. Use of good quality grammar does not mean to use tough words, that for each word the evaluator has to go through dictionary. Do not start sentence with a conjunction. Do not fragment sentences. Eliminate one-word sentences. Ignore passive voice. Do not ever use a big word when a diminutive one would suffice. Verbs have to be in agreement with their subjects. Prepositions are not expressions to finish sentences with. It is incorrect to ever divide an infinitive. Avoid clichés like the disease. Also, always shun irritating alliteration. Use language that is simple and straight forward. put together a neat summary.

21. Arrangement of information: Each section of the main body should start with an opening sentence and there should be a changeover at the end of the section. Give only valid and powerful arguments to your topic. You may also maintain your arguments with records.

22. Never start in last minute: Always start at right time and give enough time to research work. Leaving everything to the last minute will degrade your paper and spoil your work.

23. Multitasking in research is not good: Doing several things at the same time proves bad habit in case of research activity. Research is an area, where everything has a particular time slot. Divide your research work in parts and do particular part in particular time slot.

24. Never copy others' work: Never copy others' work and give it your name because if evaluator has seen it anywhere you will be in trouble.

25. Take proper rest and food: No matter how many hours you spend for your research activity, if you are not taking care of your health then all your efforts will be in vain. For a quality research, study is must, and this can be done by taking proper rest and food.

26. Go for seminars: Attend seminars if the topic is relevant to your research area. Utilize all your resources.

27. Refresh your mind after intervals: Try to give rest to your mind by listening to soft music or by sleeping in intervals. This will also improve your memory.

28. Make colleagues: Always try to make colleagues. No matter how sharper or intelligent you are, if you make colleagues you can have several ideas, which will be helpful for your research.

29. **Think technically:** Always think technically. If anything happens, then search its reasons, its benefits, and demerits.

30. Think and then print: When you will go to print your paper, notice that tables are not be split, headings are not detached from their descriptions, and page sequence is maintained.

31. Adding unnecessary information: Do not add unnecessary information, like, I have used MS Excel to draw graph. Do not add irrelevant and inappropriate material. These all will create superfluous. Foreign terminology and phrases are not apropos. One should NEVER take a broad view. Analogy in script is like feathers on a snake. Not at all use a large word when a very small one would be sufficient. Use words properly, regardless of how others use them. Remove quotations. Puns are for kids, not grunt readers. Amplification is a billion times of inferior quality than sarcasm.

32. Never oversimplify everything: To add material in your research paper, never go for oversimplification. This will definitely irritate the evaluator. Be more or less specific. Also too, by no means, ever use rhythmic redundancies. Contractions aren't essential and shouldn't be there used. Comparisons are as terrible as clichés. Give up ampersands and abbreviations, and so on. Remove commas, that are, not necessary. Parenthetical words however should be together with this in commas. Understatement is all the time the complete best way to put onward earth-shaking thoughts. Give a detailed literary review.

33. Report concluded results: Use concluded results. From raw data, filter the results and then conclude your studies based on measurements and observations taken. Significant figures and appropriate number of decimal places should be used. Parenthetical remarks are prohibitive. Proofread carefully at final stage. In the end give outline to your arguments. Spot out perspectives of further study of this subject. Justify your conclusion by at the bottom of them with sufficient justifications and examples.

34. After conclusion: Once you have concluded your research, the next most important step is to present your findings. Presentation is extremely important as it is the definite medium though which your research is going to be in print to the rest of the crowd. Care should

be taken to categorize your thoughts well and present them in a logical and neat manner. A good quality research paper format is essential because it serves to highlight your research paper and bring to light all necessary aspects in your research.

Informal Guidelines of Research Paper Writing

Key points to remember:

- Submit all work in its final form.
- Write your paper in the form, which is presented in the guidelines using the template.
- Please note the criterion for grading the final paper by peer-reviewers.

Final Points:

A purpose of organizing a research paper is to let people to interpret your effort selectively. The journal requires the following sections, submitted in the order listed, each section to start on a new page.

The introduction will be compiled from reference matter and will reflect the design processes or outline of basis that direct you to make study. As you will carry out the process of study, the method and process section will be constructed as like that. The result segment will show related statistics in nearly sequential order and will direct the reviewers next to the similar intellectual paths throughout the data that you took to carry out your study. The discussion section will provide understanding of the data and projections as to the implication of the results. The use of good quality references all through the paper will give the effort trustworthiness by representing an alertness of prior workings.

Writing a research paper is not an easy job no matter how trouble-free the actual research or concept. Practice, excellent preparation, and controlled record keeping are the only means to make straightforward the progression.

General style:

Specific editorial column necessities for compliance of a manuscript will always take over from directions in these general guidelines.

To make a paper clear

· Adhere to recommended page limits

Mistakes to evade

- Insertion a title at the foot of a page with the subsequent text on the next page
- Separating a table/chart or figure - impound each figure/table to a single page
- Submitting a manuscript with pages out of sequence

In every sections of your document

· Use standard writing style including articles ("a", "the," etc.)

· Keep on paying attention on the research topic of the paper

· Use paragraphs to split each significant point (excluding for the abstract)

· Align the primary line of each section

· Present your points in sound order

· Use present tense to report well accepted

· Use past tense to describe specific results

· Shun familiar wording, don't address the reviewer directly, and don't use slang, slang language, or superlatives

· Shun use of extra pictures - include only those figures essential to presenting results

Title Page:

Choose a revealing title. It should be short. It should not have non-standard acronyms or abbreviations. It should not exceed two printed lines. It should include the name(s) and address (es) of all authors.

Abstract:

The summary should be two hundred words or less. It should briefly and clearly explain the key findings reported in the manuscript--must have precise statistics. It should not have abnormal acronyms or abbreviations. It should be logical in itself. Shun citing references at this point.

An abstract is a brief distinct paragraph summary of finished work or work in development. In a minute or less a reviewer can be taught the foundation behind the study, common approach to the problem, relevant results, and significant conclusions or new questions.

Write your summary when your paper is completed because how can you write the summary of anything which is not yet written? Wealth of terminology is very essential in abstract. Yet, use comprehensive sentences and do not let go readability for briefness. You can maintain it succinct by phrasing sentences so that they provide more than lone rationale. The author can at this moment go straight to shortening the outcome. Sum up the study, with the subsequent elements in any summary. Try to maintain the initial two items to no more than one ruling each.

- Reason of the study - theory, overall issue, purpose
- Fundamental goal
- To the point depiction of the research
- Consequences, including definite statistics - if the consequences are quantitative in nature, account quantitative data; results of any numerical analysis should be reported
- Significant conclusions or questions that track from the research(es)

Approach:

- Single section, and succinct
- As a outline of job done, it is always written in past tense
- A conceptual should situate on its own, and not submit to any other part of the paper such as a form or table
- Center on shortening results - bound background information to a verdict or two, if completely necessary
- What you account in an conceptual must be regular with what you reported in the manuscript
- Exact spelling, clearness of sentences and phrases, and appropriate reporting of quantities (proper units, important statistics) are just as significant in an abstract as they are anywhere else

Introduction:

The **Introduction** should "introduce" the manuscript. The reviewer should be presented with sufficient background information to be

capable to comprehend and calculate the purpose of your study without having to submit to other works. The basis for the study should be offered. Give most important references but shun difficult to make a comprehensive appraisal of the topic. In the introduction, describe the problem visibly. If the problem is not acknowledged in a logical, reasonable way, the reviewer will have no attention in your result. Speak in common terms about techniques used to explain the problem, if needed, but do not present any particulars about the protocols here. Following approach can create a valuable beginning:

- Explain the value (significance) of the study
- Shield the model - why did you employ this particular system or method? What is its compensation? You strength remark on its appropriateness from a abstract point of vision as well as point out sensible reasons for using it.
- Present a justification. Status your particular theory (es) or aim(s), and describe the logic that led you to choose them.
- Very for a short time explain the tentative propose and how it skilled the declared objectives.

Approach:

- Use past tense except for when referring to recognized facts. After all, the manuscript will be submitted after the entire job is done.
- Sort out your thoughts; manufacture one key point with every section. If you make the four points listed above, you will need a least of four paragraphs.
- Present surroundings information only as desirable in order hold up a situation. The reviewer does not desire to read the whole thing you know about a topic.
- Shape the theory/purpose specifically - do not take a broad view.
- As always, give awareness to spelling, simplicity and correctness of sentences and phrases.

Procedures (Methods and Materials):

This part is supposed to be the easiest to carve if you have good skills. A sound written Procedures segment allows a capable scientist to replacement your results. Present precise information about your supplies. The suppliers and clarity of reagents can be helpful bits of information. Present methods in sequential order but linked methodologies can be grouped as a segment. Be concise when relating the protocols. Attempt for the least amount of information that would permit another capable scientist to spare your outcome but be cautious that vital information is integrated. The use of subheadings is suggested and ought to be synchronized with the results section. When a technique is used that has been well described in another object, mention the specific item describing a way but draw the basic principle while stating the situation. The purpose is to text all particular resources and broad procedures, so that another person may use some or all of the methods in one more study or referee the scientific value of your work. It is not to be a step by step report of the whole thing you did, nor is a methods section a set of orders.

Materials:

- Explain materials individually only if the study is so complex that it saves liberty this way.
- Embrace particular materials, and any tools or provisions that are not frequently found in laboratories.
- Do not take in frequently found.
- If use of a definite type of tools.
- Materials may be reported in a part section or else they may be recognized along with your measures.

Methods:

- Report the method (not particulars of each process that engaged the same methodology)
- Describe the method entirely
- To be succinct, present methods under headings dedicated to specific dealings or groups of measures
- Simplify - details how procedures were completed not how they were exclusively performed on a particular day.
- If well known procedures were used, account the procedure by name, possibly with reference, and that's all.

Approach:

- It is embarrassed or not possible to use vigorous voice when documenting methods with no using first person, which would focus the reviewer's interest on the researcher rather than the job. As a result when script up the methods most authors use third person passive voice.
- Use standard style in this and in every other part of the paper - avoid familiar lists, and use full sentences.

What to keep away from

- Resources and methods are not a set of information.
- Skip all descriptive information and surroundings - save it for the argument.
- Leave out information that is immaterial to a third party.

Results:

The principle of a results segment is to present and demonstrate your conclusion. Create this part a entirely objective details of the outcome, and save all understanding for the discussion.

The page length of this segment is set by the sum and types of data to be reported. Carry on to be to the point, by means of statistics and tables, if suitable, to present consequences most efficiently.

You must obviously differentiate material that would usually be incorporated in a study editorial from any unprocessed data or additional appendix matter that would not be available. In fact, such matter should not be submitted at all except requested by the instructor.
Content

- Sum up your conclusion in text and demonstrate them, if suitable, with figures and tables.
- In manuscript, explain each of your consequences, point the reader to remarks that are most appropriate.
- Present a background, such as by describing the question that was addressed by creation an exacting study.
- Explain results of control experiments and comprise remarks that are not accessible in a prescribed figure or table, if appropriate.
- Examine your data, then prepare the analyzed (transformed) data in the form of a figure (graph), table, or in manuscript form.

What to stay away from

- Do not discuss or infer your outcome, report surroundings information, or try to explain anything.
- Not at all, take in raw data or intermediate calculations in a research manuscript.
- Do not present the similar data more than once.
- Manuscript should complement any figures or tables, not duplicate the identical information.
- Never confuse figures with tables - there is a difference.

Approach

- As forever, use past tense when you submit to your results, and put the whole thing in a reasonable order.
- Put figures and tables, appropriately numbered, in order at the end of the report
- If you desire, you may place your figures and tables properly within the text of your results part.

Figures and tables

- If you put figures and tables at the end of the details, make certain that they are visibly distinguished from any attach appendix materials, such as raw facts
- Despite of position, each figure must be numbered one after the other and complete with subtitle
- In spite of position, each table must be titled, numbered one after the other and complete with heading
- All figure and table must be adequately complete that it could situate on its own, divide from text

Discussion:

The Discussion is expected the trickiest segment to write and describe. A lot of papers submitted for journal are discarded based on problems with the Discussion. There is no head of state for how long a argument should be. Position your understanding of the outcome visibly to lead the reviewer through your conclusions, and then finish the paper with a summing up of the implication of the study. The

purpose here is to offer an understanding of your results and hold up for all of your conclusions, using facts from your research and generally accepted information, if suitable. The implication of result should be visibly described.
Infer your data in the conversation in suitable depth. This means that when you clarify an observable fact you must explain mechanisms that may account for the observation. If your results vary from your prospect, make clear why that may have happened. If your results agree, then explain the theory that the proof supported. It is never suitable to just state that the data approved with prospect, and let it drop at that.

- Make a decision if each premise is supported, discarded, or if you cannot make a conclusion with assurance. Do not just dismiss a study or part of a study as "uncertain."
- Research papers are not acknowledged if the work is imperfect. Draw what conclusions you can based upon the results that you have, and take care of the study as a finished work
- You may propose future guidelines, such as how the experiment might be personalized to accomplish a new idea.
- Give details all of your remarks as much as possible, focus on mechanisms.
- Make a decision if the tentative design sufficiently addressed the theory, and whether or not it was correctly restricted.
- Try to present substitute explanations if sensible alternatives be present.
- One research will not counter an overall question, so maintain the large picture in mind, where do you go next? The best studies unlock new avenues of study. What questions remain?
- Recommendations for detailed papers will offer supplementary suggestions.

Approach:

- When you refer to information, differentiate data generated by your own studies from available information
- Submit to work done by specific persons (including you) in past tense.
- Submit to generally acknowledged facts and main beliefs in present tense.

Administration Rules Listed Before Submitting Your Research Paper to Global Journals

Please carefully note down following rules and regulation before submitting your Research Paper to Global Journals:

Segment Draft and Final Research Paper: You have to strictly follow the template of research paper. If it is not done your paper may get rejected.

- The **major constraint** is that you must independently make all content, tables, graphs, and facts that are offered in the paper. You must write each part of the paper wholly on your own. The Peer-reviewers need to identify your own perceptive of the concepts in your own terms. NEVER extract straight from any foundation, and never rephrase someone else's analysis.
- Do not give permission to anyone else to "PROOFREAD" your manuscript.

Written Material: You may discuss with your guides and key sources.

- Do not copy or imitate anyone else paper. (Various Methods to avoid Plagiarism is applied by us on every paper, if found guilty, you will be blacklisted by all of our collaborated research groups, your institution will be informed for this and strict legal actions will be taken immediately.)
- To guard yourself and others from possible illegal use please do not permit anyone right to use to your paper and files.

Criterion for Grading a Research Paper *(Compilation)* by Global Journals

Please note that following table is only a Grading of "Paper Compilation" and not on "Performed/Stated Research" whose grading solely depends on Individual Assigned Peer Reviewer and Editorial Board Member. These can be available only on request and after decision of Paper. This report will be the property of Global Journals.

Topics	Grades		
	A-B	**C-D**	**E-F**
Abstract	Clear and concise with appropriate content, Correct format. 200 words or below	Unclear summary and no specific data, Incorrect form Above 200 words	No specific data with ambiguous information Above 250 words
Introduction	Containing all background details with clear goal and appropriate details, flow specification, no grammar and spelling mistake, well organized sentence and paragraph, reference cited	Unclear and confusing data, appropriate format, grammar and spelling errors with unorganized matter	Out of place depth and content, hazy format
Methods and Procedures	Clear and to the point with well arranged paragraph, precision and accuracy of facts and figures, well organized subheads	Difficult to comprehend with embarrassed text, too much explanation but completed	Incorrect and unorganized structure with hazy meaning
Result	Well organized, Clear and specific, Correct units with precision, correct data, well structuring of paragraph, no grammar and spelling mistake	Complete and embarrassed text, difficult to comprehend	Irregular format with wrong facts and figures
Discussion	Well organized, meaningful specification, sound conclusion, logical and concise explanation, highly structured paragraph reference cited	Wordy, unclear conclusion, spurious	Conclusion is not cited, unorganized, difficult to comprehend
References	Complete and correct format, well organized	Beside the point, Incomplete	Wrong format and structuring

Index

Q

R

S

T

V

W

Global Journal of Computer Science and Technology

Visit us on the Web at www.GlobalJournals.org | www.ComputerResearch.org
or email us at helpdesk@globaljournals.org

ISSN 9754350